NB_Re^3: A Novel Framework for Reconstructing High-Quality Reflectance Time Series Taking Full Advantage of High-Quality NDVI and Multispectral Autocorrelations

Hongtao Shu, Zhuoning Gu, Yang Chen, Hui Chen, Xuehong Chen, and Jin Chen

***Abstract*—Multispectral reflectance—signals reflected from the Earth's surface across different wavelengths—is a primary data source for most remote sensing applications. However, obtaining complete cloud-free multispectral reflectance time series during the vegetation growing season remains challenging due to cloud contamination and limitations of existing reconstruction methods. To address the challenge, this study proposed a novel normalized difference vegetation index (NDVI)-guided bi-directional recurrent reconstruction model for multispectral reflectance time series (referred to as "NB-Re^3"), which aimed to reconstruct dense time series of reflectance images by exploiting the dependence of NDVI on multispectral reflectance. NB-Re^3 utilizes a temporal convolutional network to capture the temporal trends in the NDVI data and a bidirectional long short-term memory to integrate the temporal features of the NDVI with the cloud-free reflectance data. The architecture establishes a robust dynamic NDVI-reflectance relationship while capturing temporal dependencies and multispectral autocorrelations of multiple spectral bands. We compared the performance of NB-Re^3 with four representative methods (MNSPI, HANTS, STAIR, and U-TILSE) in reconstructing multispectral reflectance time series, ranging from the visible bands to near-infrared and short-wave infrared bands, at two challenging sites: the irrigated area of Colleambally, Australia, and the cultivated area of Rikaze on the Southern Tibetan Plateau, China. The result showed that NB-Re^3 kept superiority with the lowest root-mean-square error values and highest correlation coefficients values. The effectiveness of integrating high-quality NDVI time series and using multispectral autocorrelation to improve reflectance time-series reconstruction was further confirmed by the ablation experiments. It is concluded that NB-Re^3 shows promise for generating long-term cloud-free reflectance time-series products tailored for ecological and agricultural applications.**

***Index Terms*—Bidirectional long short-term memory (Bi-LSTM), multispectral reflectance data, normalized difference vegetation index (NDVI), spectral autocorrelation, temporal convolutional network (TCN), time series.**

I. Introduction

MULTISPECTRAL reflectance, which represents the signals reflected from the Earth's surface across different wavelength ranges, is a fundamental data source for most remote sensing applications, such as the calculation of vegetation indices [e.g., normalized difference vegetation index (NDVI)], classification and change detection, target detection, and the derivation of biophysical and biochemical parameters. An increasing number of Earth observation satellites, such as MODIS, VIIRS, Landsat, and Sentinel, have collected and released time-series surface reflectance (SR) data in different spectral bands. These high spatiotemporal resolution data provide irreplaceable opportunities to monitor the continuous, dense dynamics, and capture accurate spatial details of the Earth's surface in various applications, such as disaster management, ecosystem and biodiversity conservation, urban planning and infrastructure development, as well as agriculture and crop monitoring [1], [2], [3], [4], [5], [6], [7], [8].

Unfortunately, the actual acquisition of SR time series is often discontinuous and incomplete in time due to cloud contamination and other disturbances [9], [10], [11], [12]. More importantly, most satellite sensors are designed with tradeoffs between spatial and temporal resolution under technical and budgetary constraints. These two aspects account for the lack of cloud-free SR time series with both high spatial and temporal resolution when using single sensor data [13]. Recently, newly developed commercial satellite systems, such as the PlanetScope constellation [14], with a spatial resolution of 3 m and a daily revisit cycle, have largely alleviated this shortage. However, given the high cost of acquiring and preprocessing such constellation data, there is still a widespread need to generate cloud-free SR time series with 30 or 10-m spatial resolution, especially for historical data.

In response to these shortcomings, a number of methods have been developed in recent years to produce cloud-free SR time series with finer spatial resolution. These methods can be divided into two types based on whether they make use of multisource image data: single-sensor-based methods and multisensor-based

Manuscript received 5 May 2024; revised 17 June 2024; accepted 26 June 2024. Date of publication 1 July 2024; date of current version 24 July 2024. This work was supported by the National Key Research and Development Program of China under Grant 2022YFD2001101. *(Corresponding author: Jin Chen.)*

Hongtao Shu, Zhuoning Gu, Hui Chen, Xuehong Chen, and Jin Chen are with the State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China, and also with the Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China (e-mail: shuhongtao@mail.bnu.edu.cn; guzhn1999@gmail.com; chen_hui@mail.bnu.edu.cn; chenxuehong@bnu.edu.cn; chenjin@bnu.edu.cn).

Yang Chen is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: chenyang01@aircas.ac.cn).

The code will be available at https://github.com/HeartBread/NB_Re3.

Digital Object Identifier 10.1109/JSTARS.2024.3421237

methods. The single-sensor-based approach consists of two main strategies: the "Filling" strategy and the "Fitting" strategy. The "Filling" strategy aims to fill in the cloud-contaminated (hereafter referred to as "missing") values of a pixel using the values of cloud-free pixels in images taken by the same sensor at temporally close dates. Neighborhood similar pixel interpolator (NSPI) method is a typical one belonging to this strategy with the function of filling the missing values of pixels caused by the failure of the scan-line corrector (SLC) and cloud contamination in the Landsat ETM+ images [15], [16]. Another spectral-angle-mapper (SAM)-based spatiotemporal similarity (SAMSTS) method is designed to fill large area gaps in Landsat data using a SAM metric for identifying similar pixels [17]. Unlike the "Filling" strategy, the "Fitting" strategy assumes that there is a temporal change pattern in the original SR time series that can be fitted using a mathematic function, such as linear harmonic model [18] or the Whittaker smoother [19]. For example, Zhu et al. [18] introduced a linear harmonic function to generate synthetic Landsat time-series images based on all available Landsat reflectance time series. In general, the "Filling" strategy is more suitable for generating single synthetic image, while the "Fitting" strategy is more suitable for generating image time series. In addition, some hybrid methods have also been proposed to integrate their advantages, respectively. Yan and Roy [20] proposed a hybrid fill-and-fit approach, which integrates the "Filling" strategy and the "Fitting" strategy in two steps, and applied it to reconstruct analysis ready data time series. Nonetheless, methods based on a single sensor are inherently constrained by the number of available cloud-free images and their temporal distribution, due to the limited observations a single sensor can provide. This limitation is apparent in both the "Filling" strategy, which requires a substantial number of spatially or temporally adjacent cloud-free pixels, and the "Fitting" strategy, which heavily depends on a dense sequence of temporally cloud-free observations.

To address the limitations of single-sensor-based methods, multisensor-based methods have recently gained increasing attention by combining the advantages of data from two or more sensors. These methods employ the "Fusing" strategy, generating images with both high spatial and temporal resolution by integrating data with high spatial resolution but low temporal resolution (e.g., Landsat imagery, called fine imagery later) and data with high temporal resolution but low spatial resolution (e.g., MODIS imagery, called coarse imagery later). The spatial and temporal adaptive reflectance fusion model (STARFM) was the first method to fuse multisensor data based on spatiotemporal fusion (STF) idea [13]. Since then, more than 100 STF methods have been developed based on different assumptions [21], examples including: ESTARFM [22] as a spatial-weight-function-based method, MMT [23] and STDFA [24] as unmixing-based methods, Fit-FC [25] and FIRST [26] as regression-based methods, SPSTFM [27] and ELM [28] as machine-learning-based methods, and FSDAF [29] and Agri-Fuse [30] as hybrid methods integrating two or more aforementioned categories. In addition to these STF-based multisensor methods, some hybrid methods integrating "Fusing" strategy and other strategies have also been proposed. One notable example is the satellite data integration (STAIR) [31], a hybrid automated method that integrated time-series "Filling" and "Fusing" strategy into a two-step mode to generate high-frequency, high-spatial resolution, and high-quality reflectance time series. Although the multisensor methods show considerable advantages in combining the strengths of data from different sensors, they rely on some rigid assumptions, such as that all coarse pixels are considered pure and that temporal changes within the same land cover class are consistent across different resolutions. These assumptions do not hold in real-world scenarios, especially when dealing with heterogeneous landscapes and complex land cover changes. Moreover, they set parameters empirically, leaving room for further improvement in their fusion accuracy.

Recent advances in deep learning (DL) models have significantly improved remote sensing data reconstruction capacity [32], [33], [34]. The abilities of DL-based models to automatically extract high-dimensional features and establish nonlinear relationships have led to remarkable improvements over traditional models [35]. These DL-based approaches often incorporate ancillary information, such as synthetic aperture radar (SAR) signals [36], [37], [38] to reconstruct multispectral reflectance images. Most of these works have focused on better exploitation of the complex relationships between ancillary data and multispectral reflectance through the use of different network architectures, including multibranch networks [39], [40], [41]; attention-based networks [42], [43]; generative adversarial networks [44], [45]; transformer encoder–decoder-based networks [46], [47]; and hybrid networks integrating two or more architectures [48], [49], [50]. However, to our knowledge, only a few studies [46], [47], [51], [52], [53] have focused on exploring the distinct long-term temporal trends inherent in reflectance time series to reconstruct reflectance time series. Moreover, the introduced ancillary data, such as SAR data, may not always be beneficial due to the instability and fluctuating noise from its active coherent imaging systems, which could degrade visual appearance [54].

Compared to multispectral reflectance time series, NDVI time series, derived from near-infrared (NIR) and red (R) reflectance, are the most common representative indicators of leaf chlorophyll content and vegetation cover, and they are closely related to ecological issues under different scenarios, e.g., drought conditions and greenness exposure [55], [56], which could help to understand the ecosystem response to climate change. Therefore, the reconstructions of NDVI time series have received more attention. In addition, given the strong correlation between NDVI values and reflectance values in certain spectral bands, such as green, red, and near infrared bands [57], this shed light on the possibility of utilizing high-quality NDVI time series as a complementary data to help reconstruct multispectral reflectance time series. The principles of existing NDVI reconstruction methods are mainly based on the periodical nature of the vegetation growth state, and the consistent impacts of noise factors, such as clouds and shadows, which typically reduce NDVI values. Thus, NDVI time series can be effectively denoised and smoothed more directly by applying an upper envelope smoothing process with various temporal filters. These

temporal filters, such as Harmonic (Fourier) models [58], [59], [60], [61] and Savitzky–Golay (SG) filter [62], [63], have been widely used to obtain a continuous time series NDVI derived by single sensor. "Fusing" strategies such as the linear mixing growth model and singular value decomposition have been employed to facilitate the blending of NDVI time series from multiple sensors [64], [65]. Chen et al. [66] further proposed an innovative approach by combining the "Fusing" and "Fitting" strategies, synthesizing 30-m cloud-free NDVI time series from MODIS and Landsat observations, and then, filtering with a weighted SG filter. Beyond these traditional methods, DL-based methods have also been employed in the NDVI reconstruction recently, most introducing complementary data (e.g., meteorological and topographic data, SAR) to fill NDVI gaps [67], [68], [69], while the complexity of SAR and other ancillary data remains a challenge to exploring a stable relationship between NDVI and these ancillary data. Overall, an unclear and complex relationship between ancillary data and multispectral reflectance could increase the difficulties of constructing stable features in models, which requires the exploration of a stable and reasonable ancillary data to guide the reconstruction of multispectral reflectance time series.

In light of the aforementioned review, we proposed a novel NDVI guided bidirectional recurrent reconstruction model for multispectral reflectance time series (called NB-Re^3) to reconstruct high-quality multispectral reflectance time series by taking full advantage of high-quality NDVI time series. The new method is designed to capture the temporal trend of reflectance itself and the potential relationship between reflectance and NDVI time series. In summary, we expect the NB-Re^3 can address the following challenges in previous reconstruction methods for high-quality multispectral reflectance time series:

1) the relationship between existing introduced ancillary data and multispectral reflectance remains unclear and complex;
2) most of the existing reconstruction methods do not focus on the long-term temporal dependence of reflectance itself, as well as the dynamic relationship between multispectral reflectance and ancillary data.

In the following sections, we first introduce the NB-Re^3, and then, apply it to two commonly used datasets, Landsat-8 and Sentinel-2, to investigate its stability and validity in complex missing scenarios.

II. Methodology

To reconstruct high-quality multispectral reflectance time series, our approach attempted to establish a stable relationship between multispectral reflectance and NDVI time series, the latter providing a continuous pattern of temporal variation. Specifically, it consisted of three main modules (see Fig. 1): a cloud and shadow identification module based on the automatic time series analysis (ATSA) method [70], an NDVI reconstruction module based on the state-of-the-art gap filling and Savitzky–Golay filtering method (GF-SG) [66], which generated cloud-free NDVI time series, and a DL-based module for reconstructing high-quality multispectral reflectance time series, in which temporal convolution network (TCN) was used to extract high-level temporal trend features from NDVI time series, and bidirectional long short-term memory (Bi-LSTM) was used to build the complex relationship between NDVI and reflectance time series and the temporal trend of reflectance itself.

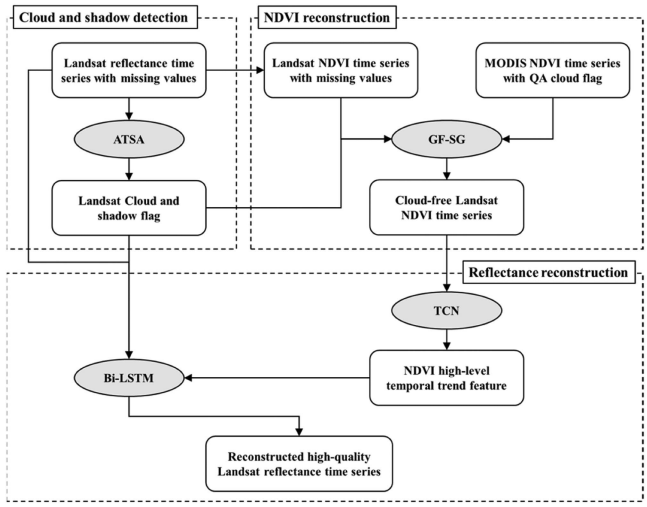

Fig. 1. Framework of the NB-Re^3 with three modules using Landsat-8 data as an example.

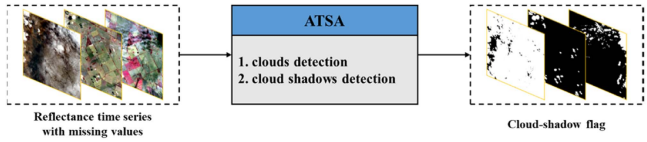

Fig. 2. Flowchart of the cloud-shadow detection using ATSA.

A. Cloud and Shadow Detection Module

For each image in the input multispectral reflectance image time series, pixels contaminated by clouds and cloud shadows must first be detected as missing values for subsequent reconstruction. To reduce cloud and shadow detection errors (e.g., misdetection of bright objects as cloud, omission of thin clouds, and failure to detect cloud shadows) in existing methods due to the imperfections of cloud detection algorithm [71], [72], [73], the ATSA method was used to automatically detect clouds and cloud shadows in this study (see Fig. 2). ATSA was specifically designed to detect clouds and cloud shadows in multispectral reflectance image time series based on the spectral characteristics of the clouds, as well as continuity feature of the land surface dynamics represented by the multispectral reflectance time series. The method first calculated the haze optimal transformation (HOT) index [74] from each multispectral reflectance image. This HOT index then served as input to the unsupervised classifier to identify clouds over land surfaces and water bodies. A thresholding method was further used to eliminate the misdetection of bright objects and missing thin clouds. In addition, ATSA calculated potential cloud shadows from the geometric relationship between clouds and cloud shadows

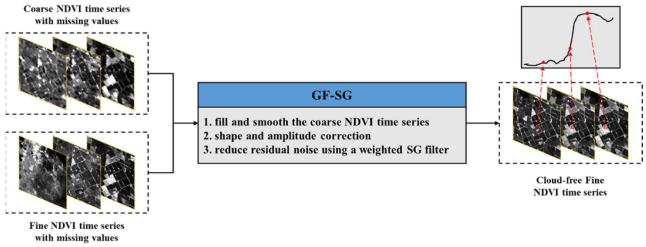

Fig. 3. Flowchart of the NDVI reconstruction using GF-SG.

using information about the sun's position [75]. The potential cloud shadows were further refined by time-series analysis of a shadow index, which was constructed based on the property that shadowed pixels are much darker than surrounding clear pixels in certain bands [73]. Thus, we adopted ATSA based on its excellent performance on Landsat-8 images and Sentinel-2 time-series images [70].

B. NDVI Reconstruction Module

GF-SG served as an effective way to generate cloud-free NDVI time series with both high spatial and temporal resolutions. It combined the advantages of coarse imagery and fine imagery of different sensors (see Fig. 3), and further applied a weighted SG filter to reduce the residual noise in the synthesized NDVI time series.

To reconstruct cloud-free NDVI time series, GF-SG involved three key steps. First, the coarse NDVI time series calculated from coarse images were filled by linear interpolation, and then, smoothed by the SG filter to remove as much cloud contamination as possible. Bicubic interpolation was used to further resample the coarse NDVI time series to fine resolution with fewer artifacts. Second, a linear transfer function was used to correct the shape and amplitude mismatch between the coarse and fine NDVI time series, which was mainly due to the mixed pixel effect and different spectral response functions between two sensors. The resampled and corrected NDVI time series were then used to fill missing values in the fine resolution NDVI time series. Finally, the residual noise was further reduced using a weighted SG filter.

We adopted GF-SG here because of its significant advantages. First, it improves the reconstruction quality of long-term continuous missing values in fine-resolution NDVI time series, whereas the other methods are less reliable for reconstructing these long data gaps. Second, the performance of GF-SG is less affected by the residual noise caused by cloud detection errors in the fine resolution images, which is due to the incorporation of the weighted SG filter. Third, GF-SG is simple and can be implemented on the computing platform Google Earth engine (GEE), which is particularly important for the practical applications at a large spatial scale.

C. Reflectance Reconstruction Module

We developed two complementary modules, TCN and Bi-LSTM, to achieve two specific objectives (see Fig. 4). The

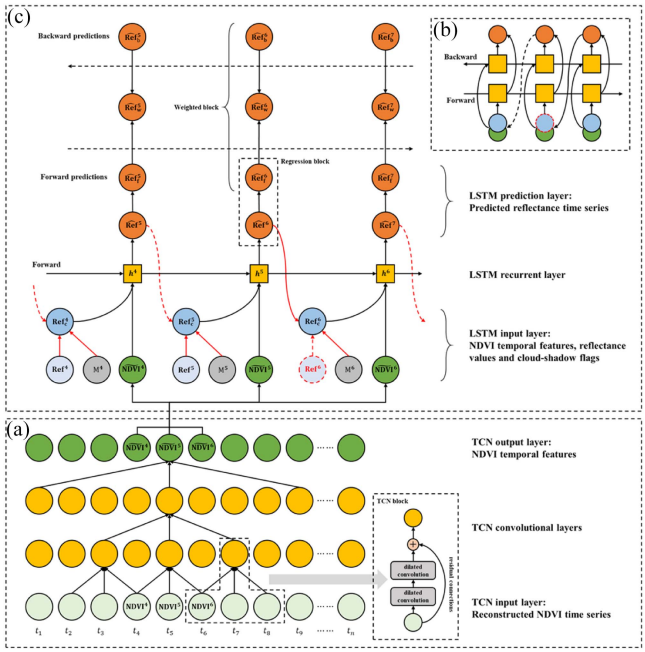

Fig. 4. Two complementary modules for reconstructing high-quality reflectance time series, taking the reconstruction of image at t_6 (Ref^6 in red presents the missing value) as an example. (a) TCN schematic diagram based on the 1-D fully convolutional network architecture. (b) Framework of Bi-LSTM in both forward and backward directions. (c) Detailed architecture of the forward direction in Bi-LSTM (the solid lines represent the paths through the input variables, and the dashed lines represent the paths not through the input variables. The red lines represent the "complement" operation, which preserves the values of the original cloud-free observations or the predictions based on the cloud-shadow flag).

TCN was tailored to extract high-level temporal trend features from NDVI time series across different temporal scales, while the Bi-LSTM was designed to capture the temporal trend of reflectance time series and to establish a robust dynamic relationship between NDVI and cloud-free reflectance values.

A TCN module was introduced to effectively extract temporal features and long-term dependencies from the reconstructed NDVI time series [76]. This module was structured based on a three-layer 1-D fully convolutional network framework, which incorporated dilated convolutions and residual connections [see Fig. 4(a)]. The first convolutional layer targeted capturing the local temporal trends of the reconstructed cloud-free NDVI time series, and as the number of layers l increased, the network gradually incorporated longer time series to learn the longer term dependencies of NDVI time series. In each layer, the dilated convolution employed a dilation factor d to effectively expand the receptive field of the network by skipping d steps in the input time series. This structure not only considered the temporal neighborhood information, but also broadened the scope of temporal features that can be captured. The residual connection was used to interlink two layers, incorporating temporal features of different layers, which helps to effectively learn long-term dependencies and improve network stabilization [77]. Specifically, each layer applied a filter $f(i) : \{i = -\frac{k-1}{2}, \ldots, 0, \ldots, \frac{k-1}{2}\}$ with the dilation factor d and the kernel size of k to the

time-series input as follows:

$$s^t = \sum_{i=-\frac{k-1}{2}}^{\frac{k-1}{2}} f(i) \cdot \text{NDVI}^{t+d\cdot i}$$

$$O\left(s^t\right) = \text{Activation}\left(s^t + \mathcal{F}\left(s^t\right)\right) \quad (1)$$

where $O(s^t)$ represents the output feature. s^t is the dilated convolution operation on the $NDVI$ value at time step t. $\mathcal{F}(s^t)$ represents a 1-D convolutional operation on s^t with the kernel size of k set to 1, which was designed to unify the channel numbers of input and output. Activation$(*)$ represents activation function from the combination of residual connections, s^t and $\mathcal{F}(s^t)$. Overall, our TCN module aimed to learn NDVI temporal features at different scales, from local temporal trends to long-term temporal trends.

The Bi-LSTM network was conducted to establish a robust long-term temporal relationship between NDVI and multispectral reflectance time series, as well as to capture the interrelations among the multispectral reflectance bands themselves. The network learned the NDVI-reflectance long-term temporal relationship in both forward and backward directions, effectively mitigating gradient vanishing, explosion problems and capturing temporal trend at different temporal scales when training long-term time series [78]. Taking the forward direction as an example, a recurrent imputation module consisted of the following components [see Fig. 4(c)]: an input layer; an LSTM recurrent layer that obtained the hidden-state features; and a prediction layer that transformed the hidden-state features to the prediction values at next time step through a fully connected layer.

The input layer provided the necessary input data for the following hidden layer. It consisted of three key input variables at time step t, NDVI temporal features $\widetilde{\text{NDVI}}^t$ from TCN, original reflectance value Ref^t and binary cloud-shadow flag M^t (with the cloud-free observations labeled as value 0 and the missing observations as value 1) Due to the missing values in the original reflectance time series, it would result in the LSTM recurrent layer not being able to back-propagate. Instead, Ref_c^t calculated from the following equation was input ("complement" operation in Fig. 4), where Ref^t with the cloud-shadow flag M^t equal to 1 (missing) was replaced by the predicted reflectance value.

$$\text{Ref}_c^t = \widehat{\text{Ref}}^t \times M^t + \text{Ref}^t \times \left(1 - M^t\right) \quad (2)$$

where $\widehat{\text{Ref}}^t$ presents the prediction reflectance value at time step t by the LSTM model from the previous time step. Ref^t presents the original reflectance at time step t, if it is cloud free.

The LSTM recurrent layer aimed to build the NDVI-reflectance dynamic relationship and update the hidden-state features h^t from Ref_c^t and $\widetilde{\text{NDVI}}^t$ as follows [79]:

$$h^t = \sigma\left(W_h \cdot h^{t-1} + U_h \cdot x^t + b_h\right) \quad (3)$$

where σ is the sigmoid function. W_h and U_h are the learnable weight parameters, and b_h is the learnable bias parameters. x^t is the concatenation of input variables Ref_c^t and corresponding $\widetilde{\text{NDVI}}^t$.

Considering that the missing patterns caused by random clouds and cloud shadows in the time series are irregular, a temporal decay factor γ^t [80], [81] was introduced based on the temporal decay interval δ^t to consider the temporal dependence that the closer cloud-free observation have a higher contribution to h^t as follows:

$$\delta^t = \begin{cases} \text{doy}^t - \text{doy}^{t-1} + M^{t-1} \times \delta^{t-1}, & \text{if } t > 1 \\ 0, & \text{if } t = 0 \end{cases}$$

$$\gamma^t = \exp\left\{-\max\left(0, W_\gamma \cdot \delta^t + b_\gamma\right)\right\} \quad (4)$$

where δ^t is defined as the time intervals at time step t between the last clear observation and the corresponding date doy^t. M^{t-1} is the cloud-shadow flag at previous time step. b_γ is the learnable bias parameters along with the learnable weight parameters W_γ. Finally, the modified recurrent network was updated from (3) as follows:

$$h^t = \sigma\left(W_h \cdot \left[h^{t-1} \times \gamma^t\right] + U_h \cdot x^t + b_h\right). \quad (5)$$

Next, the LSTM recurrent layer transformed the hidden-state features h^t to the prediction values $\widehat{\text{Ref}}^{t+1}$ at next time step $t+1$ through a fully connected layer as follows:

$$\widehat{\text{Ref}}^{t+1} = \hat{W} \cdot h^t + \hat{b} \quad (6)$$

where $\hat{b}$ is the learnable bias parameters along with the learnable weight parameters $\hat{W}$ obtained by the training process.

The process described previously is to predict individual spectral band only using its temporal self-dependence. In fact, when our task is to predict the reflectance of multiple bands, spectral autocorrelation between multiple spectral bands is valuable for improving prediction accuracy [25], [26]. Accordingly, in the prediction layer, the spectral autocorrelation between multiple bands were then considered to construct more robust predictions. These spectral autocorrelations were estimated using a linear regression block in Fig. 4 after the reflectance $\widehat{\text{Ref}}^{t+1}$ was predicted.

$$\widehat{\text{Ref}}_{f,i}^{t+1} = \sum_{k=1}^{n}\left(W_{f,i} * \widehat{\text{Ref}}_{f,k}^{t+1} + b_{f,i}\right) \quad (7)$$

where $\widehat{\text{Ref}}_{f,i}^{t+1}$ is the regressed ith band prediction in the forward direction at the time step $t+1$. Here, $\widehat{\text{Ref}}_{f,k}^{t+1}$ represents the reflectance value predicted for each spectral band according to (6); n is the band number; and $b_{f,i}$ is the learnable bias parameters along with the learnable weight parameters $W_{f,i}$.

Moreover, recognizing the potential instability of unidirectional prediction, we used a learnable weighting combination of the predictions [derived from (7)] of both forward and backward directions in the prediction layer

$$\widehat{\text{Ref}}_w^{t+1} = W_f^{t+1} \cdot \widehat{\text{Ref}}_f^{t+1} + \left(1 - W_f^{t+1}\right) \cdot \widehat{\text{Ref}}_b^{t+1} \quad (8)$$

where $\widehat{\text{Ref}}_w^{t+1}$ is the weighted prediction at the time step $t+1$. $\widehat{\text{Ref}}_f^{t+1}$ and $\widehat{\text{Ref}}_b^{t+1}$ are final predictions of all bands obtained from both forward and backward directions, respectively. W_f^{t+1}

represents the learnable weight of the forward directions learned by network itself.

The loss function $\mathcal{L}_f$ in one direction (i.e., forward direction) was estimated from the mean absolute error (MAE) between the predicted reflectance values $\widehat{\text{Ref}}_f^t$ and the corresponding original cloud-free reflectance values Ref^t.

$$\mathcal{L}_f = \frac{1}{T}\sum_{t=1}^{T}\left(\left(1-M^t\right)\times\left|\widehat{\text{Ref}}_f^t - \text{Ref}^t\right|\right) \tag{9}$$

where T is the length of the time series. At the same time, same operation was performed to calculate the loss $\mathcal{L}_b$ in the backward direction. Furthermore, the weighted prediction combined forward and backward prediction was also considered in loss calculation. Thus, we used a weighted loss, $\mathcal{L}_w$, to calculate the MAE between cloud-free reflectance values Ref^t and weighted prediction $\widehat{\text{Ref}}_w^t$ obtained from (8). The final loss function was obtained by adding the forward loss $\mathcal{L}_f$, the backward loss $\mathcal{L}_b$, and the weighted loss $\mathcal{L}_w$ as follows:

$$\mathcal{L}_w = \frac{1}{T}\sum_{t=1}^{T}\left(\left(1-M^t\right)\times\left|\widehat{\text{Ref}}_w^t - \text{Ref}^t\right|\right)$$

$$\mathcal{L}_{\text{final}} = \mathcal{L}_f + \mathcal{L}_b + \mathcal{L}_w. \tag{10}$$

D. Model Implementation

The proposed model was trained using an adaptive moment estimation (Adam) optimizer [82]. We set a batch size of 256 and a learning rate of 0.001 to optimally determine the network weights, aiming for the global minimum of the loss function. The training process spanned ten epochs, after which the training loss stabilized and stopped decreasing. For the TCN, the kernel size was configured constant to 3, and the corresponding dilated convolution was set to (1, 2, 4) over (32, 64, 128) units. The hidden unit of the LSTM was set to 96. These methods were implemented using the PyTorch DL package and trained on an Intel Core i9-10850K CPU at 3.60 GHz, equipped with 56 GB of RAM and an NVIDIA GeForce RTX 3070 GPU with 8 GB of memory.

NB-Re^3 is an end-to-end model with self-training function. It means that in practice, users only need to input cloud and shadow data generated by the ATSA method, high-quality NDVI time series generated by GF-SG and original multispectral reflectance time series, and then, obtain the reconstructed multispectral reflectance time-series images. In detail, the model was locally trained using the training pixels obtained by a uniformly distributed spatial sampling strategy (see Fig. 5), the pixels selected as the training pixels were input to NB-Re^3 to obtain the learnable parameters of the model using the cloud-free observations of these pixels to calculate loss function, and then, the well-trained NB-Re^3 was applied to predict the remaining pixels. Here, we adopted a uniformly distributed spatial sampling strategy instead of using all pixels for training, based on the fact that both clouds and land cover are locally spatially continuous, and the sampled pixels centered in the sampling window (e.g., 3 × 3) are somewhat representative of the surrounding pixels. Such a training process cannot only significantly reduce the training time, but also cover almost all possible conditions.

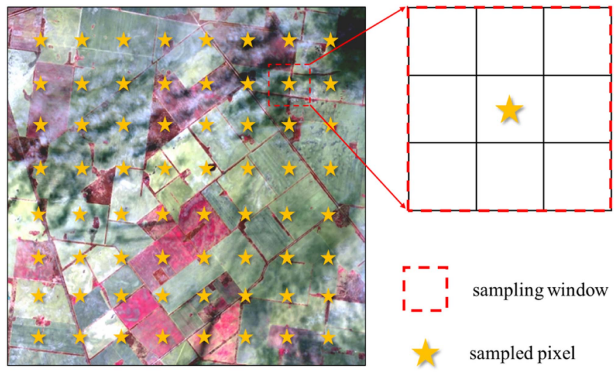

Fig. 5. Uniform spatial sampling strategy. The red dotted box represents the sampling window (the window size was set to 3 as an example). The stars in the center of the box represent the sampled pixels, which were treated as training samples and fed into the NB-Re^3 network.

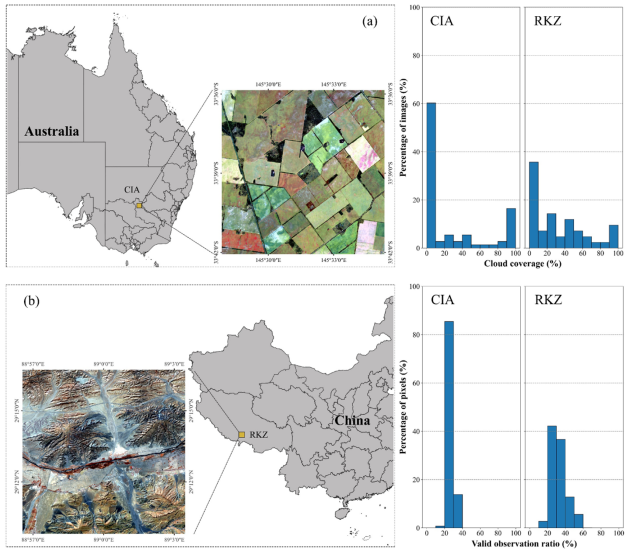

Fig. 6. Locations and observation conditions of testing areas. (a) CIA, the southern New South Wales, Australia. (b) Rikaze crop land (RKZ), the southern Qinghai-Tibetan plateau, China. All images use NIR–Red–Green as RGB. The top right plot indicates the proportional of images in the time series with different cloud coverages, and the bottom right plot indicates the percentage of pixels with different valid observation ratios in time series.

III. Data Preparation and Experiments

A. Testing Areas

We validated the NB-Re^3 at two distinct areas that represent homogeneous and heterogeneous landscape, respectively, each with different levels of cloud contamination (see Fig. 6). The first site is Colleambally Irrigation Area (hereafter called as "CIA") (145.50°E, 33.67°S), located in the southern part of New South Wales, Australia. The area is dominated by large, homogeneous paddy fields, cultivated from October one year to May the next, with significant temporal variations as the crop grows. Based on the statistics from the satellite images, about half of time series in CIA were cloud-free and only about one-fifth were heavily

TABLE I
IMAGE INFORMATION OF DIFFERENT DATASETS AT THE TWO SITES

Datasets	Spatio-temporal resolutions	Spectral bands	Available observations	Image size (m)
Landsat-8	30 m, 16-day	Blue, Green, Red, NIR, SWIR 1, SWIR 2	22 (CIA) 13 (RKZ)	15000 × 15000
Sentinel-2	10 m, 5-day	Blue, Green, Red, Red Edge 1, 2, 3, NIR, Red Edge 4, SWIR 1, SWIR 2	73 (CIA) 42 (RKZ)	
MOD09Q1	240 m, 8-day	Red, NIR	47 (CIA) 28 (RKZ)	

contaminated with cloud coverage over 80%. The second site is near Rikaze city (hereafter called as "RKZ") (89.05°E, 29.20°N), located in southern Qinghai-Tibetan plateau, China. It presents a heterogeneous mix of natural and agricultural landscapes, such as alpine shrub meadow, highland barley (grown from March to August) and rapeseed (grown from August of the current year to May of the following year). RKZ had worse observational conditions than CIA (see Fig. 6), only about 20% images in the time series were cloud-free due to the unique geomorphic and climatic features of the plateau, posing significant challenges for data reconstruction.

B. Datasets

To validate the effectiveness and robustness of NB-Re^3 across sensors with different spatiotemporal resolutions, we collected Landsat-8 and Sentinel-2 SR product from the GEE platform, respectively. All bands in Sentinel-2 were resampled to 10 m. The specific details of these images are summarized in Table I. In addition, MODIS SR product (MODIS Terra MOD09Q1 collection 6) collected from the GEE platform was used as the input of GF-SG. Since the time intervals of the generated NDVI time series do not coincide with those of the reflectance time series, we applied temporal linear interpolation to the reconstructed NDVI time series to make them coincidence.

C. Performance Comparison of Benchmark Methods

The proposed NB-Re^3 was compared with four typical approaches, MNSPI [16], HANTS [61], STAIR [31], and U-TILISE [46], as representatives of "Filling," "Fitting," "Fusing," and "DL-based" strategies, respectively. MNSPI is a typical filling approach based on the assumption that similar pixels in the spatial neighborhood have similar temporal variation patterns and predicts the missing data using weighted similar pixel values. HANTS is an improved fitting method based on the fast Fourier transform. It decomposes the time series into sine and cosine curves of different phases, frequencies, and amplitudes. These curves are then selectively integrated to remove cloud contamination. STAIR is an automatic fusing method based on the assumption that SR changes linearly in the short term, and it corrects the temporal change according to different land covers types. U-TILISE is a state-of-the-art DL-based approach that employs U-Net architecture [83] with a convolutional spatial encoder and an attention-based temporal encoder to capture the spatial and temporal dependencies, and a convolutional decoder to decode the high-level features back into the reconstructed time series. Detailed descriptions of these four methods are available in their original studies.

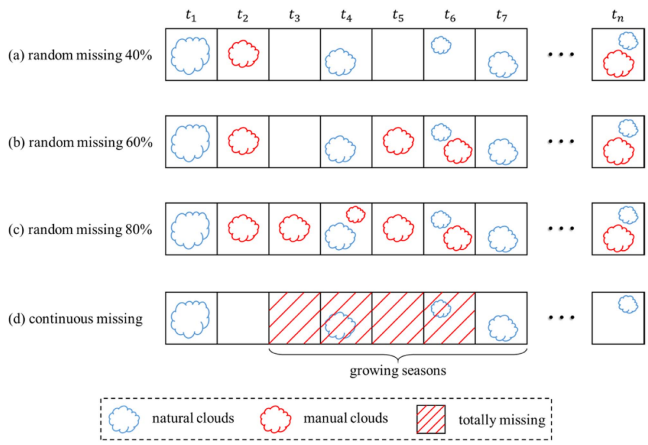

Fig. 7. Manual cloud simulation schematic for our experiments.

D. Experiments

Experiment I was designed as the reconstruction of random missing reflectance time series. To evaluate the performance of the models in scenarios with different levels of random missing data (the percentage of cloud pixels in the time-series images equal to 40%, 60%, and 80%, respectively), we performed cloud simulation process. Specifically, the missing percentage in the original time-series images was first calculated.

$$\mathrm{MP} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{X}\sum_{j=1}^{Y} M_{i,j}^{t}}{T*X*Y} \times 100\% \tag{11}$$

where MP is the ratio of the number of missing pixels to the total number of pixels. $M_{i,j}^{t}$ represents the binary cloud-shadow flag at the spatial location (i, j) and time step t (with the cloud-free observations labeled as value 0 and the missing observations as value 1). T is the length of time series. X and Y are the size of one single image. If the original MP was less than our scenarios (40%, 60%, and 80%), cloud pixels were randomly added to each image by Perlin Noise and Fractal Brownian Motion approach [84], [85] until MP, including the original missing pixels and the simulated missing pixels, reached the scenarios [see Fig. 7(a)–(c)]. The simulated cloud pixels serve two purposes, first as missing pixels to simulate our scenarios in the training process, and second as ground truth for performance validation.

The continuous missing cases are common in heavily clouded areas and also pose a challenge to the reconstruction process. Therefore, we designed Experiment II on both Landsat-8 and Sentinel-2 datasets that specifically simulated a scenario of continuous missing data during the growing season. These continuous missing gaps were simulated by setting all images to be missing during a certain period, such as the growing seasons (for CIA, from January–February and April–May in 2020, and for RKZ, from April–May and July–August in 2022) [see Fig. 7(d)], while keeping the original missing pixels for the remaining images.

To evaluate the effects of ancillary NDVI and improved network blocks on the reflectance time-series reconstruction, we conducted an ablation study (Experiment III) in four scenarios. In the first scenario, the NB-Re^3 was set as the baseline. The

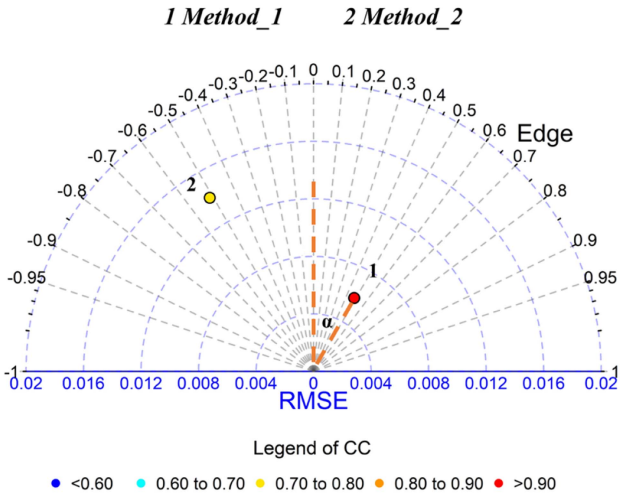

Fig. 8. Diagram illustrates the method comparison using the APA diagram. Each point represents a comparison method; the RMSE value is represented by the radial distance; the CC value is indicated by the color of the points; and the Edge index is proportional to the azimuth angle α that ranges from 180° to 0°. A point closer to the center of the circle with a redder color indicates better performance of the method.

second scenario, denoted as "baseline-no-TCN," removed the NDVI high-level features extraction block (TCN) after input NDVI time series. In the third scenario, "baseline-no-NDVI," we only input reflectance time series and considered multispectral correlations, excluding NDVI data. The final scenario was denoted as "baseline-no-regression," in which the linear regression block across multispectral bands was removed. Specifically, we selected one scenario from Experiment I (random missing 60%).

D. Evaluation Indices

Two typical spectral accuracy indices and one spatial accuracy index were used to evaluate the performance of different methods using the simulated clouds but actual cloud-free pixels. The first spectral index was the root-mean-square-error (RMSE) between the predicted reflectance value and the reference reflectance value for each simulated cloudy pixel in the time series, which shows the average prediction errors of the contaminated pixels. The second spectral index was the linear correlation coefficient (CC) between the predicted reflectance time series and the reference reflectance time series, which indicates the trend similarity between predicted and the reference reflectance time series. The spatial index was a normalized edge index (Edge), which is calculated by Robert's edge and defined as follows:

$$E_{(i,j)} = \left|D_{(i,j)} - D_{(i+1,j+1)}\right| + \left|D_{(i,j+1)} - D_{(i+1,j)}\right|$$
$$\text{Edge}_{(i,j)} = \frac{E_{\text{predicted}} - E_{\text{reference}}}{E_{\text{predicted}} + E_{\text{reference}}} \quad (12)$$

where $E_{(i,j)}$ represents the Robert's edge of the pixel at the ith row and jth column, and $D_{(i,j)}$ represents the corresponding reflectance value. This index mainly describes the edge information (e.g., boundaries of land parcels). Negative values indicate oversmoothed edges, while positive values indicate oversharpened edges.

In addition, an all-round performance assessment (APA) diagram (see Fig. 8) was employed [86] for a better visual cross comparison among different methods, with various evaluation indices into a single Taylor diagram [87].

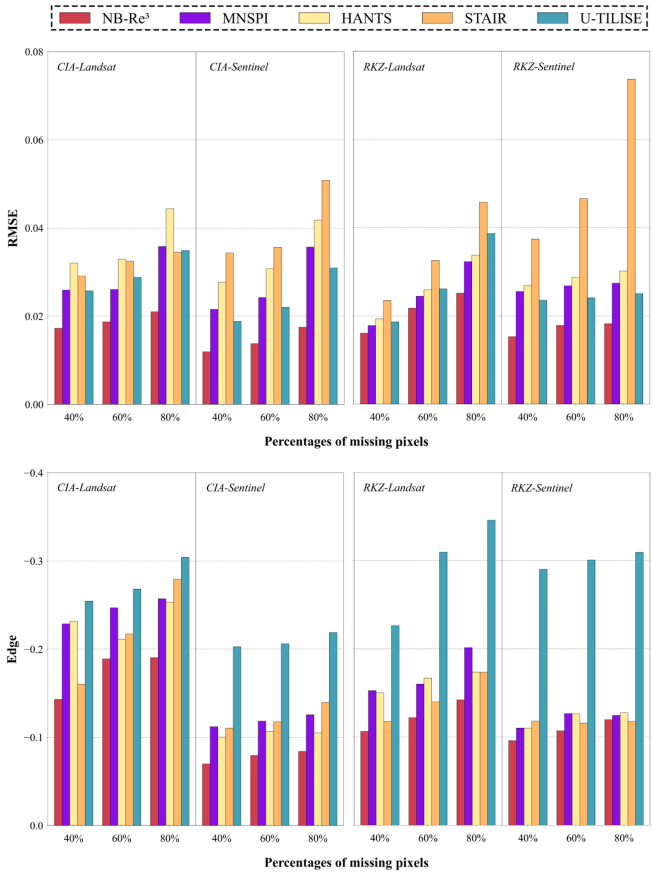

Fig. 9. Quantitative accuracies from both spectral and spatial perspectives (RMSE and Edge) on the reconstructed images by different methods at both sites for different random missing scenarios. RMSE and Edge values are the mean values of all simulated cloudy pixels.

IV. Results

A. Reconstruction of Random Missing Scenarios

Fig. 9 shows the quantitative assessments in both spectral and spatial dimensions. Despite a general trend of decreasing overall accuracy with increasing cloud coverage, the NB-Re^3 performs most robustly and excellently in different random missing scenarios (40%/60%/80%) and datasets (Landsat-8 and Sentinel-2) at two different challenge sites (CIA and RKZ), both in terms of RMSE and Edge metrics. As another DL-based approach, U-TILISE achieves the averaged second-best score on the RMSE metric, but the worst score on the Edge metric. The performance of MNSPI, HANTS, and STAIR are less satisfactory than the DL-based methods, especially STAIR, due to their inability to use high-quality NDVI data and inherent flaws in the design of the methods. More importantly, all methods generally perform better on Sentinel datasets than on Landsat datasets, mainly due

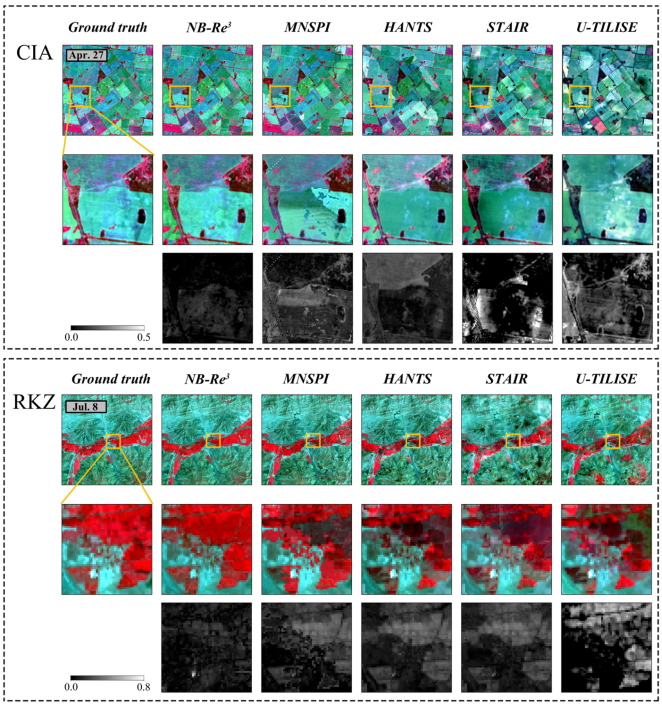

Fig. 10. Spatial patterns of the reconstructed Landsat-8 images by different comparison methods for 60% random missing scenario at two sites. The second row presents the subimages zoomed from the first row. All images are displayed in false-color composites (NIR, red, and green as RGB). The third row shows the gray-scaled cumulative absolute error for reflectance bands in the subimages.

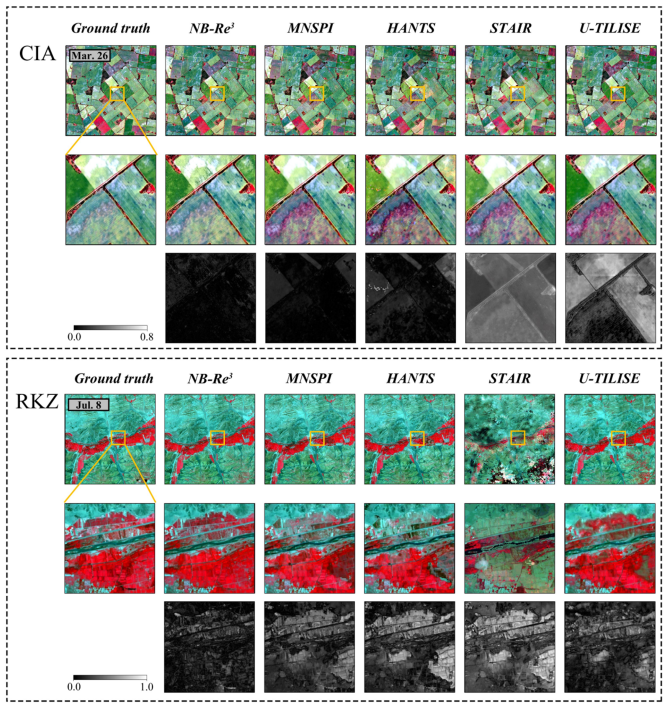

Fig. 11. Spatial patterns of the reconstructed Sentinel-2 images by different comparison methods for 60% random missing scenario at two sites. The second row presents the subimages zoomed from the first row. All images are displayed in false-color composites (NIR, red, and green as RGB). The third row shows the gray-scaled cumulative absolute error for reflectance bands in the subimages.

to the fact that Sentinel datasets has a higher spatiotemporal resolution, which increases the input information.

Fig. 10 shows the spatial patterns by different methods at two testing sites, using a set of Landsat-8 images reconstructed under the scenarios with 60% random missing images. Compared to other benchmarks, NB-Re^3 accurately captures spatial details, avoiding the "the salt and pepper noise" or "blocking effect." This superiority is attributed to the continuous spatial detail of the NDVI time series reconstructed with GF-SG and a specifically designed framework based on TCN and Bi-LSTM. However, other methods without high-quality NDVI data show inferior visual performance. For MNSPI, the selection of cloud-free neighboring pixels based on spectral similarity can introduce unexpected small patches. For HANTS, the fitting method struggles to account for all possible combinations of temporal trend patterns, resulting in oversmoothed spatial patterns at the homogeneous CIA site and degraded spatial consistency at the heterogeneous RKZ site. Although STAIR uses additional temporal information from MODIS, its low spatial resolution poses a challenge for spectral unmixing in highly heterogeneous regions, resulting in spatial inconsistencies. In contrast to NB-Re^3, U-TILISE does not use NDVI as an auxiliary guide for model training, and thus, exhibits significant discontinuity in local spatial patterns. In summary, the images reconstructed by the benchmark method show spatial discontinuities or smooth blurring of spatial details in the visual representation.

Fig. 11 illustrates examples of the spatial patterns of a reconstructed Sentinel-2 image at two testing sites, employing different methods under a scenario with 60% random missing images. Similar to the visual performance on Landsat-8 datasets, NB-Re^3 demonstrates continuous spatial details. It is noteworthy that as the spatial resolution increases from 30 to 10 m, Sentinel-2 images reveal more spatially local details than Landsat-8. This enhancement in ground complexity contributes to unstable reconstruction accuracy for traditional benchmarks such as MNSPI, HANTS, and STAIR. Among them, STAIR performs worst, because the increased coarse-fine spatial resolution ratio leads to extremely instability in pixels unmixing and residual distribution. Likewise, U-TILISE exhibits spatial discontinuity and loss of spatial details at the two testing sites without NDVI for guiding model parameters training.

B. Reconstruction of Continuous Missing Scenario

The averaged accuracies at two testing sites for simulated continuous missing scenarios are further quantified and shown in APA diagrams, respectively (see Fig. 12). Similar to the performance of random missing scenarios in Experiment I, the reflectance images reconstructed by NB-Re^3 outperform those of the other four methods in band-averaged accuracy. Especially in the heterogeneous region such as RKZ, the denser time series available in the Sentinel-2 datasets contributes to slightly higher accuracy for all methods on the Sentinel-2 datasets compared to the Landsat-8 dataset. Compared with NB-Re^3, the performance of other benchmarks is more sensitive to long-term missing gaps and varying degrees of landscape heterogeneity, ultimately limiting their reconstruction accuracy.

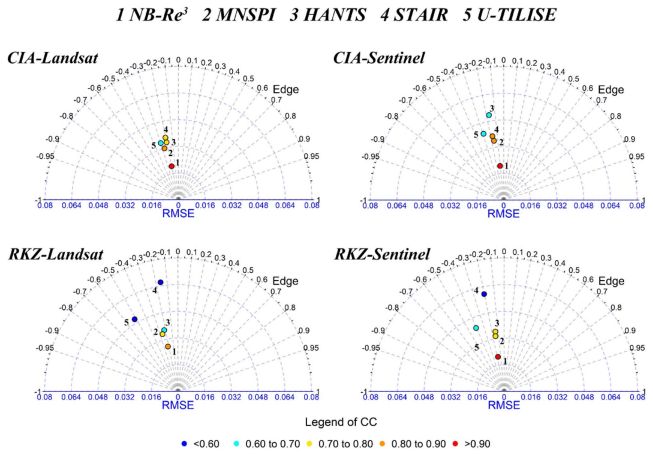

Fig. 12. APA diagrams displaying the performance of the five comparison methods for the continuous missing scenarios in different datasets at two sites. Results were averaged over all testing dates.

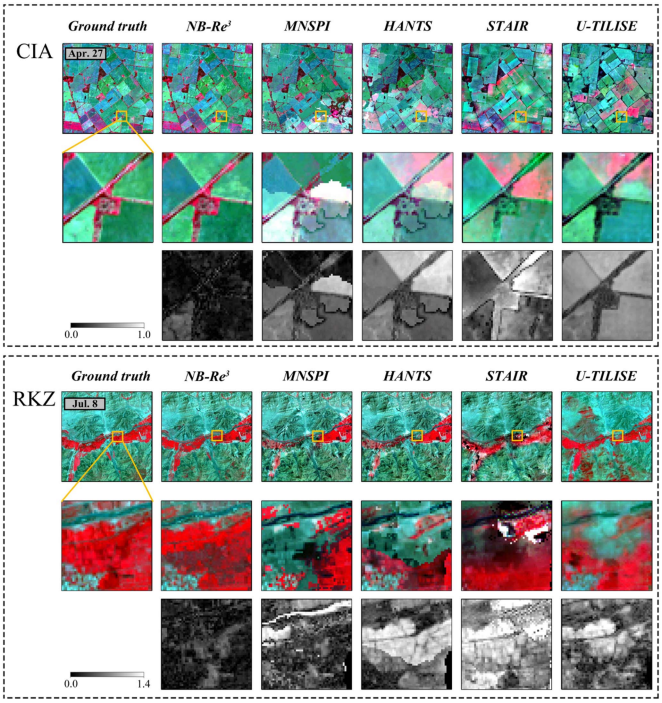

Fig. 13. Spatial patterns of the reconstructed Landsat-8 images by different comparison methods for continuous missing scenario at two sites. The second row presents the subimages zoomed from the first row. All images are displayed in false-color composites (NIR, red, and green as RGB). The third row shows the gray-scaled cumulative absolute error for reflectance bands in the subimages.

In the challenge scenario with continuous missing of time-series images, the NB-Re^3 consistently demonstrates superior visual performance, consistent with the quantitative performance (see Figs. 13 and 14). Similar to the Experiment I, images reconstructed by other benchmarks exhibit discontinuous spatial details and show great disadvantages. Long-term missing pattern is the main reason for the lower quality visual performance, especially in the region with fragment and heterogeneous landscape. For MNSPI, the limited observations lead to the increasing difficulty in finding spectrally similar neighboring pixels; for

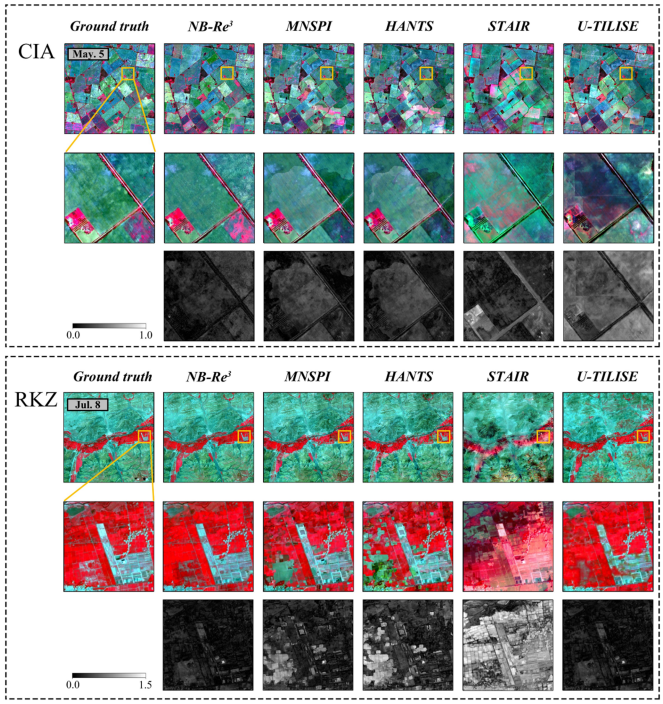

Fig. 14. Spatial patterns of the reconstructed Sentinel-2 images by different comparison methods for continuous missing scenario at two sites. The second row presents the subimages zoomed from the first row. All images are displayed in false-color composites (NIR, red, and green as RGB). The third row shows the gray-scaled cumulative absolute error for reflectance bands in the subimages.

TABLE II
ABLATION STUDY OF PROPOSED MODEL

Scenarios	Operations			Quantified accuracies		
	Regression	NDVI	TCN	RMSE	CC	Edge
baseline	+	+	+	**0.0187**	**0.9600**	**-0.1889**
				0.0218	**0.9120**	**-0.1819**
baseline-no-TCN	+	+	-	0.0188	0.9590	-0.2188
				0.0231	0.8728	-0.2120
baseline-no-NDVI	+	-	-	0.0202	0.9535	-0.2359
				0.0233	0.8697	-0.2482
baseline-no-Regression	-	+	+	0.0197	0.9558	-0.1983
				0.0222	0.8853	-0.2234

Values above represent CIA site, values below represent RKZ site. Values in bold indicate the best performance.

HANTS, it leads to the failure to capture the true temporal trends of the time series using only the early and late observations. The long missing observations also lead to the problem of linear interpolation for each fusion date in STAIR, and to the difficulty of modeling the temporal dependence of the reflectance time series in U-TILISE. Overall, benchmark methods present poorer performances with continuous missing scenario compared to those with random missing scenarios.

C. Assessment of Ablation Study

In the ablation study, "baseline" achieves the best reconstruction accuracy, while "baseline-no-NDVI" has the worst performance at both sites, in terms of both spectral and spatial accuracy (see Table II). This emphasizes the critical role of high-quality NDVI as an auxiliary input that provides valuable

insights into vegetation dynamics, especially when corresponding reflectance observations are sparse. Compared to "baseline-no-NDV," adding NDVI, and then, using either the regression module or the TCN module can improve the performance of the model to some extent, with the regression module outperforming the TCN module in homogeneous regions and the TCN module being more effective in heterogeneous regions. The best results can only be achieved by combining the TCN and regression modules to create a "baseline." That is, the regression and TCN modules have complementary strengths.

V. Discussion

A. Superiority of the NB-Re^3 Model

The proposed NB-Re^3 model aims to overcome the limited cloud-free observations by incorporating ancillary NDVI time series into the reflectance reconstruction. Compared to existing "Filling," "Fitting," and "Fusing" strategies, NB-Re^3 significantly outperformed these traditional methods in its ability to learn high-dimensional features and construct the NDVI-reflectance dynamic relationship and spectral autocorrelation between multiple spectral bands. The end-to-end framework also allows the network to learn the functional relationship directly from the data, without the assumptions made in traditional strategies. Moreover, a better NDVI auxiliary information and a suitable network structure could be the reason why NB-Re^3 outperformed other "DL-based" methods.

NB-Re^3 has two significant superiorities based on its design principle. First, it integrates high-quality NDVI time series to support the reconstruction of reflectance time series. This is necessary because relying solely on incomplete reflectance time series makes it difficult to accurately capture temporal trends, especially when the available cloud-free observations are sparse and unevenly distributed over the time series. High-quality NDVI time series provide additional information during cloudy periods when reflectance observations are not available. More importantly, high-quality NDVI time series are easier to reconstruct with the sophisticated methods. This is because the NDVI time series has an obvious seasonal periodicity and can thus be modeled more directly with higher reconstruction accuracy, regardless of whether single-sensor or multisensor methods are used [59], [64], [66]. Given the value of high-quality NDVI time series, the TCN module is applied to capture the deep temporal features of NDVI at different time scales. Then, the Bi-LSTM is used to construct the dynamic relationship between NDVI and reflectance time series. Overall, NB-Re^3 focuses on learning the temporal dependence within reflectance time series and the robust relationships between the reflectance and high-quality NDVI time series, thereby enhancing the guidance for reflectance reconstruction. The results of Experiment III highlight the critical role of NDVI as an auxiliary input that, together with the TCN module, provides valuable insights into vegetation dynamics.

Second, NB-Re^3 takes full advantage of the spectral autocorrelation of multiple spectral bands. It is very often the case that spectral values in multiple bands change simultaneously, and in a certain pattern, rather than independently. Therefore, using multiple bands in a regression block rather than a single band can provide more information, leading to more accurate prediction. Another major benefit of using multiple bands is that the method is less sensitive to the quality of the input data [26]. With Bi-LSTM, we obtain reflectance predictions for each band, but the quality of the reflectance predicted by Bi-LSTM is different for each spectral band, i.e., differences in prediction accuracy. This is because the correlation between NDVI and reflectance is not the same for each band, and the original reflectance is affected differently by noise (e.g., haze and thin clouds) across bands (typically the longer wavelength shortwave infrared bands are less affected). By building the multiple regression model, reflectance predictions by Bi-LSTM for higher accuracy bands can be used to improve reflectance predictions for lower accuracy bands. To demonstrate this advantage more clearly, we repeated the randomly missing 60% scenario in Experiment I using multiple bands (referred to "NB-Re^3_multi") versus each single band (referred to "NB-Re^3_single") as input respectively (see Fig. 15). It is evident that using multiple bands can capture more complex temporal change patterns and spatial details than using each single band.

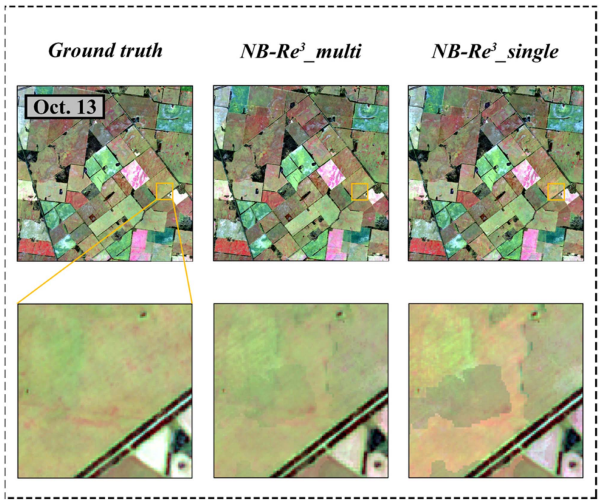

Fig. 15. Spatial patterns of the reconstructed Sentinel-2 images at CIA site in "NB-Re^3_multi" and "NB-Re^3_single" scenario. Images below each image are the zoom-in subimages. All images are false-color images (NIR, red, and green as RGB).

B. Robustness of NB-Re^3 Against False Cloud Masks

Automatic cloud and cloud shadow detection algorithms offer an effective approach for generating cloud mask data. However, these methods still suffer from omission and commission errors, where some clouds may be missed or certain bright land surface objects may be misidentified as clouds. These unavoidable errors may further affect the performance of NB-Re^3. To evaluate the sensitivity of NB-Re^3 to cloud mask errors in real-world applications, we performed sensitivity experiments using the Sentinel-2 dataset from the CIA site, with the same parameter settings as in Experiment I, but with the cloud mask that contains omission and commission errors. The results show that the NB-Re^3 reconstructed images show very similar spatial patterns to the ground

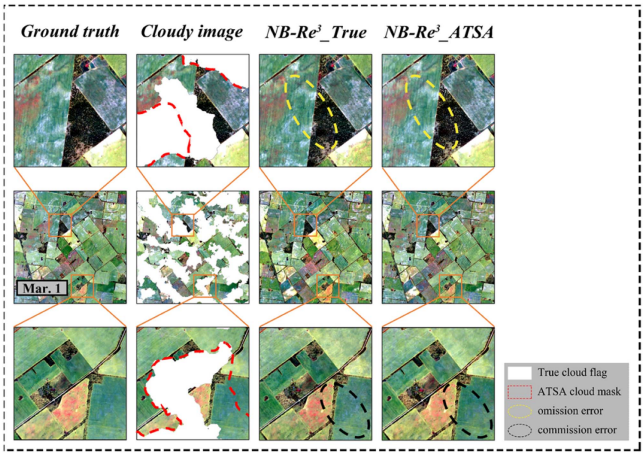

Fig. 16. Spatial patterns of the reconstructed Sentinel-2 images at CIA site using different mask flag data as input ("NB-Re^3_True" and "NB-Re^3_ ATSA"). White polygons represent the accurate true cloud flag, red dotted polygons represent the generated cloud mask by the ASTA algorithm, omission error, and commission error are marked, respectively. All images are false-color images (NIR, red, and green as RGB).

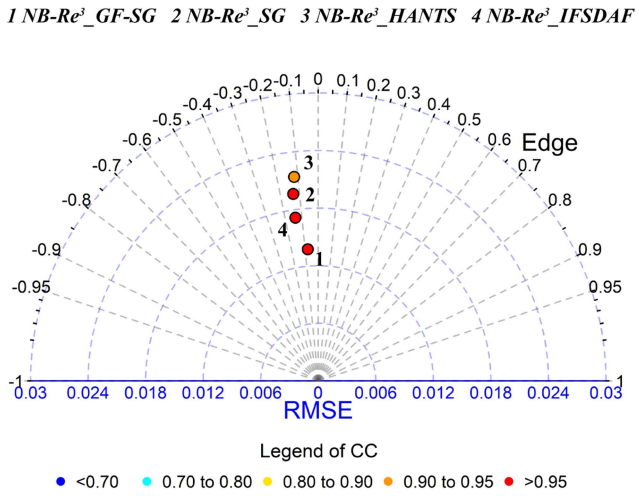

Fig. 17. APA diagrams displaying the performance of the NB-Re^3 using different NDVI time series as input for the random missing 60% images in Sentine-2 datasets at CIA site.

truth, whether using the manually generated accurate cloud masks (called "NB-Re^3_True") or the error-prone cloud masks generated by the ASTA algorithm (called "NB-Re^3_ATSA") (see Fig. 16). This suggests that NB-Re^3 can mitigate the impact of inaccurate cloud masks on prediction accuracy to a greater extent.

C. Impact of NDVI Reconstruction Quality on Reconstruction

Although high-quality NDVI time series are easier to reconstruct because of their apparent seasonal periodicity, the accuracy of NDVI time series reconstructed by different methods varies in different scenarios [88], especially in areas of heavy cloud coverage. To investigate the sensitivity of NB-Re^3 to NDVI time-series reconstruction quality, we repeated Experiment I using the CIA site Sentinel-2 dataset, where NB-Re^3 input NDVI time series reconstructed by different methods, including the four most typical NDVI reconstruction methods: the GF-SG [66], SG filter [63], HANTS [61], and IFSDAF [89]. The results show that the higher the quality of the reconstructed NDVI time series, the higher the accuracy of NB-Re^3 (see Fig. 17). This not only confirms that choosing GF-SG in NB-Re^3 is a better option, but also suggests that any method to improve the quality of NDVI reconstruction will be beneficial in improving NB-Re^3 performance.

D. Limitation and Future Work

It is important to recognize that NB-Re^3 still has its limitations; first, while high-quality NDVI time series can be very helpful in reconstructing reflectance, their effectiveness may be degraded in nonvegetated areas with minimal seasonal variation, such as bare soil and impervious surfaces. Second, some new DL architectures, such as the Transformers [90], have emerged and achieved great success in processing sequential data. These newly developed models have substantial potential to replace the role of Bi-LSTM, although they require more training samples and computational resources. We call for greater efforts to explore this potential.

VI. Conclusion

This study proposed a novel high-quality reflectance time-series reconstruction approach (NB-Re^3) taking full advantage of NDVI time series and spectral autocorrelation of multiple spectral bands. Given high-quality NDVI time series are of great value and easier to reconstruct, the TCN module is used to capture the deep temporal features of NDVI time series at different time scales. Then, a Bi-LSTM is used to construct the dynamic relationship between NDVI and reflectance time series. In addition, NB-Re^3 fully exploits the spectral autocorrelation of multiple spectral bands, which not only provides more information, but makes NB-Re^3 less sensitive to Bi-LSTM's predicted reflectance quality of each spectral band. The experiments showed that this novel NB-Re^3 model outperformed other four existing state-of-the-art methods (MNSPI, HANTS, STAIR, and U-TILISE) under different missing observation scenarios. In conclusion, NB-Re^3 has great potential to generate cloud-free reflectance time-series products for user-specific applications.

References

[1] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL, USA: CRC Press, 2019.
[2] M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, May 2012, doi: 10.1016/j.rse.2011.11.026.
[3] R. J. Frazier, N. C. Coops, and M. A. Wulder, "Boreal shield forest disturbance and recovery trends using Landsat time series," *Remote Sens. Environ.*, vol. 170, pp. 317–327, Dec. 2015, doi: 10.1016/j.rse.2015.09.015.
[4] M. A. Friedl et al., "Global land cover mapping from MODIS: Algorithms and early results," *Remote Sens. Environ.*, vol. 83, no. 1, pp. 287–302, Nov. 2002, doi: 10.1016/S0034-4257(02)00078-0.
[5] D. P. Roy et al., "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sens. Environ.*, vol. 145, pp. 154–172, Apr. 2014, doi: 10.1016/j.rse.2014.02.001.
[6] C. E. Woodcock, T. R. Loveland, M. Herold, and M. E. Bauer, "Transitioning from change detection to monitoring with remote sensing: A paradigm shift," *Remote Sens. Environ.*, vol. 238, Mar. 2020, Art. no. 111558, doi: 10.1016/j.rse.2019.111558.

[7] M. A. Wulder, J. C. White, M. D. Gillis, N. Walsworth, M. C. Hansen, and P. Potapov, "Multiscale satellite and spatial information and analysis framework in support of a large-area forest monitoring and inventory update," *Environ. Monit. Assessment*, vol. 170, no. 1, pp. 417–433, Nov. 2010, doi: 10.1007/s10661-009-1243-8.

[8] M. L. Imhoff, P. Zhang, R. E. Wolfe, and L. Bounoua, "Remote sensing of the urban heat island effect across biomes in the continental USA," *Remote Sens. Environ.*, vol. 114, no. 3, pp. 504–513, Mar. 2010, doi: 10.1016/j.rse.2009.10.008.

[9] V. Kovalskyy and D. P. Roy, "The global availability of Landsat 5 TM and Landsat 7 ETM + land surface observations and implications for global 30 m Landsat data product generation," *Remote Sens. Environ.*, vol. 130, pp. 280–293, Mar. 2013, doi: 10.1016/j.rse.2012.12.003.

[10] L. A. Santos, K. R. Ferreira, G. Camara, M. C. A. Picoli, and R. E. Simoes, "Quality control and class noise reduction of satellite image time series," *ISPRS J. Photogrammetry Remote Sens.*, vol. 177, pp. 75–88, Jul. 2021, doi: 10.1016/j.isprsjprs.2021.04.014.

[11] M. Sudmanns, D. Tiede, H. Augustin, and S. Lang, "Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass," *Int. J. Digit. Earth*, vol. 13, no. 7, pp. 768–784, Jul. 2020, doi: 10.1080/17538947.2019.1572799.

[12] H. Zhang, W. Liu, and L. Zhang, "Seamless and automated rapeseed mapping for large cloudy regions using time-series optical satellite imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 45–62, Feb. 2022, doi: 10.1016/j.isprsjprs.2021.12.001.

[13] 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.

[14] Planet Team, "Planet application program interface: In space for life on Earth," Planet, San Francisco, CA, USA, 2017. [Online]. Available: https://api.planet.com/

[15] J. Chen, X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin, "A simple and effective method for filling gaps in Landsat ETM+ SLC-off images," *Remote Sens. Environ.*, vol. 115, no. 4, pp. 1053–1064, Apr. 2011, doi: 10.1016/j.rse.2010.12.010.

[16] X. Zhu, F. Gao, D. Liu, and J. Chen, "A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 521–525, May 2012, doi: 10.1109/LGRS.2011.2173290.

[17] L. Yan and D. P. Roy, "Large-area gap filling of Landsat reflectance time series by spectral-angle-mapper based spatio-temporal similarity (SAMSTS)," *Remote Sens.*, vol. 10, no. 4, Apr. 2018, Art. no. 4, doi: 10.3390/rs10040609.

[18] Z. Zhu, C. E. Woodcock, C. Holden, and Z. Yang, "Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time," *Remote Sens. Environ.*, vol. 162, pp. 67–83, Jun. 2015, doi: 10.1016/j.rse.2015.02.009.

[19] F. Vuolo, W.-T. Ng, and C. Atzberger, "Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 57, pp. 202–213, May 2017, doi: 10.1016/j.jag.2016.12.012.

[20] L. Yan and D. P. Roy, "Spatially and temporally complete Landsat reflectance time series modelling: The fill-and-fit approach," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111718, doi: 10.1016/j.rse.2020.111718.

[21] X. Zhu, F. Cai, J. Tian, and T. K.-A. Williams, "Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions," *Remote Sens.*, vol. 10, no. 4, Apr. 2018, Art. no. 4, doi: 10.3390/rs10040527.

[22] X. Zhu, J. Chen, F. Gao, X. 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.

[23] 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.

[24] M. Wu, Z. Niu, C. Wang, C. Wu, and L. Wang, "Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model," *J. Appl. Remote Sens.*, vol. 6, Mar. 2012, Art. no. 063507, doi: 10.1117/1.JRS.6.063507.

[25] Q. 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.

[26] S. Liu, J. Zhou, Y. Qiu, J. Chen, X. Zhu, and H. Chen, "The FIRST model: Spatiotemporal fusion incorporating spectral autocorrelation," *Remote Sens. Environ.*, vol. 279, Sep. 2022, Art. no. 113111, doi: 10.1016/j.rse.2022.113111.

[27] B. Huang and 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.

[28] X. Liu, C. Deng, S. Wang, G.-B. Huang, B. Zhao, and P. Lauren, "Fast and accurate spatiotemporal fusion based upon extreme learning machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 2039–2043, Dec. 2016, doi: 10.1109/LGRS.2016.2622726.

[29] 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.

[30] Z. Gu, J. Chen, Y. Chen, Y. Qiu, X. Zhu, and X. Chen, "Agri-Fuse: A novel spatiotemporal fusion method designed for agricultural scenarios with diverse phenological changes," *Remote Sens. Environ.*, vol. 299, Dec. 2023, Art. no. 113874, doi: 10.1016/j.rse.2023.113874.

[31] Y. Luo, K. Guan, and J. Peng, "STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product," *Remote Sens. Environ.*, vol. 214, pp. 87–99, Sep. 2018, doi: 10.1016/j.rse.2018.04.042.

[32] J. Gao, Q. Yuan, J. Li, and X. Su, "Unsupervised missing information reconstruction for single remote sensing image with deep code regression," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102599, doi: 10.1016/j.jag.2021.102599.

[33] B. Jiang et al., "A deep-learning reconstruction method for remote sensing images with large thick cloud cover," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 115, Dec. 2022, Art. no. 103079, doi: 10.1016/j.jag.2022.103079.

[34] S. Wang, D. Cui, L. Wang, and J. Peng, "Applying deep-learning enhanced fusion methods for improved NDVI reconstruction and long-term vegetation cover study: A case of the Danjiang River Basin," *Ecological Indicators*, vol. 155, Nov. 2023, Art. no. 111088, doi: 10.1016/j.ecolind.2023.111088.

[35] J. Li et al., "Deep learning in multimodal remote sensing data fusion: A comprehensive review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102926, doi: 10.1016/j.jag.2022.102926.

[36] A. Jayakrishnan, M. Venkatesan, P. Prabhavathy, and M. Alkha, "MSDF-Net: A multi-scale deep fusion network with dilated convolutions for cloud removal from Sentinel-2 imagery," in *Proc. Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2023, pp. 63–70, doi: 10.1109/APSIPAASC58517.2023.10317471.

[37] A. Sebastianelli et al., "PLFM: Pixel-level merging of intermediate feature maps by disentangling and fusing spatial and temporal data for cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5412216, doi: 10.1109/TGRS.2022.3208694.

[38] F. Xu, Y. Shi, P. Ebel, W. Yang, and X. X. Zhu, "Multimodal and multiresolution data fusion for high-resolution cloud removal: A novel baseline and benchmark," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Dec. 2024, Art. no. 5600615, doi: 10.1109/TGRS.2023.3337845.

[39] H. Li, P. Ghamisi, U. Soergel, and X. X. Zhu, "Hyperspectral and LiDAR fusion using deep three-stream convolutional neural networks," *Remote Sens.*, vol. 10, no. 10, Oct. 2018, Art. no. 10, doi: 10.3390/rs10101649.

[40] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018, doi: 10.1109/JSTARS.2018.2805923.

[41] F. Xu et al., "GLF-CR: SAR-enhanced cloud removal with global–local fusion," *ISPRS J. Photogrammetry Remote Sens.*, vol. 192, pp. 268–278, Oct. 2022, doi: 10.1016/j.isprsjprs.2022.08.002.

[42] H. Liu, B. Huang, and J. Cai, "Thick cloud removal under land cover changes using multisource satellite imagery and a spatiotemporal attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jan. 2023, Art. no. 5601218, doi: 10.1109/TGRS.2023.3236106.

[43] Z. Qiu, H. Shen, L. Yue, and G. Zheng, "Cross-sensor remote sensing imagery super-resolution via an edge-guided attention-based network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 199, pp. 226–241, May 2023, doi: 10.1016/j.isprsjprs.2023.04.016.

[44] K. Enomoto et al., "Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1533–1541, doi: 10.1109/CVPRW.2017.197.

[45] J. Gao, Q. Yuan, J. Li, H. Zhang, and X. Su, "Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks," *Remote Sens.*, vol. 12, no. 1, Jan. 2020, Art. no. 1, doi: 10.3390/rs12010191.

[46] C. Stucker, V. S. F. Garnot, and K. Schindler, "U-TILISE: A sequence-to-sequence model for cloud removal in optical satellite time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Nov. 2023, Art. no. 5408716, doi: 10.1109/TGRS.2023.3333391.

[47] M. Zhao, P. Olsen, and R. Chandra, "Seeing through clouds in satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jan. 2023, Art. no. 4704616, doi: 10.1109/TGRS.2023.3239592.

[48] W. He and N. Yokoya, "Multi-temporal Sentinel-1 and -2 data fusion for optical image simulation," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 10, Oct. 2018, Art. no. 10, doi: 10.3390/ijgi7100389.

[49] C. Li, X. Liu, and S. Li, "Transformer meets GAN: Cloud-free multispectral image reconstruction via multisensor data fusion in satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Nov. 2023, Art. no. 5408513, doi: 10.1109/TGRS.2023.3326545.

[50] A. S. Namboodiri, R.K. Sanodiya, and P. Arun, "Remote sensing cloud removal using a combination of spatial attention and edge detection," in *Proc. 11th Int. Symp. Electron. Syst. Devices Comput.*, 2023, pp. 1–6, doi: 10.1109/ESDC56251.2023.10149875.

[51] I. Dumeur, S. Valero, and J. Inglada, "Self-supervised spatio-temporal representation learning of satellite image time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 4350–4367, Jan. 2024, doi: 10.1109/JSTARS.2024.3358066.

[52] P. Ebel, M. Schmitt, and X. X. Zhu, "Internal learning for sequence-to-sequence cloud removal via synthetic aperture radar prior information," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2691–2694, doi: 10.1109/IGARSS47720.2021.9554268.

[53] T. Peng, M. Liu, X. Liu, Q. Zhang, L. Wu, and X. Zou, "Reconstruction of optical image time series with unequal lengths SAR based on improved sequence–Sequence model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5632017, doi: 10.1109/TGRS.2022.3208926.

[54] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, "A tutorial on speckle reduction in synthetic aperture radar images," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 3, pp. 6–35, Sep. 2013, doi: 10.1109/MGRS.2013.2277512.

[55] A. de la I. Martinez and S. M. Labib, "Demystifying normalized difference vegetation index (NDVI) for greenness exposure assessments and policy interventions in urban greening," *Environ. Res.*, vol. 220, Mar. 2023, Art. no. 115155, doi: 10.1016/j.envres.2022.115155.

[56] J. Drisya, K. D. S., and T. Roshni, "Spatiotemporal variability of soil moisture and drought estimation using a distributed hydrological model," in *Integrating Disaster Science and Management*, P. Samui, D. Kim, and C. Ghosh, Eds. Amsterdam, The Netherlands: Elsevier, 2018, pp. 451–460, doi: 10.1016/B978-0-12-812056-9.00027-0.

[57] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, May 1979, doi: 10.1016/0034-4257(79)90013-0.

[58] G. J. Roerink, M. Menenti, and W. Verhoef, "Reconstructing cloudfree NDVI composites using Fourier analysis of time series," *Int. J. Remote Sens.*, vol. 21, no. 9, pp. 1911–1917, Jan. 2000, doi: 10.1080/014311600209814.

[59] D. P. Roy and L. Yan, "Robust Landsat-based crop time series modelling," *Remote Sens. Environ.*, vol. 238, Mar. 2020, Art. no. 110810, doi: 10.1016/j.rse.2018.06.038.

[60] B. T. Wilson, J. F. Knight, and R. E. McRoberts, "Harmonic regression of Landsat time series for modeling attributes from National Forest Inventory Data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 137, pp. 29–46, Mar. 2018, doi: 10.1016/j.isprsjprs.2018.01.006.

[61] J. Zhou, L. Jia, and M. Menenti, "Reconstruction of global MODIS NDVI time series: Performance of harmonic analysis of time series (HANTS)," *Remote Sens. Environ.*, vol. 163, pp. 217–228, Jun. 2015, doi: 10.1016/j.rse.2015.03.018.

[62] R. Cao et al., "A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter," *Remote Sens. Environ.*, vol. 217, pp. 244–257, Nov. 2018, doi: 10.1016/j.rse.2018.08.022.

[63] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter," *Remote Sens. Environ.*, vol. 91, no. 3, pp. 332–344, Jun. 2004, doi: 10.1016/j.rse.2004.03.014.

[64] Y. Qiu, J. Zhou, J. Chen, and X. Chen, "Spatiotemporal fusion method to simultaneously generate full-length normalized difference vegetation index time series (SSFIT)," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 100, Aug. 2021, Art. no. 102333, doi: 10.1016/j.jag.2021.102333.

[65] Y. Rao, X. Zhu, J. Chen, and J. Wang, "An improved method for producing high spatial-resolution NDVI time series datasets with multi-temporal MODIS NDVI data and landsat TM/ETM+ images," *Remote Sens.*, vol. 7, no. 6, Jun. 2015, Art. no. 6, doi: 10.3390/rs70607865.

[66] Y. Chen, R. Cao, J. Chen, L. Liu, and B. Matsushita, "A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky–Golay filter," *ISPRS J. Photogrammetry Remote Sens.*, vol. 180, pp. 174–190, Oct. 2021, doi: 10.1016/j.isprsjprs.2021.08.015.

[67] A. Garioud, S. Valero, S. Giordano, and C. Mallet, "Recurrent-based regression of Sentinel time series for continuous vegetation monitoring," *Remote Sens. Environ.*, vol. 263, Sep. 2021, Art. no. 112419, doi: 10.1016/j.rse.2021.112419.

[68] G. Scarpa, M. Gargiulo, A. Mazza, and R. Gaetano, "A CNN-based fusion method for feature extraction from sentinel data," *Remote Sens.*, vol. 10, no. 2, Feb. 2018, Art. no. 2, doi: 10.3390/rs10020236.

[69] W. Zhao, Y. Qu, J. Chen, and Z. Yuan, "Deeply synergistic optical and SAR time series for crop dynamic monitoring," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111952, doi: 10.1016/j.rse.2020.111952.

[70] X. Zhu and E. H. Helmer, "An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions," *Remote Sens. Environ.*, vol. 214, pp. 135–153, Sep. 2018, doi: 10.1016/j.rse.2018.05.024.

[71] S. Foga et al., "Cloud detection algorithm comparison and validation for operational Landsat data products," *Remote Sens. Environ.*, vol. 194, pp. 379–390, Jun. 2017, doi: 10.1016/j.rse.2017.03.026.

[72] K. Tarrio et al., "Comparison of cloud detection algorithms for Sentinel-2 imagery," *Sci. Remote Sens.*, vol. 2, Dec. 2020, Art. no. 100010, doi: 10.1016/j.srs.2020.100010.

[73] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images," *Remote Sens. Environ.*, vol. 159, pp. 269–277, Mar. 2015, doi: 10.1016/j.rse.2014.12.014.

[74] Y. Zhang, B. Guindon, and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images," *Remote Sens. Environ.*, vol. 82, no. 2, pp. 173–187, Oct. 2002, doi: 10.1016/S0034-4257(02)00034-2.

[75] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, Mar. 2012, doi: 10.1016/j.rse.2011.10.028.

[76] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," Apr. 19, 2018, *arXiv:1803.01271*, doi: 10.48550/arXiv.1803.01271.

[77] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[78] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1171–1179.

[79] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[80] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "BRITS: Bidirectional recurrent imputation for time series," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6775–6785.

[81] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent Neural networks for multivariate time series with missing values," *Sci. Rep.*, vol. 8, no. 1, Apr. 2018, Art. no. 6085, doi: 10.1038/s41598-018-24271-9.

[82] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Jan. 29, 2017, *arXiv.1412.6980*, doi: 10.48550/arXiv.1412.6980.

[83] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

[84] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, no. 4, pp. 422–437, Oct. 1968, doi: 10.1137/1010093.

[85] K. Perlin, "An image synthesizer," *ACM Siggraph Comput. Graph.*, vol. 19, no. 3, pp. 287–296, Jul. 1985, doi: 10.1145/325165.325247.

[86] X. 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.

[87] K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *J. Geophys. Res., Atmospheres*, vol. 106, no. D7, pp. 7183–7192, 2001, doi: 10.1029/2000JD900719.

[88] S. Li, L. Xu, Y. Jing, H. Yin, X. Li, and X. Guan, "High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102640, doi: 10.1016/j.jag.2021.102640.

[89] 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.
[90] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

Hongtao Shu received the B.S. degree in cartography and geographic information system from Sun Yat-sen University, Guangzhou, China, in 2016, and the M.S. degree in global climate change in 2019 from Beijing Normal University, Beijing, China, where he is currently working toward the Ph.D. degree in cartography and geographic information system with the Faculty of Geographical Science.

His research interests include remote sensing image processing and crop classification.

Zhuoning Gu received the B.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2021. She is currently working toward the M.S. degree in cartography and geographic information systems with Beijing Normal University, Beijing, China.

Her research interests include the agricultural application of remote sensing and related data reconstruction.

Yang Chen received the B.S. degree in environmental sciences and the M.S. degree in electronics and communications engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2016 and 2019, respectively, and the Ph.D. degree in geography from Beijing Normal University, Beijing, China, in 2023.

He is currently a postdoctoral with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include remote sensing image processing and crop monitoring.

Hui Chen received the B.E. degree in remote sensing science and technology and the M.E. degree in photogrammetry and remote sensing from the China University of Geoscience, Wuhan, China in 2016 and 2019, respectively. He is currently working toward the Ph.D. degree in geography with Beijing Normal University, Beijing, China.

His research interests include remote sensing image processing and crop classification.

Xuehong Chen received the B.S. degree in physics and the M.S. degree in civil engineering from Beijing Normal University, Beijing, China, in 2006 and 2009, respectively, and the Ph.D. degree in Earth and environmental science from Nagoya University, Nagoya, Japan, in 2012.

He is currently an Associate Professor with the Faculty of Geographical Science, Institute of Remote Sensing Science and Engineering, Beijing Normal University. His research interests include preprocessing, data fusion of satellite images, and land cover mapping by remote sensing.

Jin Chen received the B.A. and M.S. degrees in geography from Beijing Normal University, Beijing, China, in 1989 and 1992, respectively, and the Ph.D. degree in civil engineering from Kyushu University, Fukuoka, Japan, in 2000.

He was a Postdoctoral Researcher with the University of California at Berkeley, Berkeley, CA, USA, from 2000 to 2001, and also with the National Institute of Environmental Studies, Tsukuba, Japan, from 2001 to 2004. He is currently a Professor with the Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University. His research interests include remote sensing modeling and vegetation parameter retrieval through the inversion of remote sensing models.