

Incorporating Environmental Variables Into Spatiotemporal Fusion Model to Reconstruct High-Quality Vegetation Index Data

Xiangqian Li, Qiongyan Peng¹, Yi Zheng, Shangrong Lin, Bin He, Yuean Qiu, Jin Chen²,
Yang Chen, and Wenping Yuan³

Abstract—Restricted by the design of satellite sensors, the existing satellite-based normalized difference vegetation index (NDVI) cannot simultaneously have a high temporal resolution and spatial resolution, which substantially limits its applications. In recent years, several spatiotemporal fusion models have been developed to produce vegetation index datasets with both high spatial and temporal resolutions, but large uncertainties remain. This study proposes a spatiotemporal fusion model [i.e., Integrating ENvironmental VarIable spatiotemporal fusion (InENVI) model] based on a machine-learning method by incorporating environmental variables to reconstruct NDVI data. Over 14 study areas covering various vegetation types globally, the InENVI method was validated for reproducing spatiotemporal variations in NDVI. On average, the determining coefficients (R^2) of the reconstructed NDVI compared with satellite-based NDVI observations were above 0.90, reflecting the spatiotemporal variations over all study sites. In addition, we compared the performance of the InENVI model with seven other fusion models over two cropland areas with high vegetation heterogeneity. The results showed that the newly developed InENVI method had the best performance, and the reconstruction error of the InENVI method decreased about 23.68%–59.63% on average over two study areas compared to the other seven methods. Our analyses also highlighted that the integration of environmental variables into spatiotemporal fusion is necessary to improve reconstruction accuracy. The InENVI model provides an alternative approach for reconstructing NDVI datasets with both high spatial and temporal resolutions over large areas.

Index Terms—Environment variables, machine learning, spatiotemporal fusion, vegetation index.

I. INTRODUCTION

THE normalized difference vegetation index (NDVI) is a widely used satellite-based product for estimating terrestrial vegetation production [1], detecting vegetation dynamics [2], and monitoring agriculture [3]; the spatial and temporal resolutions of these data can determine the application potential of remote sensing images. There are essentially no satellite sensors that can simultaneously achieve high spatiotemporal resolution and long time series. Some remote sensing images have high temporal resolution (e.g., 0.5–2 days) but low spatial resolution (from 250 m to >1 km), such as the Moderate Resolution Imaging Spectroradiometer (MODIS) dataset. Limited by their low spatial resolution, these satellite-based NDVI datasets cannot identify vegetation changes on a fine scale [4]. Some other types of remote sensing images have low temporal resolution (e.g., >5 days) but high spatial resolution (e.g., Landsat dataset, 30 m); because of their low temporal resolution, these high spatial resolution images are easily affected by poor atmospheric conditions, such as cloudy and rainy weather [5], [6]. Consequently, observations in these datasets are often missing or contaminated [7]. In addition, these datasets are acquired relatively infrequently compared to low spatial resolution datasets, generally limiting their ability to capture key vegetation phenological phases [8].

Numerous studies have been attempting to develop methods for reproducing NDVI time-series images with both high spatial and temporal resolutions, which can be mainly divided into the temporal interpolation methods and the spatiotemporal fusion methods [9], [10]. The temporal interpolation method is based solely on a high spatial resolution dataset [11], [12], [13], [14], [15]. These interpolation methods develop a time-series method to fit the original Landsat-NDVI data and reconstruct NDVI on dates without Landsat observations to improve the temporal resolution of the dataset [16], [17]. Typical time-series fitting methods include linear harmonic [17], [18], logistic [19], and nonlinear harmonic models [10]. The performance of these interpolation methods depends heavily on the number and representation of the available Landsat observations [10]. In particular, the availability of original Landsat data during vegetation phenophase transitions and

Manuscript received 24 June 2023; revised 1 November 2023; accepted 21 December 2023. Date of publication 3 January 2024; date of current version 19 January 2024. This work was supported by the China National Funds for Distinguished Young Scientists under Grant 41925001. (Xiangqian Li and Qiongyan Peng contributed equally to this work.) (Corresponding author: Wenping Yuan.)

Xiangqian Li, Qiongyan Peng, Yi Zheng, Shangrong Lin, and Wenping Yuan are with the International Research Center of Big Data for Sustainable Development Goals, School of Atmospheric Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Zhuhai, Guangdong 519082, China (e-mail: lixq229@mail.sysu.edu.cn; pengqy25@mail2.sysu.edu.cn; zhengy335@mail.sysu.edu.cn; linshr6@mail.sysu.edu.cn; yuanwp3@mail.sysu.edu.cn).

Bin He and Jin Chen are with the State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Institute of Remote Sensing Science and Engineering, Beijing Normal University, Beijing 100875, China (e-mail: hebin@bnu.edu.cn; chenjin@bnu.edu.cn).

Yuean Qiu is with the Department of Geography, Environment and Spatial Sciences, Michigan State University, East Lansing, MI 48824 USA (e-mail: qiuyuean@msu.edu).

Yang Chen is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100045, China (e-mail: chenyang01@aircas.ac.cn).

Digital Object Identifier 10.1109/TGRS.2024.3349513

growth peak periods substantially determines the reliability of the method [6], [11]. There are many data gaps during the growing seasons in humid areas [20], which greatly restricts the application of these methods.

Spatiotemporal fusion models are another method for reconstructing NDVI time-series data with both high spatial and temporal resolutions. In contrast to temporal interpolation methods, spatiotemporal fusion models integrate high temporal resolution but low spatial resolution data (e.g., MODIS) to reconstruct NDVI [9], [21], [22], [23], and MODIS data can provide reference information for reconstruction dates to optimize the performance of the method. In general, spatiotemporal fusion models first develop the relationship between Landsat-NDVI and MODIS-NDVI on 30-m pixels on dates with both Landsat and MODIS datasets, and then, the Landsat-NDVI can be reconstructed based on the established relationship and data from the MODIS-NDVI on dates without Landsat-NDVI data. Previous studies have developed numerous spatiotemporal fusion models, such as the spatial temporal adaptive reflection fusion model (STARFM) [9], spatial temporal data fusion approach (STDFA) [24], and sparse representation-based spatiotemporal reflectance fusion model (SPSTFM) [25]. In recent years, with the continuous progresses of machine learning methods, spatiotemporal fusion models based on machine learning have been widely applied, including extreme learning [26], random forest [27], [28], convolutional neural network [29], [30], artificial neural network [31], and long-short-term memory [32].

Although many spatiotemporal fusion models have been developed, there are still two important shortcomings that significantly restrict their performance. First, almost all the existing models use one or two pairs of cloud-free Landsat and MODIS images that are adjacent to the target construction dates [10], [33], [34]. However, over humid areas, there are large ratios of temporally continuous data gaps, and it is quite difficult to find high-quality adjacent pixels, which greatly decreases the accuracy of the method [11], [35], [36]. For example, previous studies have found that available cloud-free images can be several months away from a given date [11]. Although gap-filling and Savitzky–Golay filtering (GFSG) model can be used to improve the reconstruction of time series with consecutive missing values [11], it may still be challenging to apply sparse and insufficient observations to Landsat data reconstruction.

The second shortcoming of current spatiotemporal fusion models is the neglect of environmental variables for determining the relationships between high and low spatial resolution images. Essentially, a spatiotemporal fusion model can be used to explore the relationships between the two types of satellite-based datasets [9]. However, several studies have highlighted that a linear mixing assumption is only suitable for cases of bare soil, ice, and snow, and the relationship is assumed to be nonlinear over vegetation-covered areas [37], [38], [39], [40]. Although machine-learning methods have been used to capture nonlinear relationships, they only assume that the relationship is temporally dependent [36], [41]. In fact, a nonlinear relationship between the two types of satellite-based datasets is likely to change with variations

in the environmental conditions [42]. Especially in heterogeneous vegetation regions, low spatial-resolution pixels generally include multiple vegetation types, which display diverse responses to environmental changes [43]. For instance, a previous study showed that grassland and cropland may be less adaptive to water changes than forests ecosystems because of their shallower root systems [44]; however, current machine-learning methods only consider the information contained in the base and target date images [41], [45], [46]. Machine-learning methods should incorporate environmental variables to improve the performance of the reconstructed NDVI dataset [42].

This study aims to develop a new machine-learning spatiotemporal fusion model to reconstruct high spatiotemporal resolution images during all periods, for which there are high temporal resolution images. In particular, our study utilizes all available images of high spatial resolution (Landsat) and high temporal resolution (MODIS) and integrates the impacts of environmental variables on their relationship into the fusion model. The overall objectives of this study are to 1) develop a machine-learning spatiotemporal fusion model to reconstruct high spatial resolution datasets, 2) examine the accuracy of this new model over various ecosystem types, and 3) compare the accuracy of the new model with other existing fusion models, including weighted function-based and unmixing-based models.

II. INTEGRATING ENVIRONMENTAL VARIABLE SPATIOTEMPORAL FUSION MODEL

A. Overview of InENVI

This study developed a new machine-learning fusion model, i.e., Integrating ENvironmental Variable spatiotemporal fusion (InENVI), which has the following improvements compared to previous machine-learning models. First, InENVI includes neighboring pixels that with a certain similarity to the target pixel to build the machine-learning models. InENVI is a machine-learning fusion model, and the number of training samples is important for improving model accuracy. Theoretically, there should be comparable relationships between similar high spatial-resolution pixels and the same low spatial-resolution pixels. Therefore, this study used a local window to select pixels with a certain similarity to the target pixel and included these pixels in the InENVI model to expand the number of training samples.

Second, we assumed that the relationship between high and low spatial-resolution NDVI data highly depends on environmental conditions; therefore, we included environmental variables as explanatory variables in the InENVI method. Although several fusion models have highlighted the time-varying relationship between high spatial resolution and high temporal resolution images [22], [47], none of them have used environmental variables to indicate the varying relationship. Near any given low spatial-resolution pixel, there are many neighboring pixels with high spatial resolution. In most cases, low-resolution pixels cover multiple vegetation types, and it is well established that various vegetation types response quite differently to environmental changes [42], [48]. Therefore, the time-varying relationships between high

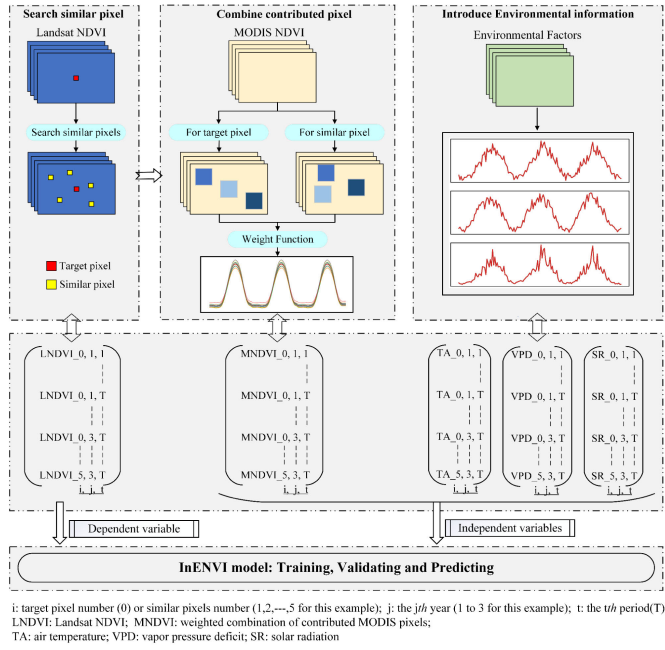


Fig. 1. Flowchart of the InENVI model.

spatial-resolution and high temporal-resolution images may be highly dependent on environmental conditions. Previous studies have analyzed the dominant environmental variables for determining temporal variations in the vegetation index and highlighted that the air temperature, vapor pressure deficit, and solar radiation are the most important variables compared to other environmental variables [42], [48]. Therefore, we selected these three environmental variables to develop the InENVI model.

Based on these assumptions, the InENVI fusion model is developed as follows:

$$\text{LNDVI}_{i,j,t} = f(\text{MNDVI}_{i,j,t}, T_{i,j,t}, \text{VPD}_{i,j,t}, R_{i,j,t}) \quad (1)$$

where $\text{LNDVI}_{i,j,t}$ is the Landsat-NDVI at the i th pixel (target pixel and similar pixels), j th year, and t th period; f is the adopted machine-learning method; $\text{MNDVI}_{i,j,t}$ is the weight-average NDVI derived from MODIS observations. $T_{i,j,t}$, $\text{VPD}_{i,j,t}$, and $R_{i,j,t}$ are the air temperature, vapor pressure deficit, and incident shortwave radiation, respectively. Fig. 1 shows the flowchart of InENVI model. The followings are the detailed method.

We first collected all Landsat images from 2001 to 2020, including Landsat5, Landsat7, and Landsat8 datasets. The revisit time of all three Landsat satellites is 16 days; therefore, the temporal resolution of combined three satellites is higher than 16 days. To match with the temporal resolution of MODIS NDVI (8 day, see Section III-A), we set each of Landsat image into the corresponding intervals of 8 day. At some intervals of 8 day, there are no Landsat images. However, there are multiple images at other intervals of 8 day, and we selected the maximum NDVI value with this interval as the NDVI value.

B. Selection of Similar Pixels

To search for similar pixels in a given target pixel (i.e., a Landsat pixel), we first set a sliding window with

1500×1500 m (i.e., 50×50 Landsat pixels). In the sliding window, we searched for similar pixels in terms of seasonal changes in the Landsat-NDVI. We then retrieved the seasonal curve (SC) of NDVI, that is, the 8-day NDVI values averaged over the last three years (the present year, previous year, and subsequent year). The averaged 8-day NDVI values of the SC should include 70% of the sample period, and if the ratio was $<70\%$, we then used the observations from the five most recent years to calculate the averaged SC of NDVI, and so on, until the ratio reached more than 70%. The purpose of selecting the closest years to generate SC is to avoid the impacts of vegetation type changes on spatiotemporal fusion, and the shorter the periods used in calculating Landsat season curve and selecting similar pixels, the higher the probability of avoiding vegetation type changes. Similar pixels were identified by calculating the determination coefficient (R^2) for the seasonal NDVI series of the target and adjacent pixels within the sliding window. If the calculated R^2 value was larger than the set threshold (i.e., 0.95), the pixel was considered to be similar to the target pixel.

This study included multiple MODIS pixels to calculate MNDVI of target and similar Landsat pixels [see (1) and Fig. 1], and these MODIS pixels are named as contribution pixels. We used the same method of selecting similar pixels to determine contribution pixels. For example, for the target Landsat pixel, we calculated the correlation coefficient of NDVI series between target Landsat pixel and all MODIS pixels within the sliding window. If the correlation coefficients (R) are larger than 0.8, the MODIS pixels are set as contribution pixels. The same method was used to determine contribution pixels for all similar pixels. In addition, we assumed the larger contributions from the closer MODIS contribution pixels to reconstruct NDVI of target pixel. Therefore, the MNDVI was used as

$$\text{MNDVI}_t = \sum_{l=1}^M \text{Con_NDVI}_{l,t} \times w_l \quad (2)$$

$$w_l = \frac{R_l}{\sum_{l=1}^M R_l} \quad (3)$$

where MNDVI_t is the weighted combined NDVI value at the t th period derived from all contribution pixels ($\text{Con_NDVI}_{l,t}$) for target pixel or similar pixel; M indicates the number of all contribution pixels; w_l is the weight value of the l th contribution pixel; R_l indicates the correlation coefficients between target pixel (or similar pixels) and contribution pixels.

C. XGBoost Model

This study used the Extreme Gradient Boosting (XGBoost) algorithm to explore the nonlinear relationship between the MODIS- and Landsat-NDVI. XGBoost was proposed by Chen and Guestrin [49] and is an optimized integrated tree algorithm. The main idea is to continuously add new trees and constantly split features to grow a tree. In fact, adding one tree each time teaches the algorithm a new function to fit the residual of the last prediction. Each tree falls into a corresponding leaf node, and each leaf node corresponds to a score. XGBoost is a great improvement over the traditional gradient boosting decision tree (GBDT) algorithm, especially

in the optimization of the objective function. Previous studies have demonstrated the good performance of XGBoost in predicting vegetation growth [42], [48].

In general, we collected all Landsat and MODIS images of the same period from 2001 to 2020, but we used only 80% of the available paired images during the last three years (the present year, previous year, and subsequent year) to train the InENVI model each time. The trained InENVI model was first used to predict high-resolution images at any desired date using only MODIS images and to predict the high spatial resolution images in the other 20% of periods with both Landsat and MODIS datasets for method validation. We recycled the aforementioned products to obtain the predicted high spatial resolution NDVI values for all periods with the cloud-free paired Landsat and MODIS images. For validation, we compared the predicted NDVI values for each period with the corresponding Landsat observations.

III. DATA AND EXPERIMENTAL SETUP

A. Satellite Data and Processing

In this study, we used MODIS-NDVI data as low spatial-resolution images. The NDVI was calculated based on the surface spectral reflectance for red (620–670 nm) and near-infrared (NIR) (841–876 nm) bands at 250-m resolution. In order to minimize the noise, such as disturbances from clouds and the atmosphere, for each MODIS pixel, the maximum value composite method was performed for the derived NDVI values from the MOD09Q1 and MYD09Q1 surface reflectance products within an 8-day period. We used the weighted Whittaker smoother to preprocess the initial MODIS-NDVI product to avoid uncertainties induced by the MODIS data. Compared with other filtering methods, the Whittaker smoother can balance fidelity and roughness well by minimizing the fitting error and penalizing the roughness of the smooth curve [50], [51].

Landsat-NDVI data were used as high spatial-resolution images. This study included surface reflectance data from the Landsat 5 ETM, 7 ETM+, and 8 OLI sensors. These sensors have near-polar orbits that can collect reflectance imagery every 16 days at a spatial resolution of 30 m. We selected these three satellites together in the study to create an 8-day return frequency for a special region. Furthermore, considering the sensor differences, we converted the NDVI data extracted from Landsat 5 or 7 sensors ($NDVI_{L5,7}$) to match the Landsat 8 OLI sensor ($NDVI_{L8}$) with the following linear transformation [52]:

$$NDVI_{L8} = 0.0235 + 0.9723 \times NDVI_{L5,7}. \quad (4)$$

B. Meteorological Data

This study used the ERA5 reanalysis dataset to drive the InENVI model, which is the latest generation of European Centre for Medium-Range Weather Forecasts (ECWMF) reanalysis data that has been substantially improved in terms of its spatiotemporal resolution, radiative transfer model, and assimilation method [53]. ERA5 dataset provides daily meteorological variables at a spatial resolution of 0.1° . In this study, we calculated the mean values of each 8-day interval to match

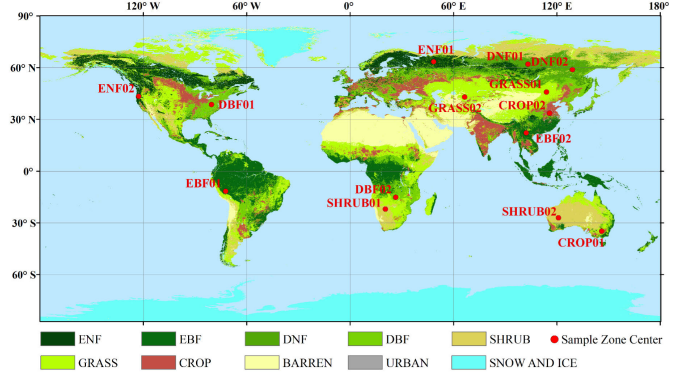


Fig. 2. Geographical distribution of the 14 study areas. The abbreviations indicate evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needleleaf forest (DNF), deciduous broadleaf forest (DBF), cropland (CROP), shrubland (SHRUB), and grassland (GRASS).

the temporal interval of Landsat and MODIS NDVI datasets, and we also resampled meteorological data to 30 m using cubic spline interpolation to match the resolution of Landsat. Air temperature and incident shortwave radiation were directly extracted from the reanalysis dataset. Notably, the ERA5 reanalysis dataset does not provide vapor pressure deficit data, which requires further calculation based on relative humidity and temperature [2].

C. Study Areas

Fourteen heterogeneous study areas were selected to run the InENVI model and examine its performance. The study area was dominated by various vegetation types, including evergreen needleleaf forests (ENFs), evergreen broadleaf forests (EBFs), deciduous needleleaf forests (DNFs), deciduous broadleaf forests (DBFs), croplands (CROPS), shrublands (SHRUBS), and grasslands (GRASSs) (see Fig. 1). All study areas covered an area of 30×30 km, and the specific location information of all study areas is labeled in Fig. 2.

D. Model Comparison and Assessment

This study assessed the accuracy of the InENVI model in reconstructing the spatiotemporal variations of the reconstructed NDVI. As discussed in Section II, this study conducted cross-validation, and 80% of the paired cloud-free MODIS and Landsat observations were used to train the InENVI model, while the other 20% of the paired observations were used for validation. In this research, the determination coefficient (R^2) was selected to evaluate the variations in observations simulated by the InENVI model. Furthermore, the root mean squared error (RMSE) and relative predictive error (BIAS) were also selected to participate in the statistical assessment, which is expressed as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (5)$$

$$BIAS = \frac{\sum_{i=1}^n |O_i - P_i|}{\sum_{i=1}^n O_i} \times 100\%. \quad (6)$$

In addition, structural similarity (SSIM) was further applied to denote the similarity between the fusion image and the true

image, ranging from -1 to 1 , and it can be calculated using the following formula:

$$\text{SSIM} = \frac{(2\mu_f\mu_t + C_1)(2\sigma_{ft} + C_2)}{(\mu_f^2 + \mu_t^2 + C_1)(\sigma_f^2 + \sigma_t^2 + C_2)} \quad (7)$$

where f and t represent the fused and true images, respectively; μ and σ refer to the mean and standard deviation, respectively; σ_{ft} is the covariance of the f and t images; and C_1 and C_2 are set as constants to avoid computation errors when divided by 0, which are set to 0.0001 and 0.0009, respectively. The higher the SSIM value, the more similar the two images are, indicating better performance of the fused model.

Additionally, to assess the performance of the InENVI model, seven typical spatiotemporal fusion models were selected for comparison with the InENVI model: the STARFM [9], enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) [22], combination of downscaling mixed pixel algorithm and spatial and temporal adaptive reflectance fusion model (CDSTARFM) [54], flexible spatiotemporal data fusion (FSDAF) [23], regression model fitting, spatial filtering and residual compensation (Fit-FC) [55], simultaneously generated full-length NDVI time series (SSFIT) [6], and GFSG [11]. The comparisons were only conducted in two cropland areas (i.e., CROP01 and CROP02) because of their high heterogeneity.

E. Sensitivity Analysis and Ablation Experimental Setup

In order to better identify the relative importance of parameters, we conducted sensitivity analysis on the developed model. In machine learning algorithms, sensitivity analysis is a crucial method for assessing how changes in the parameter settings affect the quality of the results [56]. In this study, each parameter was systematically changed, and each time accuracy was computed, in order to determine the sensitivity of the results to variation. The key variables tested include the SCs used for training, the sliding window size, and the Landsat correlation coefficient threshold. Among them, the SCs were incorporated into three different combinations as 2016–2018, 2017–2019, and 2018–2020, and 2019 was considered as the target year. The size of the sliding window was set from 10×10 to 65×65 Landsat pixels, with a step size of 5, and a total of 12 tests were conducted. The threshold for the determination coefficient (R^2) used in the identification of similar pixels was configured at five levels: 0.8, 0.85, 0.9, 0.95, and 0.98, resulting in a total of five tests.

These tests were conducted using CROP01 region as an example, and MODIS- and Landsat-NDVI and environmental variables from 2018 to 2020 were used for model training and prediction. We used only 80% of the available paired images during the selected three years to train the InENVI model each time. The trained InENVI model was used to predict the high spatial resolution images in the other 20% of periods. The above process was recycled until the predicted high spatial-resolution NDVI values for all periods were obtained. For validation, we compared the predicted NDVI values for each period with the corresponding Landsat observations.

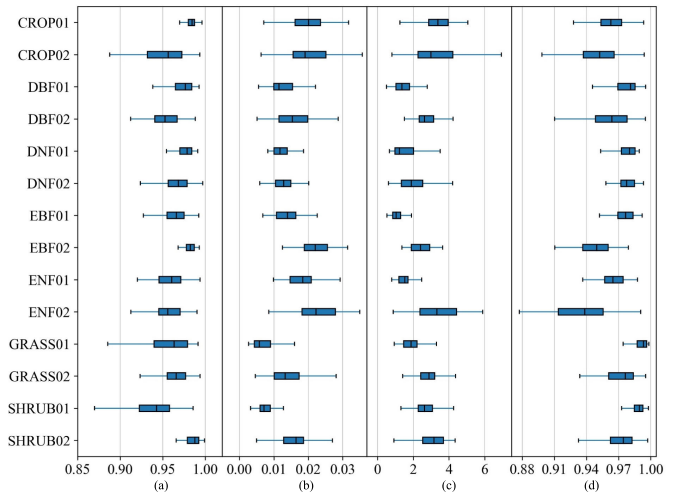


Fig. 3. Comparisons of (a) R^2 , (b) RMSE, (c) BIAS (%), and (d) SSIM among all cloud-free NDVI observations and the reconstructed NDVI over all 14 study areas from 2001 to 2020.

We also conducted ablation experiments on three important innovative steps of the InENVI model, which include searching similar pixels of Landsat, searching MODIS contributed pixels, and introducing environmental information. Ablation experiments explore the impact on model performance by removing different steps of the model, which is a very labor-saving method to study causal relationships [57]. To verify the importance of each step, the contributed pixels of MODIS, the similar pixels of Landsat, and environmental variables were removed, respectively. In each experiment, the verification kept consistency with previous tests, and the results were also compared with the original InENVI model, which did not remove any steps.

IV. RESULTS

A. Accuracy Assessment of the InENVI Fusion Model

First, this study examined the accuracy of the reconstructed NDVI using the InENVI fusion model to reproduce spatial variations. In each study area, we selected all cloud-free Landsat images during the study period (2001–2020), which were not used to train the InENVI model for validation. We calculated four statistical metrics (see Section III-D) for each period and then calculated the mean values of these four metrics over all investigated periods to validate the model performance. The results indicate that the InENVI model can accurately reflect the spatial variations in NDVI for various vegetation types (see Fig. 3). The mean values of R^2 were above 0.90 over all 14 study areas, and the highest R^2 reaches 0.99 [see Fig. 3(a)]. The average RMSE and BIAS values were below 0.03 and 5% over all areas [see Fig. 3(b) and (c)], respectively, and the average SSIM values were above 0.9 [see Fig. 3(d)].

To better evaluate the actual performance of the InENVI model, the study area of CROP01, with strong vegetation heterogeneity, was selected for a detailed assessment. In this region, four Landsat-NDVI images on March 13, June 9, September 21, and December 18 in 2020 served as reference data for independent verification under cloudy and cloud-free

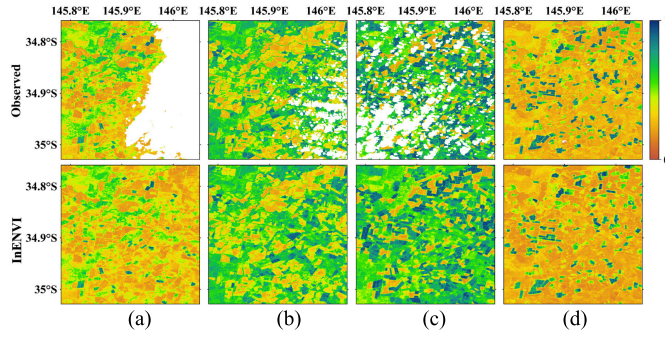


Fig. 4. Comparison between Landsat-NDVI observations and the reconstructed NDVI at the CROP01 site on (a) March 13, (b) June 9, (c) September 21, and (d) December 18 in 2020.

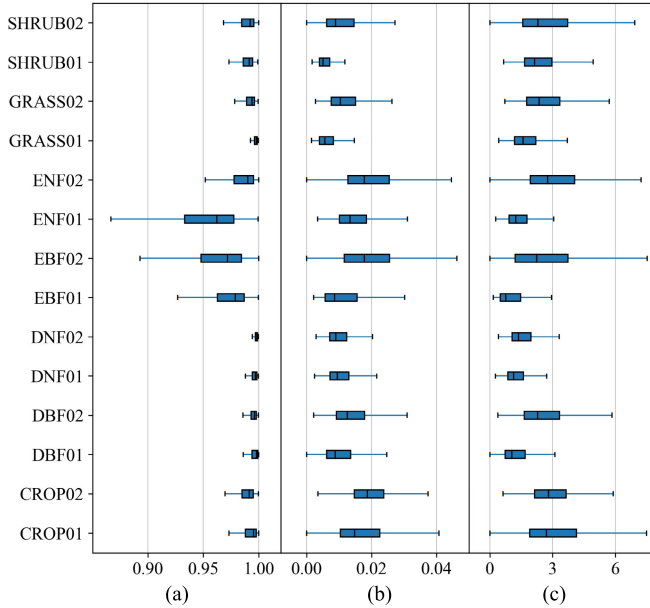


Fig. 5. Comparisons of (a) R^2 , (b) RMSE, and (c) BIAS (%) among all cloud-free observed NDVI values and the reconstructed NDVI over 14 study sites from 2001 to 2020.

conditions (see Fig. 4). It is evident from Fig. 4 that the InENVI model can accurately reconstruct the NDVI. Spatial details were captured correctly in the fused images [see Fig. 4(a)–(c)].

We further examined the performance of the InENVI model in reproducing temporal variations in NDVI over all study areas. Based on the 8-day reconstructed NDVI values and observations from 2001 to 2020, we calculated three metrics (i.e., R^2 , RMSE, and BIAS) at each pixel and used the mean values of the three metrics over the entire study area to indicate the model performance for reproducing the temporal variations of NDVI. The results revealed that the reconstructed NDVI accurately indicated the temporal variations in NDVI in all 14 study areas (see Fig. 5). The average overall R^2 was above 0.95, and it almost reached 1 in several study areas [see Fig. 5(a)]. Compared to the other ecosystem types, the two EBFs showed a lower average R^2 and higher RMSE [see Fig. 5(a) and (b)].

As an example, this study showed a detailed validation of temporal variations in the reconstructed NDVI in the CROP01 area. The mean values of the reconstructed

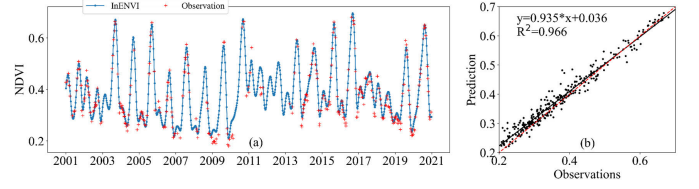


Fig. 6. Validation of the temporal variations of the Landsat-NDVI observations and reconstructed NDVI at the CROP01 site. (a) Temporal changes of the Landsat-NDVI observations and constructed NDVI averaged over all pixels at the CROP01 site. (b) Correlation between the observed and reconstructed NDVI at the CROP01 site.

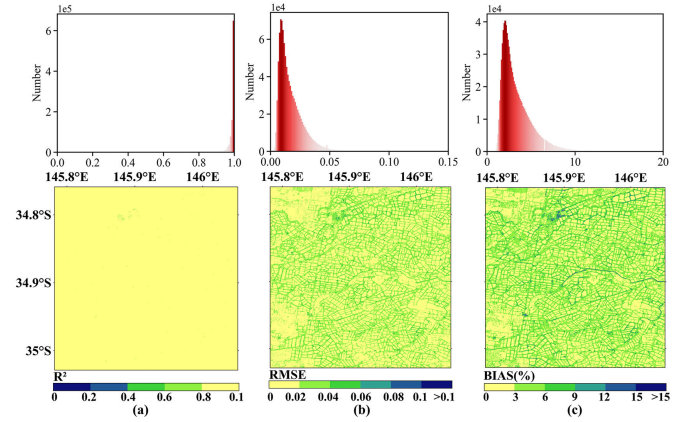


Fig. 7. Comparisons between the observed and reconstructed NDVI at the CROP01 site from 2001 to 2020 showing (a) R^2 , (b) RMSE, and (c) BIAS.

NDVI over the entire study area agreed very well with the Landsat-NDVI observations in terms of temporal variations on an 8-day scale from 2001 to 2020 [see Fig. 6(a)], and the temporal correlation between the average reconstructed NDVI and the observations reached approximately 0.97 [see Fig. 6(b)]. In addition, for almost all pixels, the reconstructed NDVI indicated very good temporal variation in the observed NDVI. The BIAS of most pixels (>84%) was <5% [see Fig. 7(c)]. The RMSE is below 0.04 for over 95.98% of the pixels, and the mean RMSE is approximately 0.02, with high values mainly distributed at the border of fields and roads [see Fig. 7(b)].

B. Method Comparison

This study compared the performance of the InENVI model with seven other spatiotemporal fusion models (see Section III-D) in two cropland regions (i.e., CROP01 and CROP02) because of high vegetation heterogeneity. These two areas have been selected by previous studies for data fusion method validation [6], [11]. First, we compared the model performance in reproducing the spatial pattern of the NDVI in the two study areas. The mean values of the four metrics (i.e., R^2 , RMSE, BIAS, and SSIM) showed the best performance for the InENVI model among all seven investigated models (see Fig. 8). The mean values of R^2 reached 0.98 and 0.93, the RMSE values were 0.03 and 0.03, and the bias values were 3.85% and 4.53% for the InENVI model at the CROP01 and CROP02 areas, respectively. In addition, the fact that SSIM (0.99) is closer to 1 than the SSIM of the other methods indicates that the InENVI model has the

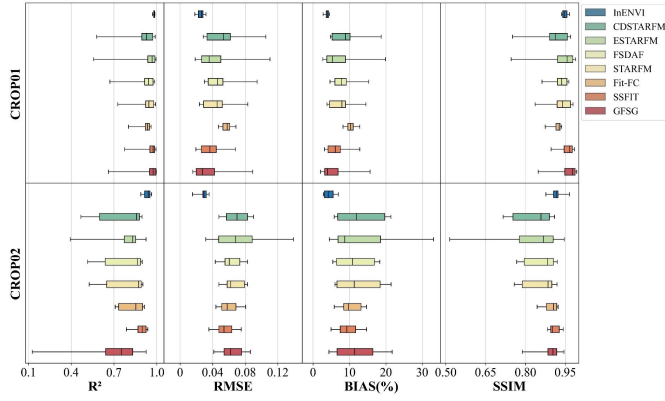


Fig. 8. Comparisons of the InENVI model with the other seven models for reproducing spatial variations of NDVI in CROP01 (2017) and CROP02 (2018). The four metrics were calculated for every cloud-free moment.

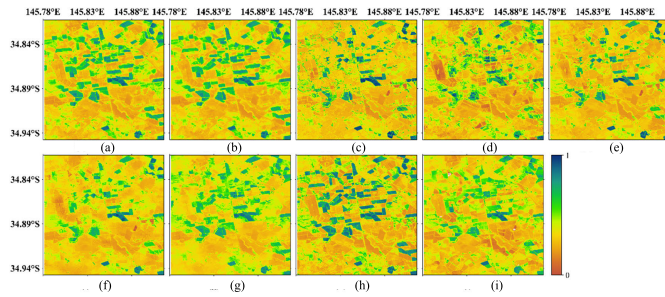


Fig. 9. Comparisons of the method performance for reproducing spatial variabilities at the CROP01 site on December 11, 2017. (a) Observed NDVI, (b) InENVI, (c) CDSTARFM, (d) ESTARFM, (e) STARFM, (f) FSDAF, (g) Fit-FC, (h) SSFIT, and (i) GFSG.

highest similarity with the observations compared to the other methods.

Figs. 9 and 10 show a detailed comparison of the spatial patterns derived from the eight models in these two cropland study areas. Compared with the other seven methods, the NDVI reconstructed using the InENVI model more closely represented the actual spatial pattern of the NDVI [see Fig. 9(b)]. It can capture land-cover borders correctly and distinguish different patches of vegetation. In contrast, the CDSTARFM, ESTARFM, and STARFM methods had large deviations compared to the observations (see Fig. 9). The reconstructed NDVI of SSFIT showed clear borders of certain patches but significantly overestimated the NDVI [see Fig. 9(h)]. Quantitative indicators also demonstrated that the InENVI model provided the most accurate predictions in these two areas (see Figs. 9 and 10). It closely replicated the actual spatial pattern of NDVI, with $R^2 = 0.99$ and $RMSE = 0.02$, which was superior to the other models. Similarly, in the CROP02 study area, the spatial patterns of six existing methods differed significantly from the actual observations. Except for the fit-FC model, all other methods had different degrees of overestimation, which was especially visible in the ESTARFM method [see Fig. 11(d)]. In contrast, the InENVI model reproduced the Landsat observations well and accurately depicted the spatial variation in the NDVI [see Fig. 11(b)]. The scatterplots in Fig. 12(a)–(e) show that the InENVI model with observations had a closer adherence to

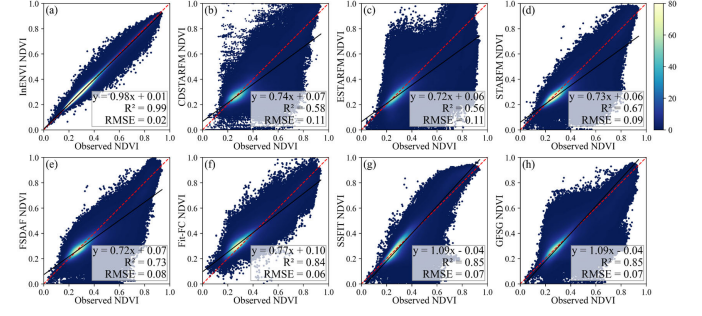


Fig. 10. Correlations between the observed and reconstructed NDVI at the CROP01 site on December 11, 2017. (a) InENVI, (b) CDSTARFM, (c) ESTARFM, (d) STARFM, (e) FSDAF, (f) Fit-FC, (g) SSFIT, and (h) GFSG.

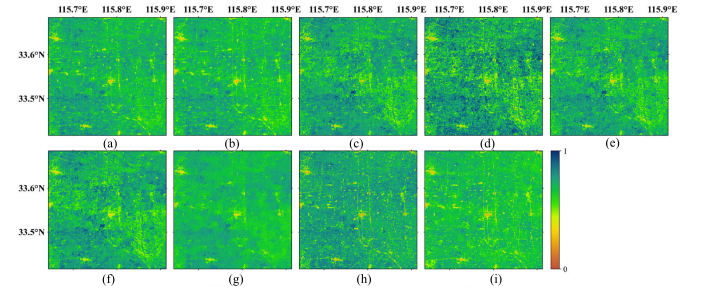


Fig. 11. Comparisons of the method performance for reproducing the spatial variability at the CROP02 site on September 6, 2018. (a) Observed NDVI, (b) InENVI, (c) CDSTARFM, (d) ESTARFM, (e) STARFM, (f) FSDAF, (g) Fit-FC, (h) SSFIT, and (i) GFSG.

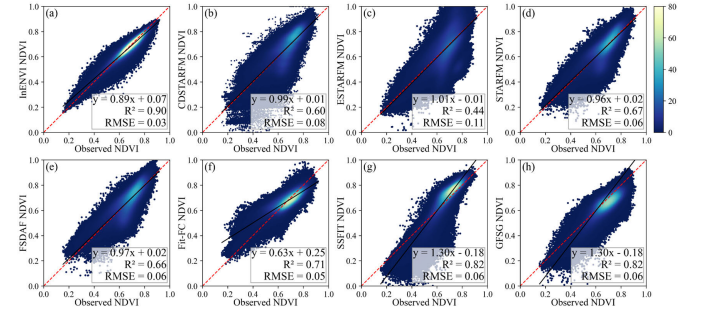


Fig. 12. Correlations between the observed and reconstructed NDVI at the CROP02 site on September 6, 2018. (a) InENVI, (b) CDSTARFM, (c) ESTARFM, (d) STARFM, (e) FSDAF, (f) Fit-FC, (g) SSFIT, and (h) GFSG.

the 1:1 line and also had the lowest RMSE (0.03) and the highest R^2 (0.90).

V. DISCUSSION

A. Sensitivity Analysis and Ablation Experiments

1) *Parameter Sensitivity Analysis:* First, we incorporated three different SCs into our model and compared the reconstruction accuracy. SC1 indicates the reconstruction accuracy using SC derived from satellite-based observations of 2018–2020, SC2 indicates the accuracy using SCs from the data of 2017–2019, and SC3 indicates the accuracy using SCs from the data of 2016–2018 [see Fig. 13(a)]. In general, the model accuracies of three tests are not statistically different [see Fig. 13(b)], which implied the InENVI model is reliable and predictable.

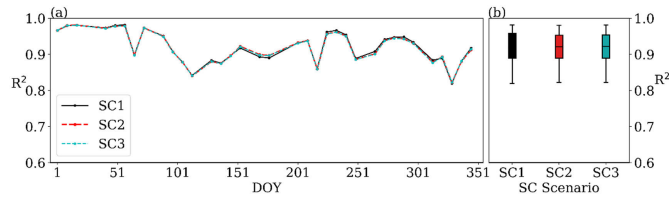


Fig. 13. Comparisons of reconstruction accuracy under different seasonal curve scenarios at the CROP01 site in 2019. (a) R^2 and (b) averaged R^2 among all cloud-free NDVI observations and the reconstructed NDVI in all periods.

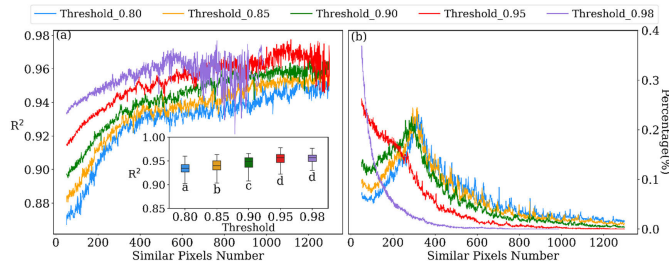


Fig. 14. Comparison of reconstruction accuracy and similar pixels number for each target pixel at the CROP01 site for different thresholds. (a) Relationship between accuracy and similar pixels number. The letters in the inset indicate the statistically significant differences ($p < 0.05$) over five groups with different thresholds. (b) Frequency distribution of similar pixels number.

Second, we evaluated the influence of Landsat determination coefficient threshold for the InENVI model. This threshold indicates a tradeoff of number and similarity of similar pixels. The higher threshold means the higher similarity of similar pixels, but the smaller number of similar pixels [see Fig. 14(b)] and vice versa. It can be seen from Fig. 14(a), as the threshold increases, the reconstruction accuracy of the model also increases. When the threshold reaches above 0.95, the average $R^2 > 0.9$. However, a higher threshold does not necessarily equate to better results. When the threshold reaches 0.98, the number of searchable similar pixels decreases, and the reconstruction accuracy also becomes unstable [see Fig. 14(a)]. The results indicated that our model is sensitive to the threshold of Landsat correlation coefficient, and the current threshold (i.e., 0.95) can get the best reconstruction accuracy over all pixels [see Fig. 14(a)].

Third, the influence of the sliding window size for the InENVI model was tested. As shown in Fig. 15, at most times, as the size of the sliding window increases, the fusion accuracy of the model also increases. When the sliding window size reaches 50×50 Landsat pixels, the fusion accuracy of the model almost no longer increases, and further increases in window size lead to a decrease in accuracy. This decline might be attributed to excessively large windows that introducing redundant information, thus affecting the accuracy of the model. Besides, as the sliding window increases to 55×55 Landsat pixels, its computation time increases rapidly (see Fig. 15). Therefore, after comprehensive consideration, we chose a 50×50 sliding window, which can well balance high accuracy and efficiency.

2) *Ablation Experiments*: We further conducted ablation experiments on three innovative steps of the developed model to determine the relative importance of each step. It can

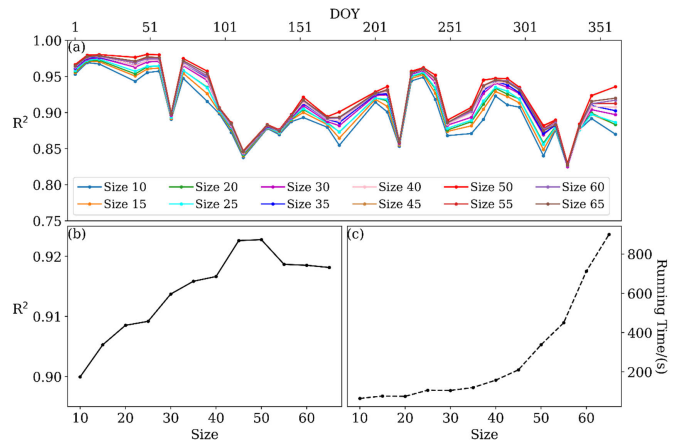


Fig. 15. Comparisons of R^2 trend (a), averaged R^2 (b) and computation efficiency (c) among all cloud-free NDVI observations and the reconstructed NDVI at the CROP01 site with different sliding window.

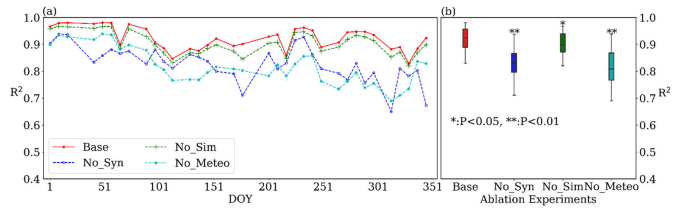


Fig. 16. Comparisons of reconstruction accuracy (R^2) trend (a) and variations (b) in different ablation experiments. Base indicates employing our optimized InENVI model for reconstruction and validation. No_Sim indicates implementing the InENVI model while removing the step of searching for similar pixels. No_Syn indicates applying the InENVI model with the exclusion of the steps involving the search for MODIS contribution pixels and the synthesis of the weighted combined curve. No_Meteo indicates executing the InENVI model by omitting environmental variables.

be seen from Fig. 16(a) that when the three steps were removed, respectively, the accuracy of the model decreased in all periods. After ignoring the MODIS contribution pixels and removing environmental variables, respectively, R^2 of the model decreased significantly ($p < 0.01$), and the average R^2 decreased by 0.10 and 0.11 compared with the original model [see Fig. 16(b)]. When removing similar pixels of Landsat, the average R^2 decreased about 0.02, showing a significant decrease ($p < 0.05$). The results showed that the input of MODIS contribution pixels, Landsat similar pixels, and environmental variables can effectively improve the accuracy of the model.

B. Superiority of the InENVI Model

In this study, a new machine-learning fusion model (InENVI) was developed by incorporating environmental variables to combine the MODIS- and Landsat-NDVI datasets. The validation in various global vegetated areas indicated the good performance of the InENVI model for reconstructing high-quality spatiotemporal resolution datasets. In addition, a comparison with seven other spatiotemporal fusion models showed that our new model can more accurately produce a synthetic high spatial-resolution NDVI dataset. In particular, the InENVI model shows several improvements over the available weighted-function-based, unmixing-based, and machine-learning methods.

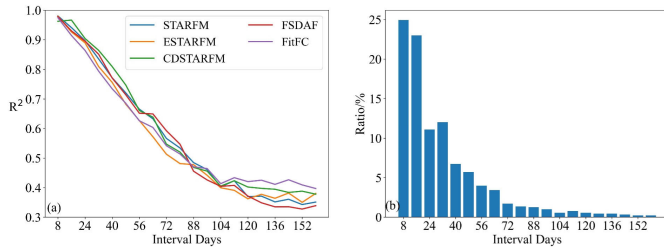


Fig. 17. Comparing the applicability of five typical models. The figure shows the average R^2 of the five models for predicting the target date NDVI image at different intervals.

First, most existing methods use adjacent pairs of cloud-free high and low spatial-resolution images (base images) in the training model to produce high spatial-resolution datasets [9], [22], [23], [54], [55]. In this study, five fusion models (except InENVI, SSFIT, and GFSG) used one or several pairs of cloud-free images. However, there is a high proportion of cloud contamination in numerous regions, particularly in humid areas [24]. Therefore, it is difficult to obtain cloud-free satellite images in adjacent periods for a target study period, and the available images may be obtained during a time that is a couple of months away from the study period [58]. For example, over the study area of CROP01, the interval number of days should be set to 88 to ensure that most of the given dates can obtain the available cloud-free images [see Fig. 17(b)].

Furthermore, our study showed decreased accuracy of the five investigated models with the time intervals between the target date and the date closest to the available paired satellite images (see Fig. 17). For example, for the STARFM model, the R^2 values between the observed and reconstructed NDVI were more than 0.9 when the base images were 24 days apart, but R^2 decreased to below 0.5 when the time intervals were longer than 88 days [see Fig. 17(a)]. The InENVI model generated a fused NDVI dataset by integrating all available paired cloud-free high and low spatial-resolution images over a long-term period (i.e., three years in this study).

Second, similar pixels were included within the sliding window to train the InENVI model. Previous fusion models (e.g., ESTARFM and FSDAF) have also shown that integrating information from additional images can improve fusion accuracy [22], [23]. A recent learning-based method also included information from several neighboring images to predict the reflectance for the target date, which showed substantial improvement in producing spatiotemporal fusion data [59]. The potential cause for this improvement may be that neighboring images increase the training dataset size and are helpful for learning-based methods to develop better models [48].

Third, the InENVI model shows a higher computational efficiency than the other six state-of-the-art fusion models (see Fig. 18). Although numerous fusion methods have been developed over the past few decades [9], [33], [34], [35], [54], [55], no regional or global fused datasets have been generated. One of the most important reasons for this is the high computation requirements given the high spatiotemporal resolution of the reconstructed datasets [60]. Therefore, it is crucial to quantify the computational efficiency of the fusion

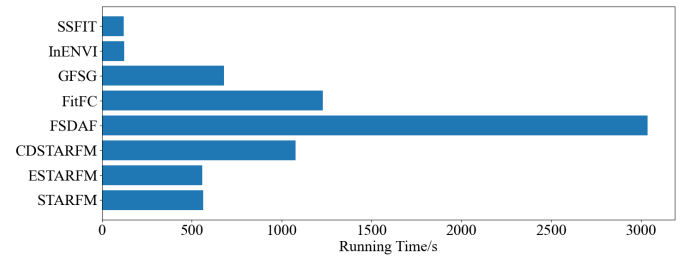


Fig. 18. Comparison of the computation times for InENVI and the other seven fusion models. All models are run to reconstruct the 8-day NDVI of one year with a spatial resolution of 30×30 m over a 30×30 km region.

models to determine the capability of producing regional and global fusion datasets. Based on a desktop computer with a CPU of eight processors at 2.7 GHz, the InENVI requires 124 s to reconstruct the 8-day NDVI data for one year with 30×30 m spatial resolution over a region of 30×30 km, which has the best computational efficiency in addition to SSFIT (see Fig. 18). The InENVI method also supports cluster parallelism and has the potential to produce fused NDVI datasets at regional and global scales.

C. Limitations of the InENVI Model

Most state-of-the-art fusion models explicitly reconstruct the reflectance values of multiple spectral bands [9], [33], [34], [35], [54]. Considering the important role of NDVI, InENVI was designed to reconstruct the NDVI dataset, and this study did not take any efforts to examine its performance for reconstructing other reflectance data. In particular, NDVI indicates vegetation growth; therefore, InENVI was designed to incorporate environmental variables to improve fusion accuracy, as the environmental variables dominate vegetation growth. The reflectance of spectral bands is also highly dependent on other factors, such as the water conditions of the land surface; therefore, the capability of InENVI to process the reflectance of spectral bands still needs further validation.

Second, the InENVI model, such as other spatiotemporal fusion models, integrated MODIS-NDVI data to reconstruct high spatiotemporal resolution NDVI data. MODIS-NDVI data are less affected by land cover or other noise and have high temporal resolution compared to Landsat-NDVI data [61], [62]. However, there are still a large number of data gaps for MODIS-NDVI data over many areas, especially in humid areas [63], [64]. Therefore, it is challenging to acquire high-quality MODIS-NDVI data for spatiotemporal fusion [11]. This study used the weighted Whittaker smoothing method to fill the data gaps in the MODIS-NDVI data [50]. Although the weighted Whittaker smoothing method has been examined as a good candidate for improving the quality of MODIS-NDVI data, some uncertainties still remain [50]. Any noise in the MODIS-NDVI propagates into the InENVI model.

VI. CONCLUSION

Traditional fusion models have long been constrained by the assumption of linear mixture theory and the restriction of input data. These issues limit their ability to generate long-term consistent time series of high-resolution images. Consequently,

this study proposes a data-driven fusion algorithm (InENVI) based on machine learning to address these challenges. In comparison with previous studies, the InENVI model does not require any cloud-free high-resolution images as necessary input, allowing it to fully utilize all the collected MODIS- and Landsat-NDVI time-series images and consider the environmental conditions to solve the nonlinear problems. Furthermore, in contrast to the six typical algorithms, InENVI shows a more stable and accurate fusion at all characteristic regions of interesting (ROIs), suggesting that it can be used to generate better NDVI products and improve related studies. In conclusion, the InENVI model improves the capability of generating long time-series NDVI images with both high spatial and temporal resolution. This capability is beneficial for monitoring terrestrial ecosystem dynamics at finer spatiotemporal scales.

APPENDIX

See Figs. 13–15.

REFERENCES

- [1] W. Cai et al., “Large differences in terrestrial vegetation production derived from satellite-based light use efficiency models,” *Remote Sens.*, vol. 6, no. 9, pp. 8945–8965, Sep. 2014, doi: [10.3390/rs6098945](https://doi.org/10.3390/rs6098945).
- [2] W. Yuan et al., “Increased atmospheric vapor pressure deficit reduces global vegetation growth,” *Sci. Adv.*, vol. 5, no. 8, Aug. 2019, Art. no. eaax1396, doi: [10.1126/sciadv.aax1396](https://doi.org/10.1126/sciadv.aax1396).
- [3] Y. Fu et al., “A satellite-based method for national winter wheat yield estimating in China,” *Remote Sens.*, vol. 13, no. 22, p. 4680, 2021.
- [4] X. Huang et al., “High spatial resolution vegetation gross primary production product: Algorithm and validation,” *Sci. Remote Sens.*, vol. 5, Jun. 2022, Art. no. 100049, doi: [10.1016/J.SRS.2022.100049](https://doi.org/10.1016/J.SRS.2022.100049).
- [5] J. Li and D. Roy, “A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring,” *Remote Sens.*, vol. 9, no. 9, p. 902, Aug. 2017, doi: [10.3390/RS9090902](https://doi.org/10.3390/RS9090902).
- [6] 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](https://doi.org/10.1016/J.JAG.2021.102333).
- [7] J. Dong et al., “Early-season mapping of winter wheat in China based on Landsat and sentinel images,” *Earth Syst. Sci. Data*, vol. 12, no. 4, pp. 3081–3095, Nov. 2020, doi: [10.5194/ESSD-12-3081-2020](https://doi.org/10.5194/ESSD-12-3081-2020).
- [8] R. Shen, H. Lu, W. Yuan, X. Chen, and B. He, “Regional evaluation of satellite-based methods for identifying end of vegetation growing season,” *Remote Sens. Ecol. Conservation*, vol. 7, no. 4, pp. 685–699, Dec. 2021, doi: [10.1002/RSE2.223](https://doi.org/10.1002/RSE2.223).
- [9] 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](https://doi.org/10.1109/TGRS.2006.872081).
- [10] 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](https://doi.org/10.1016/J.RSE.2018.06.038).
- [11] 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. Photogramm. Remote Sens.*, vol. 180, pp. 174–190, Oct. 2021, doi: [10.1016/J.ISPRSJPRS.2021.08.015](https://doi.org/10.1016/J.ISPRSJPRS.2021.08.015).
- [12] D. Chu et al., “Long time-series NDVI reconstruction in cloud-prone regions via spatio-temporal tensor completion,” *Remote Sens. Environ.*, vol. 264, Oct. 2021, Art. no. 112632, doi: [10.1016/J.RSE.2021.112632](https://doi.org/10.1016/J.RSE.2021.112632).
- [13] 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](https://doi.org/10.1016/J.JAG.2021.102640).
- [14] J. M. Chen, F. Deng, and M. Chen, “Locally adjusted cubic-spline capping for reconstructing seasonal trajectories of a satellite-derived surface parameter,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2230–2238, Aug. 2006, doi: [10.1109/TGRS.2006.872089](https://doi.org/10.1109/TGRS.2006.872089).
- [15] H. Shen et al., “A spatiotemporal constrained machine learning method for OCO-2 solar-induced chlorophyll fluorescence (SIF) reconstruction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4413817, doi: [10.1109/TGRS.2022.3204885](https://doi.org/10.1109/TGRS.2022.3204885).
- [16] H. Carrao, P. Gonçalves, and M. Caetano, “A nonlinear harmonic model for fitting satellite image time series: Analysis and prediction of land cover dynamics,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1919–1930, Apr. 2010, doi: [10.1109/TGRS.2009.2035615](https://doi.org/10.1109/TGRS.2009.2035615).
- [17] D. Pouliot and R. Latifovic, “Reconstruction of Landsat time series in the presence of irregular and sparse observations: Development and assessment in North–Eastern Alberta, Canada,” *Remote Sens. Environ.*, vol. 204, pp. 979–996, Jan. 2018, doi: [10.1016/J.RSE.2017.07.036](https://doi.org/10.1016/J.RSE.2017.07.036).
- [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](https://doi.org/10.1016/J.RSE.2015.02.009).
- [19] V. Kovalskyy, D. Roy, X. Y. Zhang, and J. Ju, “The suitability of multi-temporal web-enabled Landsat data NDVI for phenological monitoring—A comparison with flux tower and MODIS NDVI,” *Remote Sens. Lett.*, vol. 3, no. 4, pp. 325–334, Jul. 2011, doi: [10.1080/01431161.2011.593581](https://doi.org/10.1080/01431161.2011.593581).
- [20] B. Pan et al., “High resolution distribution dataset of double-season paddy Rice in China,” *Remote Sens.*, vol. 13, no. 22, p. 4609, Nov. 2021, doi: [10.3390/RS13224609](https://doi.org/10.3390/RS13224609).
- [21] T. Hilker et al., “A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS,” *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613–1627, Aug. 2009, doi: [10.1016/J.RSE.2009.03.007](https://doi.org/10.1016/J.RSE.2009.03.007).
- [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](https://doi.org/10.1016/J.RSE.2010.05.032).
- [23] 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](https://doi.org/10.1016/J.RSE.2015.11.016).
- [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, no. 1, Mar. 2012, Art. no. 063507, doi: [10.1117/1.JRS.6.063507](https://doi.org/10.1117/1.JRS.6.063507).
- [25] 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](https://doi.org/10.1109/TGRS.2012.2186638).
- [26] X. Wen, H. Liu, G. Yan, and F. Sun, “Weakly paired multimodal fusion using multilayer extreme learning machine,” *Soft Comput.*, vol. 22, no. 11, pp. 3533–3544, Jun. 2018, doi: [10.1007/s00500-018-3108-y](https://doi.org/10.1007/s00500-018-3108-y).
- [27] N. Kausar and A. Majid, “Random forest-based scheme using feature and decision levels information for multi-focus image fusion,” *Pattern Anal. Appl.*, vol. 19, no. 1, pp. 221–236, Feb. 2016, doi: [10.1007/s10044-015-0448-4](https://doi.org/10.1007/s10044-015-0448-4).
- [28] X. Li et al., “Unmixing-based spatiotemporal image fusion based on the self-trained random forest regression and residual compensation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5406319, doi: [10.1109/TGRS.2023.3308902](https://doi.org/10.1109/TGRS.2023.3308902).
- [29] D. Jia, C. Cheng, S. Shen, and L. Ning, “Multitask deep learning framework for spatiotemporal fusion of NDVI,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5616313, doi: [10.1109/TGRS.2021.3140144](https://doi.org/10.1109/TGRS.2021.3140144).
- [30] 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, p. 236, Feb. 2018, doi: [10.3390/RS10020236](https://doi.org/10.3390/RS10020236).
- [31] Y. Zhao, P. Hou, J. Jiang, J. Zhao, Y. Chen, and J. Zhai, “High-spatial-resolution NDVI reconstruction with GA-ANN,” *Sensors*, vol. 23, no. 4, p. 2040, Feb. 2023, doi: [10.3390/S23042040](https://doi.org/10.3390/S23042040).
- [32] B. Mishra and T. B. Shahi, “Deep learning-based framework for spatiotemporal data fusion: An instance of Landsat 8 and Sentinel 2 NDVI,” *J. Appl. Remote Sens.*, vol. 15, no. 3, Sep. 2021, Art. no. 034520, doi: [10.1117/1.JRS.15.034520](https://doi.org/10.1117/1.JRS.15.034520).
- [33] 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](https://doi.org/10.1016/J.RSE.2019.03.012).

- [34] 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](https://doi.org/10.1016/j.rse.2018.04.042).
- [35] S. N. Goward et al., "Landsat's enduring legacy: Pioneering global land observations from space," *Photogrammetric Eng. Remote Sens.*, vol. 88, no. 6, pp. 357–358, Jun. 2022, doi: [10.14358/PERS.88.6.357](https://doi.org/10.14358/PERS.88.6.357).
- [36] J. Ju and D. P. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1196–1211, Mar. 2008, doi: [10.1016/j.rse.2007.08.011](https://doi.org/10.1016/j.rse.2007.08.011).
- [37] C. C. Borel and S. A. W. Gerstl, "Nonlinear spectral mixing models for vegetative and soil surfaces," *Remote Sens. Environ.*, vol. 47, no. 3, pp. 403–416, 1994, doi: [10.1016/0034-4257\(94\)90107-4](https://doi.org/10.1016/0034-4257(94)90107-4).
- [38] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014, doi: [10.1109/JSTARS.2014.2320576](https://doi.org/10.1109/JSTARS.2014.2320576).
- [39] T. W. Ray and B. C. Murray, "Nonlinear spectral mixing in desert vegetation," *Remote Sens. Environ.*, vol. 55, no. 1, pp. 59–64, Jan. 1996, doi: [10.1016/0034-4257\(95\)00171-9](https://doi.org/10.1016/0034-4257(95)00171-9).
- [40] C. Small and D. Sousa, "The Sentinel 2 MSI spectral mixing space," *Remote Sens.*, vol. 14, no. 22, p. 5748, Nov. 2022, doi: [10.3390/rs14225748](https://doi.org/10.3390/rs14225748).
- [41] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 821–829, Mar. 2018, doi: [10.1109/JSTARS.2018.2797894](https://doi.org/10.1109/JSTARS.2018.2797894).
- [42] X. Li, W. Yuan, and W. Dong, "A machine learning method for predicting vegetation indices in China," *Remote Sens.*, vol. 13, no. 6, p. 1147, Mar. 2021, doi: [10.3390/rs13061147](https://doi.org/10.3390/rs13061147).
- [43] K. L. Roth, D. A. Roberts, P. E. Dennison, S. H. Peterson, and M. Alonzo, "The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data," *Remote Sens. Environ.*, vol. 171, pp. 45–57, Dec. 2015, doi: [10.1016/j.rse.2015.10.004](https://doi.org/10.1016/j.rse.2015.10.004).
- [44] W. Yuan et al., "Multiyear precipitation reduction strongly decreases carbon uptake over Northern China," *J. Geophys. Res., Biogeosciences*, vol. 119, no. 5, pp. 881–896, May 2014, doi: [10.1002/2014JG002608](https://doi.org/10.1002/2014JG002608).
- [45] V. Moosavi, A. Talebi, M. H. Mokhtari, S. R. F. Shamsi, and Y. Niazi, "A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature," *Remote Sens. Environ.*, vol. 169, pp. 243–254, Nov. 2015, doi: [10.1016/j.rse.2015.08.015](https://doi.org/10.1016/j.rse.2015.08.015).
- [46] Y. Ke, J. Im, S. Park, and H. Gong, "Downscaling of MODIS one kilometer evapotranspiration using Landsat-8 data and machine learning approaches," *Remote Sens.*, vol. 8, no. 3, p. 215, Mar. 2016, doi: [10.3390/rs8030215](https://doi.org/10.3390/rs8030215).
- [47] S. Yang and X. Wang, "Sparse representation and SRCNN based spatio-temporal information fusion method of multi-sensor remote sensing data," *J. Netw. Intell.*, vol. 6, no. 1, pp. 40–53, Feb. 2021. Accessed: Sep. 30, 2022. [Online]. Available: <https://bit.kuas.edu.tw/~jni/2021/vol6/s1/04-v6n1-0085.pdf>
- [48] Q. Peng et al., "How well can we predict vegetation growth through the coming growing season?" *Sci. Remote Sens.*, vol. 5, Jun. 2022, Art. no. 100043, doi: [10.1016/j.srs.2022.100043](https://doi.org/10.1016/j.srs.2022.100043).
- [49] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [50] D. Kong, Y. Zhang, X. Gu, and D. Wang, "A robust method for reconstructing global MODIS EVI time series on the Google Earth engine," *ISPRS J. Photogramm. Remote Sens.*, vol. 155, pp. 13–24, Sep. 2019, doi: [10.1016/j.isprsjprs.2019.06.014](https://doi.org/10.1016/j.isprsjprs.2019.06.014).
- [51] P. H. C. Eilers, V. Pesendorfer, and R. Bonifacio, "Automatic smoothing of remote sensing data," in *Proc. 9th Int. Workshop Anal. Multi-temporal Remote Sens. Images (MultiTemp)*, Jun. 2017, pp. 1–3, doi: [10.1109/MULTI-TEMP.2017.8076705](https://doi.org/10.1109/MULTI-TEMP.2017.8076705).
- [52] D. P. Roy et al., "Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity," *Remote Sens. Environ.*, vol. 185, pp. 57–70, Nov. 2016, doi: [10.1016/j.rse.2015.12.024](https://doi.org/10.1016/j.rse.2015.12.024).
- [53] L. Hoffmann et al., "From ERA-interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations," *Atmos. Chem. Phys.*, vol. 19, no. 5, pp. 3097–3124, Mar. 2019, doi: [10.5194/acp-19-3097-2019](https://doi.org/10.5194/acp-19-3097-2019).
- [54] X. Dengfeng, Z. Jinshui, S. Peijun, P. Yaozhong, Y. Ya, and Y. Zhoumiqi, "Remote sensing data fusion by combining STARFM and downscaling mixed pixel algorithm," *Nat. Remote Sens. Bull.*, vol. 20, no. 1, pp. 62–72, Jan. 2016, doi: [10.11834/JRS.20165058](https://doi.org/10.11834/JRS.20165058).
- [55] 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](https://doi.org/10.1016/j.rse.2017.10.046).
- [56] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 2, pp. 135–149, Jun. 2007, doi: [10.1016/j.isprsjprs.2007.03.001](https://doi.org/10.1016/j.isprsjprs.2007.03.001).
- [57] W. Han et al., "Geological remote sensing interpretation using deep learning feature and an adaptive multisource data fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4510314, doi: [10.1109/TGRS.2022.3183080](https://doi.org/10.1109/TGRS.2022.3183080).
- [58] L. Wang et al., "A summary of the special issue on remote sensing of land change science with Google Earth Engine," *Remote Sens. Environ.*, vol. 248, Oct. 2020, Art. no. 112002, doi: [10.1016/j.rse.2020.112002](https://doi.org/10.1016/j.rse.2020.112002).
- [59] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4273–4286, May 2021, doi: [10.1109/TGRS.2020.3010530](https://doi.org/10.1109/TGRS.2020.3010530).
- [60] H. Liu, P. Gong, J. Wang, X. Wang, G. Ning, and B. Xu, "Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020—iMap World 1.0," *Remote Sens. Environ.*, vol. 258, Jun. 2021, Art. no. 112364, doi: [10.1016/j.rse.2021.112364](https://doi.org/10.1016/j.rse.2021.112364).
- [61] 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, pp. 7865–7891, Jun. 2015, doi: [10.3390/rs70607865](https://doi.org/10.3390/rs70607865).
- [62] L. Zeng et al., "A novel strategy to reconstruct NDVI time-series with high temporal resolution from MODIS multi-temporal composite products," *Remote Sens.*, vol. 13, no. 7, p. 1397, Apr. 2021, doi: [10.3390/rs13071397/S1](https://doi.org/10.3390/rs13071397/S1).
- [63] 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](https://doi.org/10.1016/j.rse.2015.03.018).
- [64] S. K. Padhee and S. Dutta, "Spatio-temporal reconstruction of MODIS NDVI by regional land surface phenology and harmonic analysis of time-series," *GIScience Remote Sens.*, vol. 56, no. 8, pp. 1261–1288, Nov. 2019, doi: [10.1080/15481603.2019.1646977](https://doi.org/10.1080/15481603.2019.1646977).



Xiangqian Li received the Ph.D. degree from Sun Yat-sen University, Zhuhai, China, in 2022.

He is currently working as a Post-Doctoral Researcher with Sun Yat-sen University. His research interests include spatiotemporal fusion and agriculture mapping.



Qiongyan Peng received the B.S. degree from the College of Life and Environmental Sciences, Minzu University of China, Beijing, China, in 2020, and the M.S. degree from the School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai, China, in 2023, where she is currently pursuing the Ph.D. degree.

Her research interests include agriculture mapping and spatiotemporal fusion.



Yi Zheng received the Ph.D. degree from Sun Yat-sen University, Zhuhai, China, in 2022.

She is currently a Post-Doctoral Researcher with Sun Yat-sen University. Her research interests include agriculture mapping and yield estimation and forest change and its climate effects.



Shangrong Lin received the B.S. degree in geographic information system from South China Normal University, Guangzhou, China, in 2015, and the Ph.D. degree in cartography and geographic information system from the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2020.

He is currently working as a Research Associate Professor with Sun Yat-sen University, Zhuhai, China. His research interests include remote sensing application of vegetation, high-spatial resolution remote sensing, gross primary production modeling, and plant photosynthesis.



Bin He received the M.S. degree from the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Beijing, China, in 2007, and the Ph.D. degree from the Institute of Geographic Sciences and Resources, Chinese Academy of Sciences, in 2010.

He is currently working as an Associate Professor with the Institute of Global Change and Earth System Science, Beijing Normal University, Beijing. His main research focuses on climate change and extreme climate events, remote sensing of ecological environment, and carbon cycling in terrestrial ecosystems.



Yuean Qiu received the M.S. degree from Beijing Normal University, Beijing, China, in 2022. He is currently pursuing the Ph.D. degree with Michigan State University, East Lansing, MI, USA.

His research interests include remote sensing product derivation and assessment, time series analysis, and forest cover change monitoring.



Jin Chen received the M.S. degree in natural geography from the Department of Geography, Beijing Normal University, Beijing, China, in 1992, and the Ph.D. degree in environmental engineering from the Department of Civil Engineering, Kyushu University, Fukuoka, Japan, in 2000.

He is currently working as a Professor with the State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Institute of Remote Sensing Science and Engineering, Beijing Normal University. His research interests include remote sensing and emergency management.



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 working as the Post-Doctoral Researcher with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include remote sensing time-series analysis and crop monitoring by using multisource data.



Wenping Yuan received the Ph.D. degree from the Institute of Botany, Chinese Academy of Sciences, Beijing, China, in 2007.

After graduation, he worked as a Post-Doctoral Researcher with the University of Oklahoma, Norman, OK, USA, and the Earth Resources Observation and Scientific Research Center, United States Geological Survey, Colorado-Denver, USA. After returning to China, he worked with Beijing Normal University, Beijing, and Sun Yat-sen University, Zhuhai, China, with a research focus on terrestrial ecosystem modeling. He is currently working as a Professor with the Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, International Research Center of Big Data for Sustainable Development Goals, School of Atmospheric Sciences, Sun Yat-sen University. His research interests include development and improvement of models, such as CO₂, CH₄, and N₂O, regional and national carbon source sink assessments, and estimation of vegetation productivity and crop yield.