## RESEARCH ARTICLE

# Effectiveness of Spatiotemporal Data Fusion in Fine-Scale Land Surface Phenology Monitoring: A Simulation Study

**Jiaqi Tian[1,2], Xiaolin Zhu[1*], Miaogen Shen[3], Jin Chen[3], Ruyin Cao[4], Yuean Qiu[3,5], and Yi Nam Xu[1]**

[1]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China. [2]Department of Geography, National University of Singapore, Singapore, Singapore. [3]State Key Laboratory of Earth Surface Processes and Resource Ecology, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China. [4]School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China. [5]Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI 48824, USA.

*Address correspondence to: xiaolin.zhu@polyu.edu.hk

Spatiotemporal data fusion technologies have been widely used for land surface phenology (LSP) monitoring since it is a low-cost solution to obtain fine-resolution satellite time series. However, the reliability of fused images is largely affected by land surface heterogeneity and input data. It is unclear whether data fusion can really benefit LSP studies at fine scales. To explore this research question, this study designed a sophisticated simulation experiment to quantify effectiveness of 2 representative data fusion algorithms, namely, pair-based Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) and time series-based Spatiotemporal fusion method to Simultaneously generate Full-length normalized difference vegetation Index Time series (SSFIT) by fusing Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) data in extracting pixel-wise spring phenology (i.e., the start of the growing season, SOS) and its spatial gradient and temporal variation. Our results reveal that: (a) STARFM can improve the accuracy of pixel-wise SOS by up to 74.47% and temporal variation by up to 59.13%, respectively, compared with only using Landsat images, but it can hardly improve the retrieval of spatial gradient. For SSFIT, the accuracy of pixel-wise SOS, spatial gradient, and temporal variation can be improved by up to 139.20%, 26.36%, and 162.30%, respectively; (b) the accuracy improvement introduced by fusion algorithms decreases with the number of available Landsat images per year, and it has a large variation with the same number of available Landsat images, and (c) this large variation is highly related to the temporal distributions of available Landsat images, suggesting that fusion algorithms can improve SOS accuracy only when cloud-free Landsat images cannot capture key vegetation growth period. This study calls for caution with the use of data fusion in LSP studies at fine scales.

## Introduction

Land surface phenology (LSP), i.e., the seasonality of vegetated land surface derived from remote sensing data, can provide more comprehensive spatiotemporal coverage of phenological metrics than plot-based and in situ observations [1,2]. The LSP metrics (e.g., spring phenology and autumn phenology) have been widely used to explore the effects of environmental variables (e.g., urban warming and climate change) on vegetation growth and dynamics [3–5]. A majority of existing LSP studies conventionally applied daily satellite images (e.g., Moderate Resolution Imaging Spectroradiometer [MODIS] and Advanced Very High Resolution Radiometer [AVHRR]) [6–8], but due to mixed pixel effects caused by coarse spatial resolutions ranging from 250 m to 8 km, their results may not be reliable, especially complex landscapes (e.g., urban areas and fragmented croplands). For example, the coarse-resolution satellite images can overestimate the urban–rural difference of spring phenology [9], and the spring phenological dates detected by coarse images are more controlled by the earlier spring phenological pixels at fine images [10,11]. On the contrary, Landsat imagery is the only data source for studying long-term LSP in complex landscapes thanks to its over 40-year archives and relatively high spatial resolution of 30 m [12]. However, Landsat-based LSP studies face a big challenge of data sparsity, since it has a 16-d revisit cycle and cloud-free images that are very limited in many regions [13]. Although some Landsat-based LSP studies reconstructed 16-d Landsat time series using multi-year composites

[14] or smoothing filters [15] before LSP extraction, the former cannot explore interannual variations of LSP while the latter may result in a false early spring phenology detection [15], and neither can address the issue of data sparsity directly.

To solve the sparsity problem of Landsat images, various spatiotemporal data fusion algorithms by fusing 16-d fine-resolution Landsat images with daily coarse-resolution satellite images (e.g., MODIS and AVHRR) were proposed and developed [16]. Based on the modeling strategies, these fusion algorithms can be generally divided into 2 major categories, pair-based (e.g., STARFM, ESTARFM, and Fit-FC) [17–19] and time series-based (e.g., SSFIT and STAIR) [20,21]. The pair-based fusion algorithms generate a target individual fused image by using one or 2 pairs of cloud-free fine- and coarse-resolution images as input data. Each pair of images is composed of one fine-resolution image and one coarse-resolution image acquired at the same time or a very close time. These pair-based fusion algorithms follow a similar principle that they model relationship between fine and coarse images from the pairs and then apply this relationship to predict fine images at the time there were no available fine-resolution images. STARFM, the most popular and representative pair-based fusion method, uses weight functions to model the relationship between fine- and coarse-resolution images. These pair-based fusion algorithms need to be implemented by multiple times to predict fine images at multiple time points to form time series data for LSP studies [9,10]. The possible limitations of applying pair-based fusion methods into LSP studies include their low efficiency and low accuracy if only a few cloud-free fine images are available, and they cannot fully utilize the information contained in fine and coarse images. On the contrary, time series-based fusion methods input all available fine- and coarse-resolution images including those images partly covered by clouds. They output cloud-free fine time series at given time intervals, such as every 16 d, instead of predicting a single fine image. Taking the recently developed SSFIT as an example, time series-based fusion methods generally use the temporal profile extracted from coarse time series of various land cover types to predict fine time series. Combining cloud-free Landsat images and Landsat-like images produced by either pair-based or time series-based fusion algorithms can construct dense vegetation index (VI) time series at 30 m and high frequency (e.g., 8 and 16 d) to support LSP retrieval.

In recent years, a large number of LSP studies employed fine-resolution data derived from spatiotemporal fusion algorithms. For example, the STARFM fusion algorithm was used for extracting crop phenology from 2001 to 2014 in central Iowa [22], characterizing phenology patterns in dryland forests in 2006 [23], investigating dryland vegetation phenology across an elevation gradient in Arizona, USA [24], improving the mapping of crop types using phenological features in the Midwestern USA [25], estimating phenology metrics of sub-Alpine forests in South Korea [26], supporting a phenology-based classification of paddy rice in Northeastern China [27], and exploring the scaling effects of LSP detection [10]. Nevertheless, several studies suggested that the performance of fusion algorithms is strongly related to the land surface heterogeneity and the input images [28,29]. However, it is unclear to what extent the fusion of Landsat and MODIS images can help LSP studies, and no study has explored this question.

To this end, this study aims to investigate the effectiveness of representative spatiotemporal data fusion algorithms in

LSP studies at fine scales. To achieve this objective, this study designed a sophisticated simulation experiment using 30-m cloud-free Harmonized Landsat Sentinel-2 (HLS) data to assess the performance of 2 representative fusion algorithms (i.e., classic pair-based STARFM and recently developed time series-based SSFIT) for spring phenology (i.e., the start of the growing season [SOS]) detection in a heterogeneous area. More specifically, we evaluated the accuracy improvement introduced by fusion algorithms in terms of pixel-wise SOS, spatial gradient, and temporal variation, under different cloud frequencies. It is worth noting that we do not intend to compare many fusion algorithms since there are more than 100 fusion algorithms published in the past 2 decades [29]. Our main objective in this study is to raise people's awareness of the real benefit brought by data fusion to phenology monitoring and provide an approach that can be used to evaluate the effectiveness of any fusion algorithms in phenology studies. The findings of this study could help users to select appropriate fusion algorithms in LSP studies and inspire future development and improvement for fusion algorithms.

## Materials and Methods

### Study area and data used
The study area (central coordinate: 41.26° N, 111.98° W) is in Ogden, Utah, USA (Fig. 1A). This area was selected because of the following criteria: (a) a less cloudy region can make it possible to collect dense cloud-free HLS time series as reference data for the simulation experiment [30]; (b) abundant land cover types including forest, grassland, shrubland, cropland, artificial surface, wetland and water body according to a global land cover product [31] (Fig. 1B) and (c) large spatial heterogeneity of spring phenology [6]. Therefore, the selected study area is an ideal location for conducting this simulation study.

The HLS V1.4 dataset covering the study area from 2018 to 2020 was used in this study. HLS data are 30-m high-frequency optical images combining Sentinel-2A, Sentinel-2B and Landsat 8 images [32,33], which reduces the cloud effects to the largest extent. The harmonized procedures for producing HLS data include atmospheric correction, geographic registration, normalization of solar and viewing angles and spectral bandpass adjustment between sensors [32]. By validation with ground observations, existing studies suggested that HLS data are promising for extracting LSP metrics [34,35].

### Experiment design
In this study, the modified Enhanced Vegetation Index (EVI2) derived from HLS time series was used to extract SOS. EVI2 is not sensitive to the background reflectance (e.g., bright soils and photosynthetically active vegetation), which can well capture the biophysical properties of the land surface and has been used by agencies to produce official vegetation phenology products (e.g., MCD12Q2 and VNP22Q2) [1,34]. Figure 2 shows the steps of the experiment in this study. First, to mimic Landsat time series with a 16-d temporal resolution, the original HLS data were processed to produce cloud-free 16-d reference Landsat EVI2 time series. Then, 2 typical spatiotemporal data fusion algorithms (i.e., STARFM and SSFIT) were used to fuse Landsat and MODIS EVI2 images by manipulating input data to simulate various cloud situations, respectively. Meanwhile, the linearly interpolated Landsat time series without fusing with MODIS was produced as a benchmark [36,37]. Lastly, the dates of SOS were extracted from 2 fused time series and
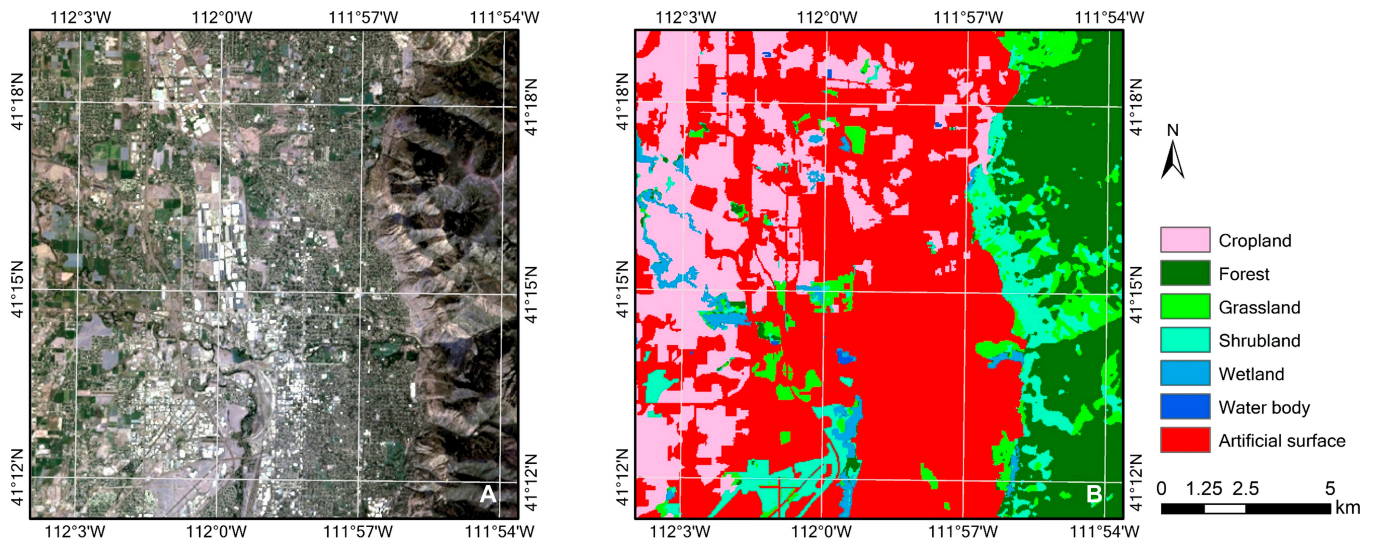
**Fig. 1.** True-color image and land cover map of the study area. (A) Landsat true-color image of the study area on 2020 September 16 and (B) its land cover map.
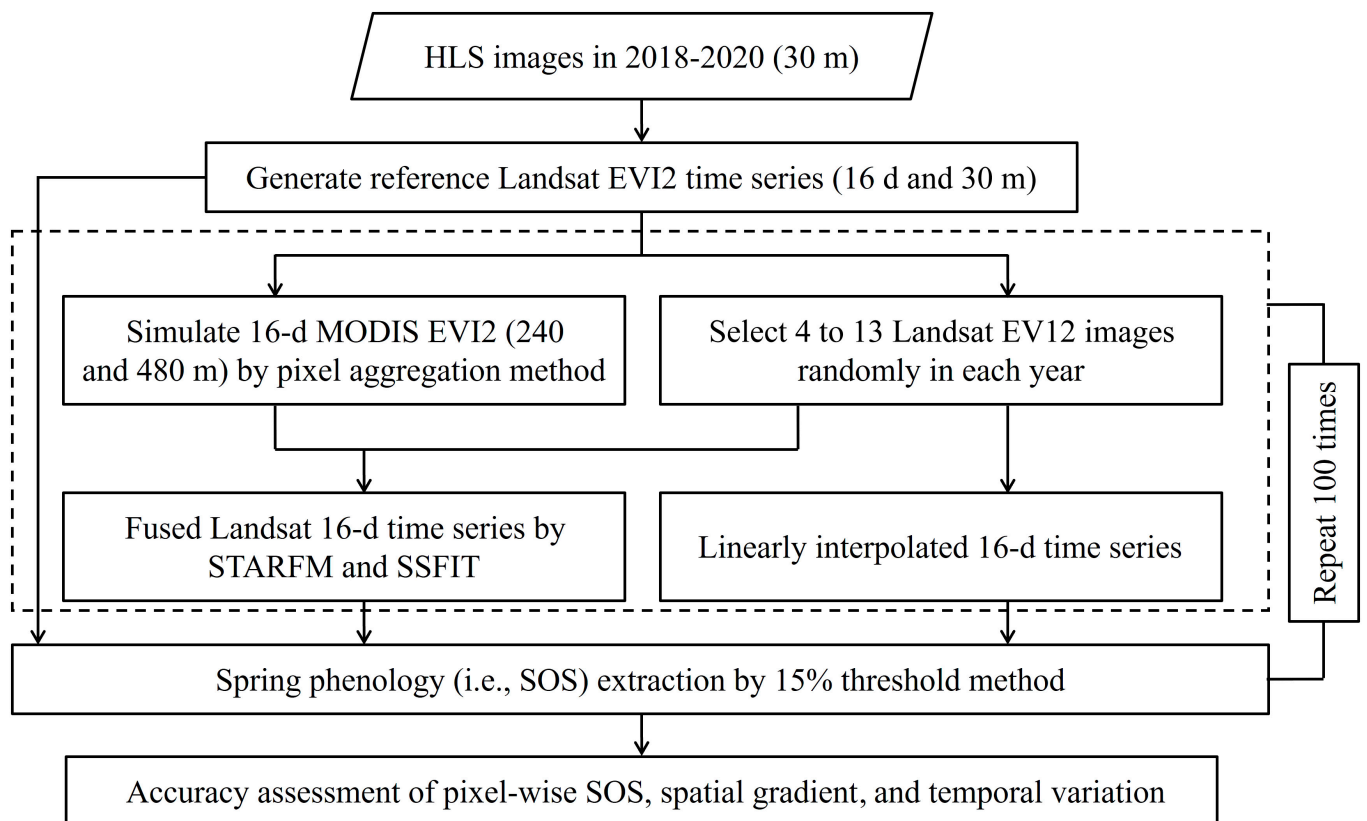


**Fig. 2.** Flowchart of assessing the performance of STARFM and SSFIT for SOS detection.

linearly interpolated time series for assessing the effectiveness of STARFM and SSFIT.

### Generation of reference Landsat EVI2 time series

Reference Landsat time series with a 16-d temporal resolution was derived from the original HLS dataset by a series of standardized processing, as shown in Fig. 3. First, we used the Automatic Time-Series Analysis, an automatic cloud and cloud shadow detection algorithm based on time-series images, to screen cloud and cloud shadow of the HLS images [38], since the cloud quality assurance (QA) band of the HLS dataset has considerable errors [34]. To refine clear observations, we followed the approaches adopted in [34], i.e., the MACCS-ATCOR Joint Algorithm cloud detection and 3-point despiking methods, to further eliminate the outliers namely the bright anomalies from undetected clouds and smoke and the negative
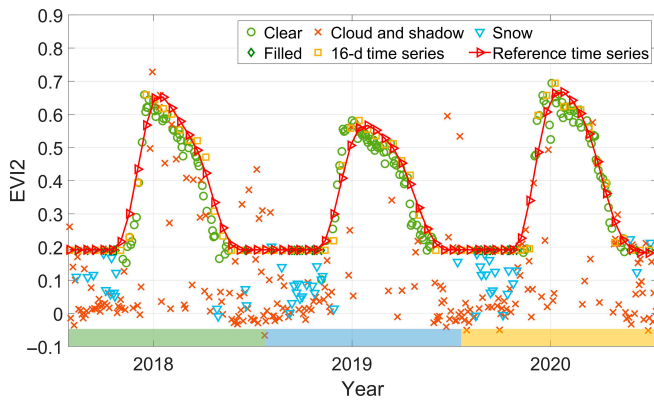
**Fig. 3.** Data preprocessing of HLS EVI2 time series for a pixel (central coordinate: 41.29° N, 111.91° W). The Reference time series shown in red means the smooth time series using the penalized cubic smoothing splines method.

anomalies from undetected cloud shadows of HLS time series. Second, we excluded snow-contaminated pixels identified by the HLS snow QA band and the Normalized Difference Forest Snow Index (NDFSI > 0.4) [34,39]. Snow-contaminated pixels were filled by a background value that refers to the fifth percentile of all snow-free EVI2 values from 2018 to 2020 at each pixel [34]. Third, the clear and filled observations together were used to produce a 16-d EVI2 time series by the maximum value composite (MVC) method. The MVC method is operated by outputting EVI2 maximum within a predetermined composite period (e.g., 16 d), which can further reduce the noises of time-series data [40,41]. Last, to generate a high-quality reference Landsat EVI2 time series (30 m and 16 d), the 16-d EVI2 composites were smoothed by the penalized cubic smoothing splines to eliminate noises and fill missing data [1,34], as red curve (i.e., namely Reference time series) shown in Fig. 3.

### Reconstruction of Landsat EVI2 time series by STARFM and SSFIT

To represent 2 types of fusion algorithms, STARFM, a classic pair-based fusion algorithm [17], and SSFIT, a recently developed time series-based fusion algorithm [21], were used to reconstruct fused Landsat-like images at a spatial resolution of 30 m, respectively.

1. STARFM

STARFM is a pair- and weight-based algorithm that the systematic error between 2 satellite sensors does not vary with time. Although STARFM was initially proposed and developed to fuse surface reflectance, its reliability and capability for fusing VIs (e.g., NDVI) have been verified [42]. Therefore, we used the EVI2 time series as input data for STARFM, and for upscaling a coarse-resolution EVI2 image to a fine image with the same spatial resolution and coordinate system, the fine-resolution EVI2 is:

$$F(x, y, t_k) = C(x, y, t_k) + \varepsilon_k \tag{1}$$

where $(x, y)$ is a pixel location, $t_k$ is the acquisition date for both fine- and coarse-resolution data, $C(x, y, t_k)$ means the coarse-upsampled EVI2 at date $t_k$, and $\varepsilon_k$ refers to the difference between 2 observed EVI2 values due to systematic bias [17]. Furthermore, STARFM assumes that the land cover type and system bias at

pixel $(x, y)$ do not change between the prediction date $t_0$ and the reference date $t_k$, namely $\varepsilon_0 = \varepsilon_k$, and therefore:

$$F(x, y, t_0) = C(x, y, t_0) + F(x, y, t_k) - C(x, y, t_k) \tag{2}$$

By integrating additional information from similar adjacent pixels within a searching window, the EVI2 of a fine-resolution pixel at date $t_0$ can be evaluated with a weight function [17], as Eq. 3:

$$F(x, y, t_0) = \tag{3}$$

$$\sum_{i=1}^{n_s} \sum_{k=1}^{n_k} W_i \times \left( C(x_i, y_i, t_0) + F(x_i, y_i, t_k) - C(x_i, y_i, t_k) \right)$$

where $(x_i, y_i)$ is the location of the $i_{th}$ similar adjacent pixels, $W_i$ represents its weight, $n_s$ is the total number of similar adjacent pixels, and $n_k$ is the number of reference dates [17]. The weight $W_i$ is determined by considering both the spatial distance between pixels and the spectral difference between the coarse- and fine-resolution images at $t_k$ [17].

2. SSFIT

SSFIT is a time series-based fusion algorithm that can produce VI time series with fine spatial resolution and frequent coverage at multiple dates simultaneously [21]. It builds a VI time series as a linear combination of multiple bases (i.e., basic components) and aims to determine the bases and coefficients:

$$F(x, y, t_0) = \sum_{c=1}^{n_b} a_c(x_i, y_i) \times b_c(x_i, y_i, t_0) + \varepsilon(x_i, y_i, t_0) \tag{4}$$

where $b_c(x_i, y_i, t_0)$ is the value of the cth base for the pixel located at $(x_i, y_i)$ at date $t_0$, $a_c(x_i, y_i)$ is its base coefficient, $\varepsilon$ is the model residual, and $n_b$ is the number of the bases [21]. Considering that the temporal patterns of time series are locally transferable between coarse and fine resolutions, SSFIT employs the singular value decomposition method to determine the bases from coarse images locally with a moving window, and the bases are then applied by the fine pixels within the window [21]. As for the base coefficients, they can be generated by typical curve fitting method. To guarantee enough sample points for curve fitting and the robustness of results, SSFIT uses simple time series-based and gap-filling processing to construct an initial time series $F_{GF}$ before fitting:

$$F_{GF}(x, y, t_0) =$$

$$C_R(x, y, t_0) + \frac{1}{n_k} \sum_{k=1}^{n_k} \left( F(x, y, t_k) - C_R(x, y, t_k) \right) \tag{5}$$

where $C_R(x, y, t_k)$ and $C_R(x, y, t_0)$ are EVI2 values at date $t_k$ and $t_0$ of the coarse reference time series, respectively, which is regarded as the most similar coarse-resolution time series to the observed fine-resolution time series [21]. The base coefficients are then produced by the weighted least squares, where the weight is derived from the similarity between the coarse reference time series and the observed fine time series:

$$a_c(x, y) = \tag{6}$$

$$\arg \min_{a_c} \sum_{j=1}^{n_t} \left( w(x, y, t_j) \times \left( F_{GF}(x, y, t_j) - \widehat{F}(x, y, t_j) \right) \right)^2$$

where $t_j$ is the $j_{th}$ date (either $t_0$ or $t_k$), $n_t$ is the length of the complete time series, and $\hat{F}(x, y, t_j)$ is the reconstructed EVI2 value by the linear combination using the extracted bases and the estimated coefficients [21]. With both the bases and the optimal coefficients, the entire EVI2 time series of the fine-resolution pixel is reconstructed by the linear combination. It should be noted that in STARFM, $t_k$ and $t_0$ represent a specific date, respectively, whereas in SSFIT, they represent multiple dates.

Considering that actual cloud conditions (i.e., frequency and seasonal distribution) vary with regions and the percent of global average cloud cover from Landsat-8 daytime observations is normally 40% to 50% [13,43]. We randomly selected the numbers of Landsat EVI2 images ranging from 4 to 13 (annual cloud cover from around 80% to 40%, as shown in Fig. 4) from the 16-d Landsat EVI2 time series as the input data for STARFM and SSFIT in this study. Figure 4 shows the DOY (i.e., day of year) of selected Landsat images (represented by black square dots) for different numbers of available Landsat images. To avoid the impact of confounding factors (e.g., registration error and observing geometry) on the study [42], the cloud-free MODIS EVI2 at 240 and 480 m required by STARFM and SSFIT were simulated by aggregating 16-d Landsat EVI2 time series instead of using real MODIS EVI2 time series, which can better explore the performance of data fusion algorithms exclusively. This strategy has been widely used by previous studies to compare the performance of different fusion algorithms [42,44], and it was verified that the experiments using simulated MODIS data could generally receive consistent results with real MODIS data [42]. The selected Landsat EVI2 images and MODIS EVI2 images were input into STARFM and SSFIT to produce Landsat-like 16-d EVI2 time series (named as fused time series). To represent different cloud conditions and obtain statistically meaningful results, the above simulation experiment was repeated 100 times (Figs. 2 and 4).

### Spring phenology detection and accuracy assessment

Linear interpolation of cloud-free observations is a common practice to generate time series for detecting vegetation phenology [36,37]. Therefore, to evaluate whether fusing Landsat with MODIS by fusion algorithms (i.e., STARFM and SSFIT) can benefit the LSP studies, the input Landsat EVI2 images used in each simulation were linearly interpolated to generate a 16-d time series as a benchmark (named as interpolated time series). As a result, we collected 3 types of time series, i.e., the reference, fused, and interpolated ones. Spring phenology (i.e., SOS) was extracted from the 3 types of time series for vegetation pixels in the study area. To focus on the vegetated pixels with observable phenology, the pixels meeting one of the following criteria were treated as nonvegetation (e.g., water and bare soil): the annual average EVI2 was less than 0.1 [45], the annual maximum EVI2 was less than 0.2 or less than 1.35 times of the annual minimum EVI2 [46], and the annual EVI2 maximum did not occur in summer seasons [46,47]. The dates of SOS were extracted by the 15% dynamic threshold method, in accordance with previous studies [1,34] and some official phenology products (e.g., MODIS MCD12Q2).
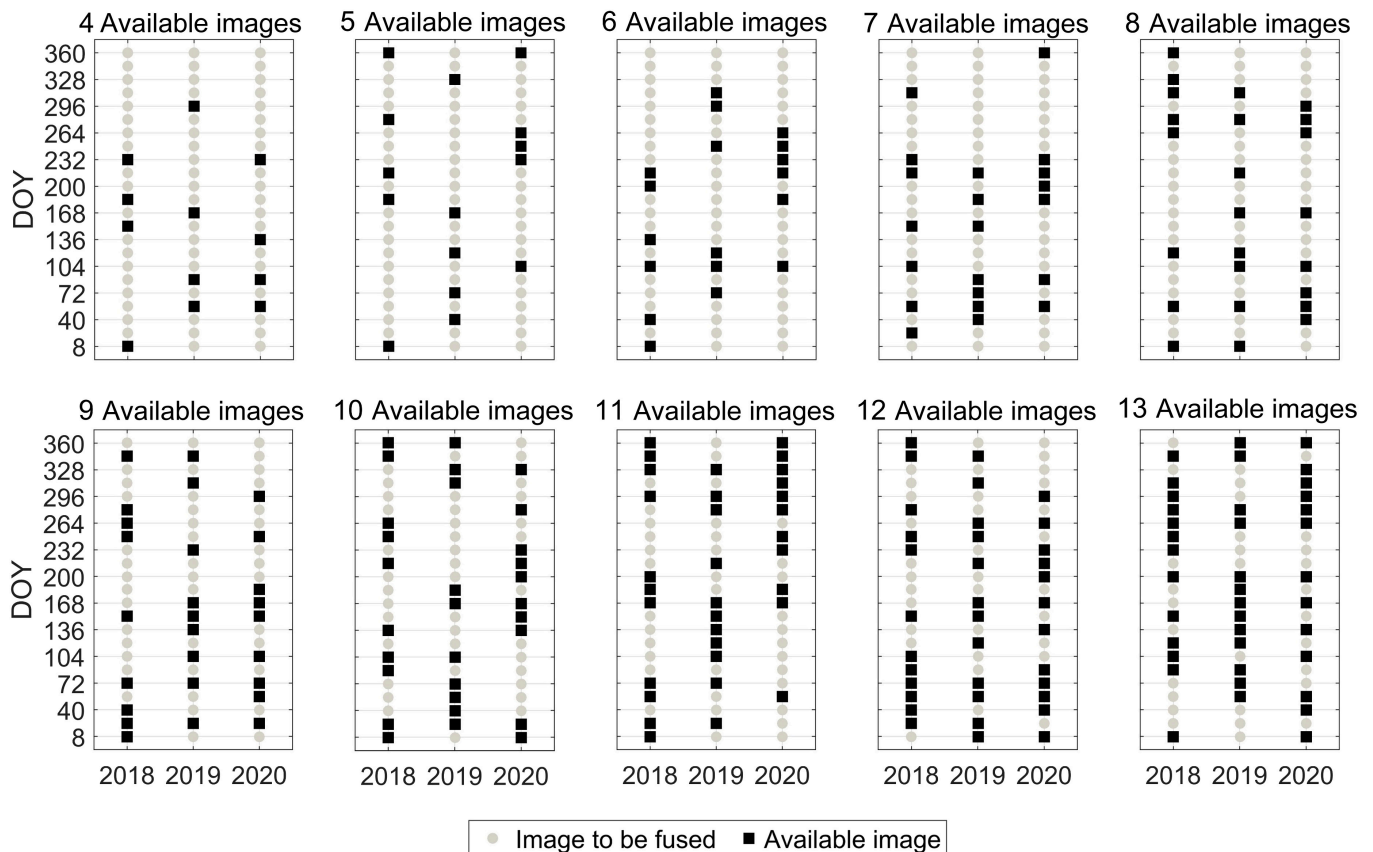
**Fig. 4.** Schematic diagram of selecting 4 to 13 HLS images randomly each year as input data of STARFM and SSFIT. DOY, day of year.

In addition, the threshold of 15% can produce highly similar phenology extraction results with other phenology extraction methods (e.g., curvature-based method) [41,46]. The threshold method has minimal manipulation on the original VI time series, while other methods, such as double logistic function fitting and spline fitting, smooth the raw VI time series to some extent which distracts our research question on investigating the performance of VI reconstruction by fusion.

The dates of SOS extracted from the fused and interpolated time series were compared with that from the reference time series for assessing the accuracy of fusion algorithms. In addition to the pixel-wise SOS, spatial gradient and temporal variation were also evaluated since they are the targets of many phenology studies [48]. For each vegetation pixel, its spatial gradient is calculated as the difference between itself and its eight neighboring pixels in a 3×3 window. Its temporal variation is computed as the anomaly of SOS values from 2018 to 2020. For each of the pixel-wise SOS, spatial gradient, and temporal variation, its pixel-level error is quantified by the difference between the estimated and reference values (i.e., estimated values minus reference values), and its image-level accuracy was quantified by a basic and commonly used index, root mean square error ($RMSE$):

$$RMSE = \sqrt{\frac{\sum_i^m (y_i - \hat{y}_i)^2}{m}} \qquad (7)$$

where $y_i$ and $\hat{y}_i$ are reference and estimated phenology variables of $i_{th}$ vegetation pixel, respectively, and $m$ is the total number of vegetation pixels. Moreover, we quantified the improvement of STARFM and SSFIT over the benchmark, $\Delta I$, according to Eq. 8:

$$\Delta I = \frac{RMSE_{interpolated} - RMSE_{fused}}{RMSE_{interpolated}} \times 100\% \qquad (8)$$

where $RMSE_{interpolated}$ and $RMSE_{fused}$ are $RMSE$ values of the benchmark and fused results respectively. Positive $\Delta I$ values indicate STARFM and SSFIT receiving better results than those of the benchmark.

To investigate the effects of the temporal distribution of input Landsat images in the growing period on SOS detection, we adopted an index of evenness to measure the temporal distributions of input Landsat images in the growing period [49]. Considering that observations around the start of a year and EVI2 maximum are more important for SOS detection, these observations were given higher weights in calculating the index of evenness. An example (e.g., 4 available images each year) of computing the evenness index is as below. In this case, we first assigned the weight of each position using a quadratic function (i.e., high weights to the first and last positions, while the low weights to the middle ones), as Fig. 5. We further calculated the evenness index ($E$) using Eq. 9 [49].

$$E = \frac{\sum min(ew_i, ow_i)}{\sum ew_i} \qquad (9)$$

where $ew_i$ is the expected weight (i.e., expected perfect evenness), and $ow_i$ is the observed weight. $E$ refers to the evenness index, ranging from 0 to 1. The $E$ values close to 1 indicate higher evenness. For example, in Fig. 5, the $E$ values of expected distribution, observed#1 and observed#2 are 1, 0.8, and 0.5, respectively.
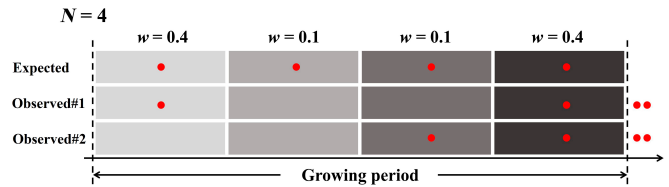


**Fig. 5.** Schematic diagram of calculating the evenness index for an example of 4 available images each year ($N = 4$). Red points mean the image distributions. $w$ refers to the weight of each available image.

## Results

### Spatial patterns of SOS errors

To visualize the pixel-level SOS errors in the study area, the spatial distributions and histograms of SOS errors when using 4, 8, and 12 Landsat images in one of the 100 simulations were shown in Fig. 6. We can see that the results of STARFM and SSFIT using MODIS images with 240 m have slightly higher accuracy than those of 480 m (i.e., more SOS errors close to 0 in the histograms). The errors of STARFM significantly decreased with the increasing number of input Landsat images, while the errors of SSFIT have muted changes for increasing numbers of input Landsat images (Fig. 6). In addition, the errors of fused results (i.e., STARFM and SSFIT) were smaller than those of interpolation when 4 input Landsat images were used (i.e., more dark-blue pixels in spatial patterns and wider error ranges in the histograms), but the errors of SOS derived from fused results were larger than those derived from interpolation when more input data were used (e.g., 8 and 12 Landsat images). Regarding the spatial distributions of SOS errors, the fused results have higher spatial heterogeneity than those interpolated ones, suggesting that the accuracy of time series reconstruction by data fusion algorithms varies pixel by pixel. It should be noted that Fig. 6 only represents one of the 100 simulations for exemplification, which cannot reach a general conclusion about the effectiveness of fused algorithms in LSP studies.

### Accuracy of pixel-wise SOS, spatial gradient, and temporal variation

To comprehensively evaluate the performance of data fusion algorithms, the $RMSE$ values of all 100 simulations and the improvement of fused results over the benchmark interpolation (i.e., $\Delta I$) were summarized in boxplots (Fig. 7). Regarding the median $RMSE$ values of all simulations, STARFM and SSFIT using 240-m MODIS images can produce slightly better results than using 480-m MODIS images. For pixel-wise SOS, both STARFM and SSFIT can obtain lower $RMSE$ (i.e., higher accuracy) than the benchmark interpolation when the number of input Landsat images was less than or equal to 6 (Fig. 7A and D), while for spatial gradient, only SSFIT with 4 input Landsat images can produce lower $RMSE$. Regardless of how many Landsat images were used as input, STARFM cannot achieve a lower $RMSE$ than those of interpolation (Fig. 7B and E). For temporal variation, both STARFM and SSFIT can produce SOS results of higher accuracy than those of benchmark when the number of input Landsat images was fewer than or equal to 6 and 7, respectively (Fig. 7C and F). The median $\Delta I$ values of all simulations in Fig. 7 suggested that with the use of 4 input Landsat images, 2 fusion algorithms can improve the accuracy
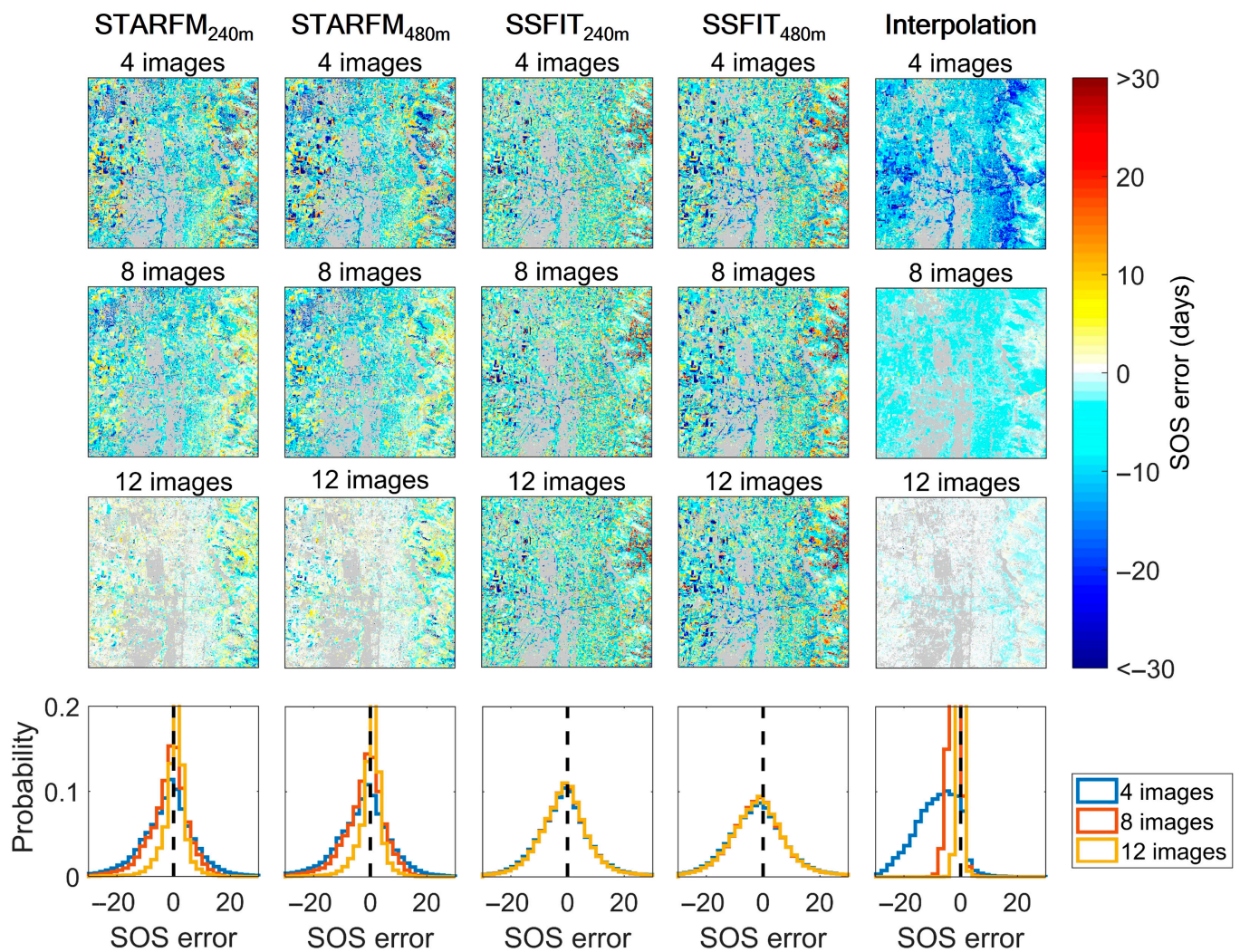
**Fig. 6.** The spatial patterns of SOS errors (estimated values minus reference values) of STARFM and SSFIT using 240-m MODIS images and 480-m MODIS images, and the linear interpolation with the use of 4, 8, and 12 Landsat images as input data in one simulation. The bottom panels are histograms of SOS errors. The temporal distributions of available Landsat images are 4 images (4th, 9th, 11th, and 20th), 8 images (4th, 6th, 9th, 12th, 14th, 17th, 18th, and 20th) and 12 images (1st, 2nd, 4th, 5th, 8th, 10th, 11th, 14th, 16th, 17th, 20th, and 22nd), respectively.

of pixel-wise SOS and temporal variation of the benchmark by up to 74.47% and 59.13% for STARFM and 139.2% and 162.3% for SSFIT, respectively. For spatial gradient, the median $\Delta I$ values of STARFM were less than zero regardless of the number of input Landsat images were used (Fig. 7H), but the accuracy of SSFIT can be increased by up to 26.36% when 4 Landsat images were used (Fig. 7K), suggesting that fusing Landsat with MODIS images by pair-based fusion algorithm (i.e., STARFM) cannot improve studies of the phenology spatial gradient but a time series-based algorithm (i.e., SSFIT) can.

It is worth noting that among the 100 simulations for each number of input Landsat images, the accuracy (i.e., *RMSE*) of STARFM shows large variations (e.g., *RMSE* ranging from 8 to 35 d for STARFM pixel-wise SOS when 4 input Landsat images were used), while the accuracy of SSFIT has minimum variations among 100 stimulations. This suggested that the accuracy of the pair-based fusion algorithm is sensitive to the distribution of input Landsat images, but the time series-based fusion algorithm is not (Fig. 7A to F). Nevertheless, as the accuracy and variations of benchmark interpolation increase with the

number of input data, the $\Delta I$ values (i.e., accuracy improvement) of STARFM and SSFIT also show large variations among 100 simulations for each number of input Landsat images (Fig. 7G to L), which implies that the accuracy improvement of STARFM and SSFIT are strongly affected by the temporal distribution of input Landsat images, which will be discussed in Discussion.

## Impacts of temporal distribution of image pairs

Since the detection of SOS primarily uses the EVI2 observations distributed in the growing period, it is reasonable to hypothesize that fusing Landsat with MODIS images by STARFM and SSFIT can benefit spring phenology studies when cloud-free Landsat images are rarely available in the growing period. To test this hypothesis, the numbers of Landsat images within the growing period (i.e., from the beginning of a year to the date of EVI2 maximum) were counted for simulations with the same amount of input Landsat images, and then they were divided into 2 groups, i.e., $\Delta I > 0$ (fusion better than interpolation) and $\Delta I < 0$ (fusion worse than interpolation), followed
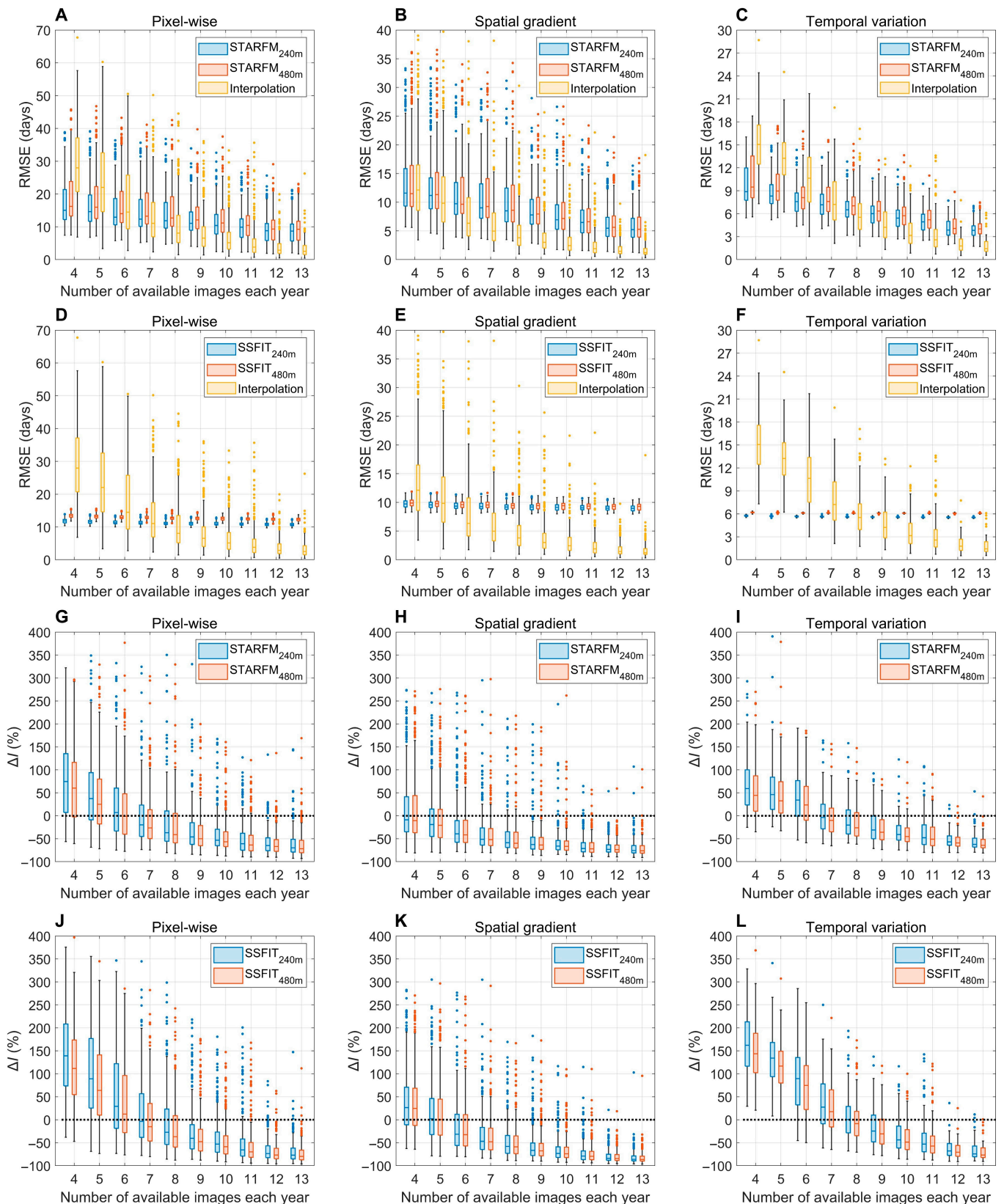
**Fig. 7.** Accuracy of phenology results. (A to F) Boxplots of *RMSE* values of phenology results by fused algorithms and interpolation in all 100 simulations. (G to L) Boxplots of Δ*I* values showing the improvement of STARFM and SSFIT over the benchmark interpolation using different numbers of input Landsat images.

by testing whether the number of images in growing period has a significant difference between 2 groups using a 2-sample $t$ test (Fig. 8). In general, both fusion algorithms (i.e., STARFM and SSFIT) showed that the cases of $\Delta I<0$ have significantly more input Landsat images in the growing period than cases of $\Delta I>0$ in terms of pixel-wise SOS, spatial gradient, and temporal variation. This suggests that fusion algorithms can improve the accuracy of SOS detection in terms of pixel-wise SOS, spatial gradient, and temporal variation even when fewer Landsat images are available during the growing period.

Although Fig. 8 indicated better performance of fusion algorithms when available Landsat images in the growing period are fewer, the larger overlap between 2 groups (i.e., $\Delta I>0$ and $\Delta I<0$) cannot be ignored. This potentially suggested that the number of available Landsat images in the growing period is not the only factor determining the effectiveness of fusion algorithms. Therefore, we further adopted an index of evenness [49] to investigate whether fused algorithms can greatly improve the phenology extraction accuracy when temporal distributions of Landsat images are uneven so that they cannot delineate the critical stages of vegetation growth (Fig. 9). The negative relationship (i.e., slope) between $\Delta I$ and evenness for pixel-wise

SOS, spatial gradient, and temporal variation confirmed that fusion algorithms have greater improvement when the temporal distribution of input Landsat images was more uneven (Fig. 9). This relationship became stronger when the total amount of available Landsat images decreased, thereby suggesting that fusing Landsat with MODIS images using both pair-based and time series-based fused algorithms can indeed improve phenology studies in heterogeneous areas where persistent clouds exist during the growing season, but the improvement is limited when there are sufficient Landsat images to capture the growing stages.

## Discussion

### Inspection of reconstructed EVI2 time series

The accuracy improvement (i.e., $\Delta I$) of SOS results derived from the STARFM and SSFIT over the benchmark interpolation is greatly related to the number of input Landsat images (i.e., the more input Landsat images, the lower accuracy improvement), and even the fusion algorithms fail to improve SOS accuracy when input Landsat images are sufficient (Fig. 7). As
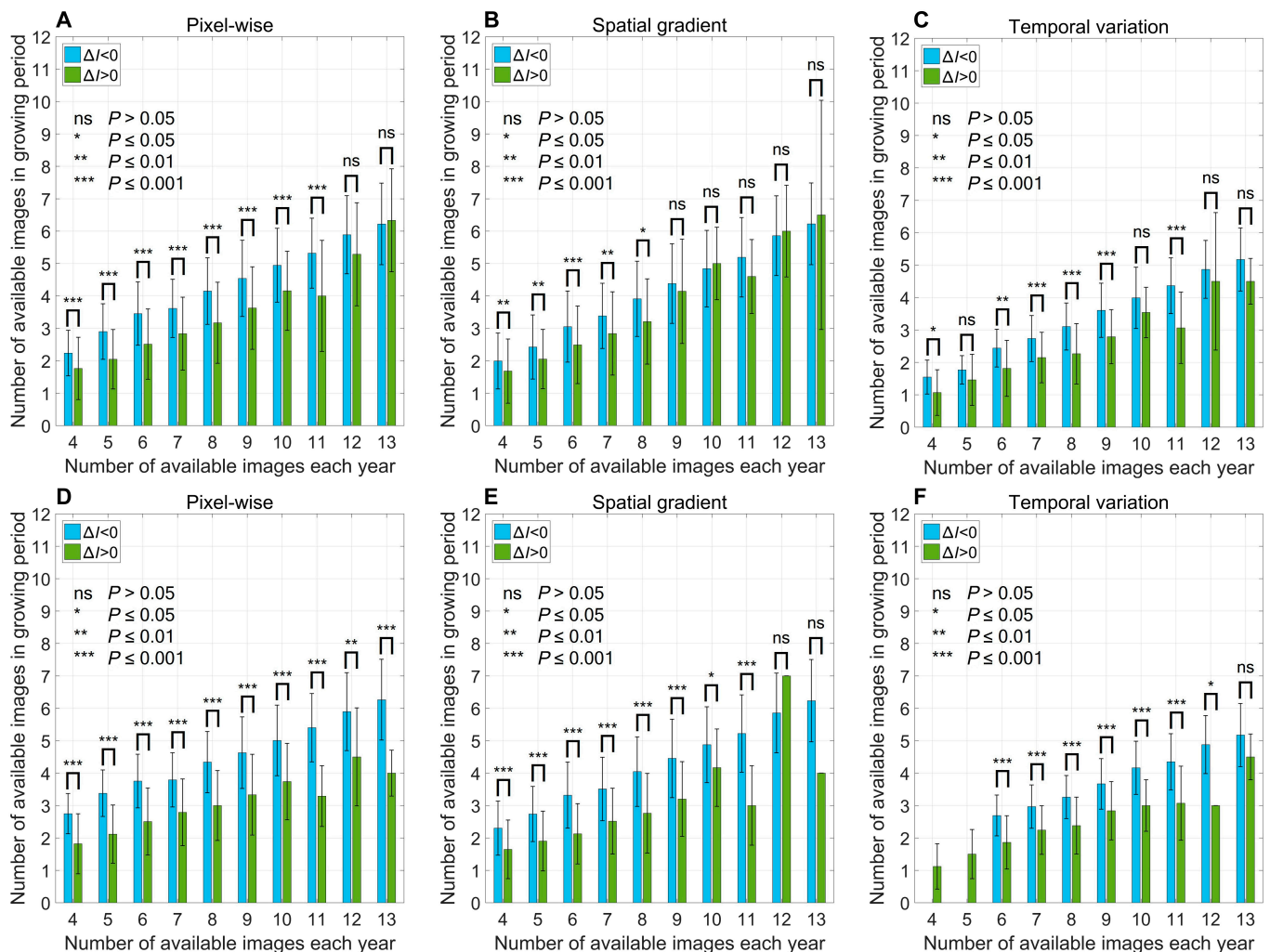
**Fig. 8.** Number of available Landsat images in the growing period for cases of $\Delta I<0$ and $\Delta I>0$. (A to C) STARFM and (D to F) SSFIT. Asterisk marks and ns indicate the significance of t test between the 2 groups, and the error bar is the standard deviation. ns, not significant.
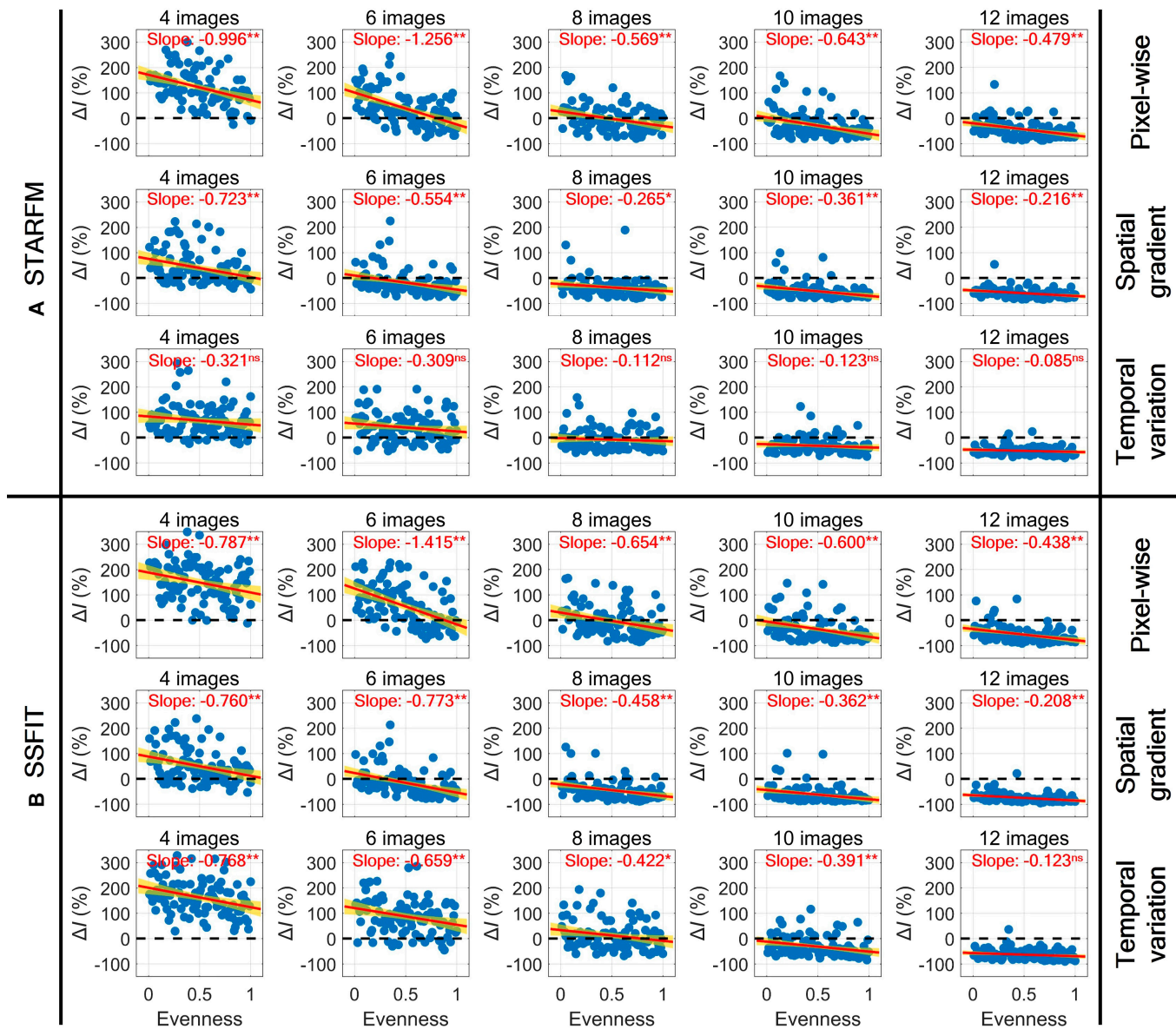
**Fig. 9.** The scatterplots between the improvement of fused algorithms over interpolation ($\Delta I$) and the evenness of temporal distribution of input Landsat images. (A) STARFM and (B) SSFIT. Red lines represent the linear regression of $\Delta I$ against the evenness, and yellow shades represent the 95% confidence interval. \*\*: $P$ value < 0.01, \*: $P$ value < 0.05. ns, not significant.

the SOS results are extracted from EVI2 time series (see Materials and Methods), it is reasonable to hypothesize that the accuracy of SOS results is linked with the accuracy of EVI2 time series reconstructed by interpolation, STARFM and SSFIT. Figure 10A to C shows the reference and 3 constructed EVI2 profiles for one pixel. It demonstrates that all 3 constructed EVI2 profiles are closer to the reference profile with an increased number of input Landsat images, but the interpolated one improves faster than the 2 fused ones. Based on this, we further investigated the impacts of the number of input Landsat images on the accuracy (i.e., RMSE) of reconstructed EVI2 time series, as shown in Fig. 10D. We found that interpolated EVI2 time series have higher accuracy than fused ones when the number of input Landsat images is sufficient (e.g., 7 well-distributed Landsat images), thereby suggesting that the improvement brought by the STARFM and SSFIT is highly dependent on the availability of Landsat images (Fig. 7).

## Impacts of resolution difference between fine and coarse images

For assessing SOS accuracy in terms of pixel-wise SOS, spatial gradient, and temporal variation, 2 fusion algorithms showed that 240-m MODIS images can produce slightly better results than using 480-m MODIS images (Figs. 6 and 7). We thus further quantified how much the accuracy of SOS can be improved using 240-m MODIS compared with 480-m MODIS (Fig. 11). It indicated that 240-m MODIS images can produce more accurate SOS results than those of 480-m MODIS images regardless of the number of input Landsat images used, which indicates that fusion algorithms can capture more spatial details and information from 240-m MODIS images for improving the accuracy of SOS detection. Meanwhile, for STARFM, the accuracy improvement brought by 240-m MODIS images decreased with the number of input Landsat images in terms of pixel-wise SOS, spatial gradient, and temporal variation, and vice versa
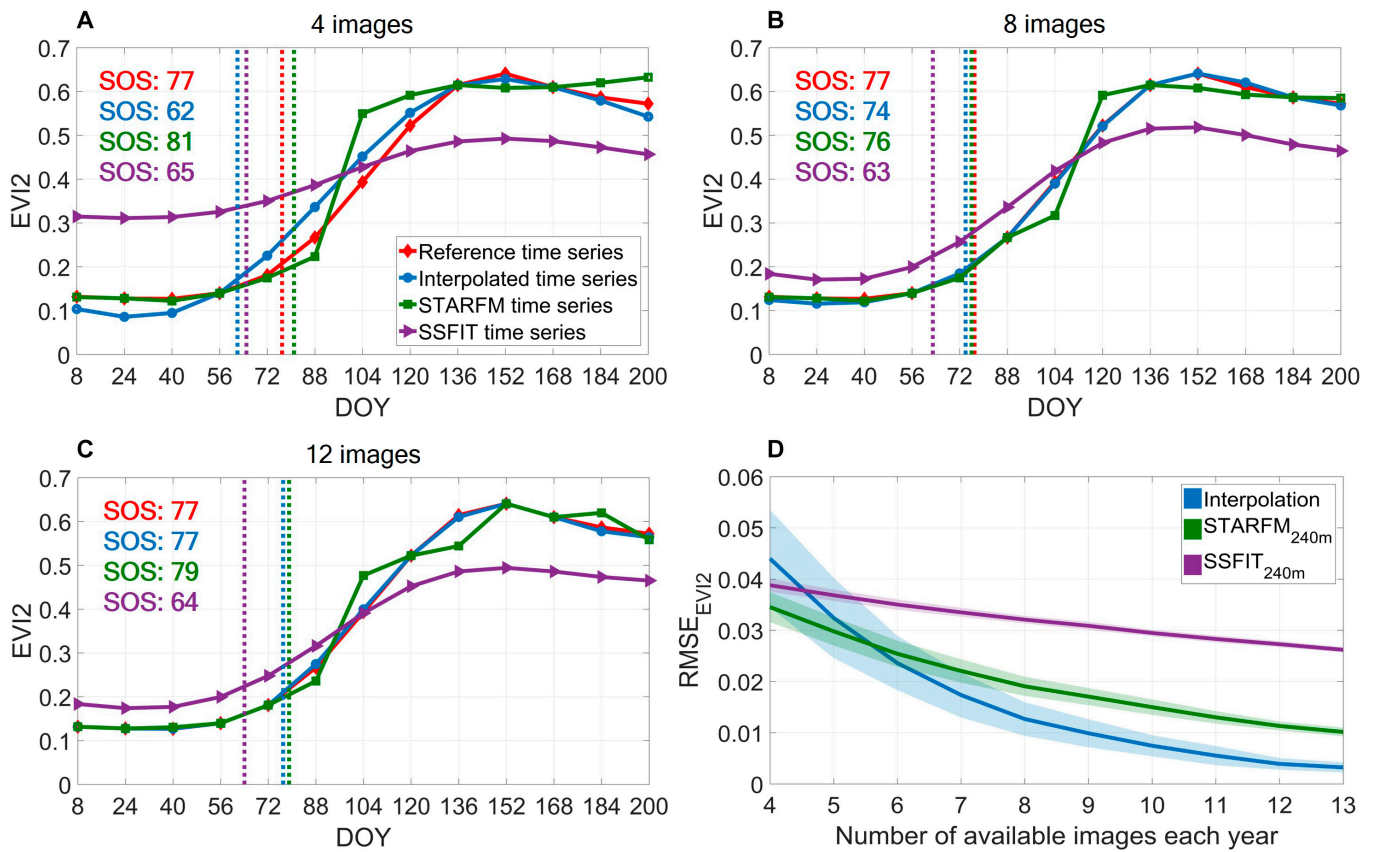
**Fig. 10.** Comparison of reference and reconstructed EVI2 time series of a pixel. (A to C) The use of 4, 8, and 12 Landsat images as input data in one simulation. (D) The accuracy (i.e., RMSE) of EVI2 time series reconstructed by interpolation, STARFM, and SSFIT in all 100 simulations.
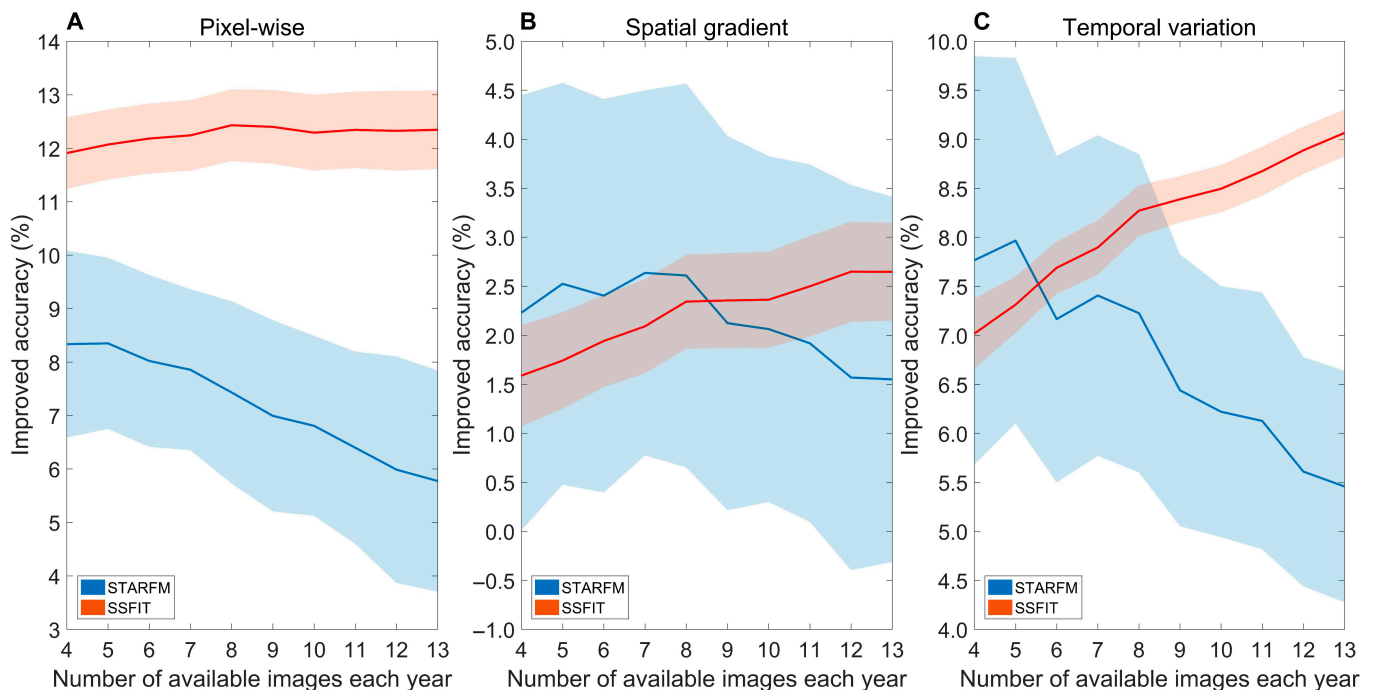


**Fig. 11.** Improved accuracy, i.e., $(RMSE_{480m} - RMSE_{240m})/RMSE_{480m}$, derived from 240-m SOS results and 480-m SOS results. (A) Pixel-wise, (B) spatial gradient, and (C) temporal variation. The solid lines refer to the mean, and shaded errors are 0.5*standard deviation.

**Fig. 13.** Comparison of the SOS accuracy (i.e., RMSE) extracted by fused images using real and simulated MODIS input data with the use of 4, 8, and 12 Landsat input images. (A) STARFM and (B) SSFIT. The error bars are one standard deviation.

images have high similarity (i.e., the more input Landsat images used, the higher SOS accuracy) for both fusion algorithms no matter what input images were used (i.e., real and simulated MODIS images). However, the SOS extraction results derived from real MODIS images have lower accuracy (i.e., higher RMSE) and greater variations (i.e., a wider range of standard deviation) than those of simulated MODIS images, which are primarily caused by the effects of confounding factors under real conditions of data acquisition.

## Implications and limitations

In recent years, an increasing number of LSP studies have used fused fine-resolution images derived from spatiotemporal data fusion algorithms [9,10,22,23]. However, several existing studies reported that the fused images are highly affected by the land surface heterogeneity and the input images [28,29]. Therefore, whether the fused fine-resolution images can be used to extract LSP accurately is still unknown. In this study, we explored the effectiveness of 2 typical data fusion algorithms (i.e., pair-based STARFM and time series-based SSFIT) for fine-scale spring phenology detection in terms of pixel-wise SOS, spatial gradient, and temporal variation, respectively (Fig. 7). We also investigated the effects of the number and temporal distribution of input Landsat images on the accuracy of SOS detection, respectively (Figs. 8 and 9), which may benefit future LSP studies at fine scales in complex landscapes. According to our results and findings, Table summarize the recommended strategies in different scenarios to serve as a guideline for future LSP studies.

In this study, there are several limitations. First, to exclude the effects of temporal resolutions on spring phenology detection, we unified the fused, interpolated and reference images into 16 d as same as Landsat archives, but satellite time series with shorter temporal intervals (e.g., 3, 5, and 8 d) have also been broadly applied into LSP studies [1,41,50]. Second, as spring phenology is a hot spot of current LSP studies [2,3,47],

**Table.** Recommended strategies in different scenarios. The severe, moderate, and mild cloud covers are around 80%, 60%, and 40%, respectively.

| Scenarios | Interpolation | STARFM | SSFIT |
|---|---|---|---|
| Severe cloud cover and bad data evenness in the growing season | × | √ | √ |
| Severe cloud cover and good data evenness in the growing season | × | √ | √ |
| Moderate cloud cover and bad data evenness in the growing season | × | × | √ |
| Moderate cloud cover and good data evenness in the growing season | × | √ | × |
| Mild cloud cover and bad data evenness in the growing season | √ | × | × |
| Mild cloud cover and good data evenness in the growing season | √ | × | × |

we only explored the effectiveness of fusion algorithms for spring phenology detection. However, our conclusions can be extended to other phenological metrics (e.g., autumn phenology). Third, in this study, we selected and tested the performance of 2 representative algorithms (i.e., classic pair-based STARFM and recently developed time series-based SSFIT) for spring
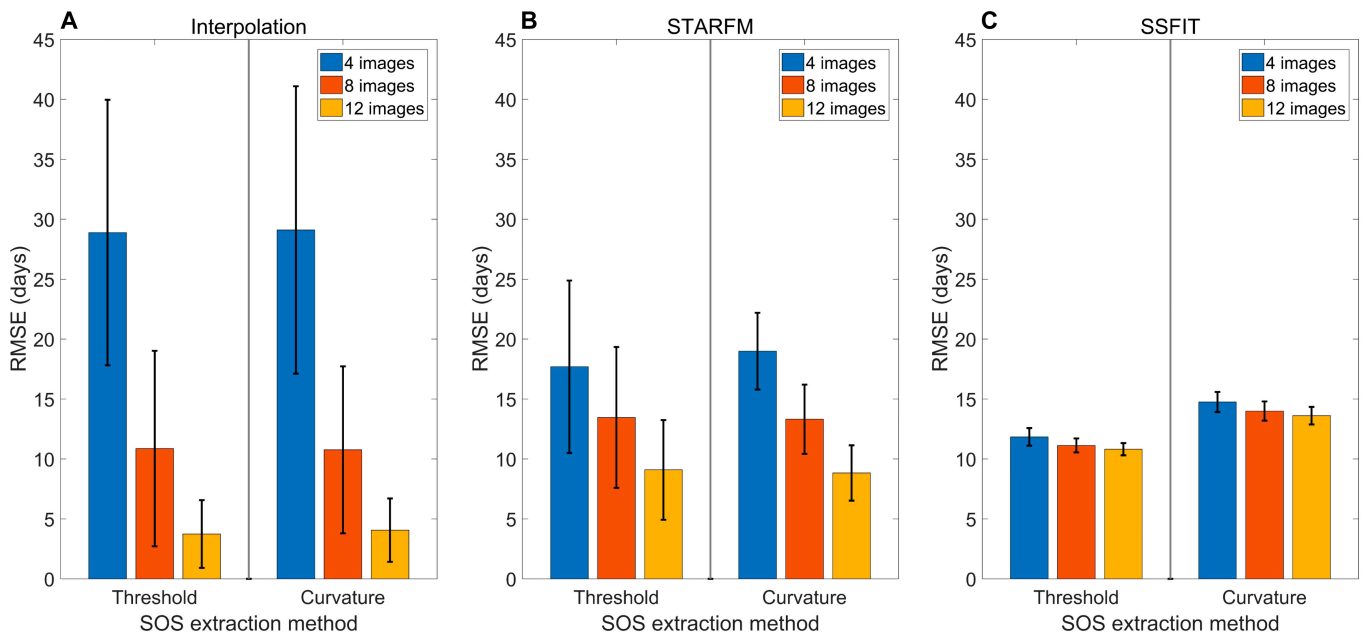
**Fig. 14.** Comparison of the SOS accuracy (i.e., RMSE) using threshold- and curvature-based extraction methods with the use of 4, 8, and 12 Landsat input images. (A) Interpolation, (B) STARFM, and (C) SSFIT. The error bars are one standard deviation.

phenology detection. According to the statistics of a literature review [29], there are more than 100 methodology papers regarding spatiotemporal data fusion published by 2021. Instead of including many spatiotemporal fusion algorithms into the comparison, this study provided a framework and an example to investigate the effectiveness of data fusion algorithms on LSP detection. Fourth, we primarily explored the effectiveness of data fusion algorithms in fine-scale LSP studies by the threshold-based phenology extraction method at a threshold of 15%, which is widely used to define spring phenology in plenty of LSP studies [1,41], and even official phenology products (e.g., MCD12Q2). Even so, we also compared the SOS accuracy (i.e., RMSE) extracted by threshold- and curvature-based methods (Fig. 14). The curvature-based method is used to determine spring phenology when the rate of change of curvature of the EVI2 time series reaches its local first maximum value [51]. Some previous studies have reported that the SOS extraction results derived from the threshold- and curvature-based methods have high consistency [15,41], and we further found that the accuracy gradients of SOS results derived from fused images are not affected by the selection of extraction methods (Fig. 14). We highly recommend future studies to assess the performance of more data fusion algorithms on more phenological metrics using more extraction methods, which can promote the applications of spatiotemporal fusion technologies and benefit some LSP studies (e.g., detecting phenological differences on both sides of the Qinling ridgeline).

## Conclusion

It is still unknown whether spatiotemporal fusion technologies can accurately capture spatial and temporal variations of LSP at fine scales. Accordingly, this study compared the spring phenology (i.e., SOS) extracted from four 16-d EVI2 time series datasets, i.e., cloud-free Landsat time series derived from HLS data (used as reference), interpolated time series from Landsat images with clouds (used as benchmark), and 2 fused time series by

STARFM and SSFIT. In light of our results, we found that both fusion algorithms (i.e., STARFM and SSFIT) can generally capture more accurate spring phenology information than the benchmark, in terms of pixel-wise SOS and temporal variation at fine scales when the amount of available Landsat images is very limited. However, for spatial gradient, STARFM can hardly obtain more accurate results than those of benchmark, and only SSFIT with 4 input Landsat images can improve the accuracy of SOS spatial gradient. Further investigation of the temporal distribution of Landsat images suggested that fusion algorithms can efficiently assist the LSP studies when available Landsat images cannot cover the critical phenological stages in the vegetation growing season. This is the first study to explicitly assess the effectiveness of spatiotemporal fusion technologies in LSP studies, even though they have been widely used in this field. Considering the cloud situation varying with regions significantly, our study provides guidance to future fine-scale LSP studies for applying spatiotemporal fusion in complex landscapes.

## Data Availability

The HLS data can be downloaded via: https://hls.gsfc.nasa.gov/hls-data/. The 30-m land cover product (GLC) can be accessed

from: https://datacore-gn.unepgrid.ch/geonetwork/srv/api/records/c67fa135-4360-463b-b2e6-b6d1def3a515. The MODIS MOD09GQ v061 can be found from https://lpdaac.usgs.gov/products/mod09gqv061/.

## References

1. Moon M, Zhang X, Henebry GM, Liu L, Gray JM, Melaas EK, Friedl MA. Long-term continuity in land surface phenology measurements: A comparative assessment of the MODIS land cover dynamics and VIIRS land surface phenology products. *Remote Sens Environ*. 2019;226:74–92.

2. White MA, De Beurs KM, Didan K, Inouye DW, Ricahrdson AD, Jensen OP, O'Keefe J, Zhang G, Nemani RR, Van Leeuwen WJD, et al. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Glob Change Biol*. 2009;15(10): 2335–2359.

3. Meng L, Mao J, Zhou Y, Richardson AD, Lee X, Thornton PE, Ricciuto DM, Li X, Dai Y, Shi X, et al. Urban warming advances spring phenology but reduces the response of phenology to temperature in the conterminous United States. *Proc Natl Acad Sci USA*. 2020;117(8):4228–4233.

4. Zhang X, Friedl MA, Schaaf CB, Strahler AH. Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data: CLIMATE CONTROLS ON VEGETATION PHENOLOGICAL PATTERNS. *Glob Change Biol*. 2004;10(7):1133–1145.

5. Tian J, Zhu X, Shen Z, Wu J, Xu S, Liang Z, Wang J. Investigating the urban-induced microclimate effects on winter wheat spring phenology using Sentinel-2 time series. *Agric For Meteorol*. 2020;294:108153.

6. Li X, Zhou Y, Asrar GR, Mao J, Li X, Li W. Response of vegetation phenology to urbanization in the conterminous United States. *Glob Change Biol*. 2017;23(7):2818–2830.

7. Gim H-J, Ho C-H, Jeong S, Kim J, Feng S, Hayes MJ. Improved mapping and change detection of the start of the crop growing season in the US Corn Belt from long-term AVHRR NDVI. *Agric For Meteorol*. 2020;294:108143.

8. Zhou D, Zhao S, Zhang L, Liu S. Remotely sensed assessment of urbanization effects on vegetation phenology in China's 32 major cities. *Remote Sens Environ*. 2016;176:272–281.

9. Tian J, Zhu X, Wu J, Shen M, Chen J. Coarse-resolution satellite images overestimate urbanization effects on vegetation spring phenology. *Remote Sens*. 2020;12(1):117.

10. Zhang X, Wang J, Gao F, Liu Y, Schaaf C, Friedl M, Yu Y, Jayavelu S, Gray J, Liu L, et al. Exploration of scaling effects on coarse resolution land surface phenology. *Remote Sens Environ*. 2017;190:318–330.

11. Chen X, Wang D, Chen J, Wang C, Shen M. The mixed pixel effect in land surface phenology: A simulation study. *Remote Sens Environ*. 2018;211:338–344.

12. Wulder MA, Loveland TR, Roy DP, Crawford CJ, Masek JG, Woodcock CE, Allen RG, Anderson MC, Belward AS, Cohen WB, et al. Current status of Landsat program, science, and applications. *Remote Sens Environ*. 2019;225:127–147.

13. Ju J, Roy DP. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sens Environ*. 2008;112(3):1196–1211.

14. Zhu X, Helmer EH, Gwenzi D, Collin M, Fleming S, Tian J, Marcano-Vega H, Meléndez-Ackerman EJ, Zimmerman JK. Characterization of dry-season phenology in tropical forests by reconstructing cloud-free Landsat time series. *Remote Sens*. 2021;13(23):4736.

15. Tian J, Zhu X, Wan L, Collin M. Impacts of satellite revisit frequency on spring phenology monitoring of deciduous broad-leaved forests based on vegetation index time series. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2021;14: 10500–10508.

16. Zhu X, Cai F, Tian J, Williams T. Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. *Remote Sens*. 2018;10(4):527.

17. Gao F, Masek J, Schwaller M, Hall F. On the blending of the Landsat and MODIS surface reflectance : Predicting daily Landsat surface reflectance. *IEEE Trans Geosci Remote Sens*. 2006;44(8):2207–2218.

18. Wang Q, Atkinson PM. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sens Environ*. 2018;204:31–42.

19. Zhu X, Chen J, Gao F, Chen X, Masek JG. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens Environ*. 2010;114(11):2610–2623.

20. Luo Y, Guan K, Peng J. 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*. 2018;214:87–99.

21. Qiu Y, Zhou J, Chen J, Chen X. Spatiotemporal fusion method to simultaneously generate full-length normalized difference vegetation index time series (SSFIT). *Int J Appl Earth Obs Geoinfo*. 2021;100:102333.

22. Gao F, Anderson MC, Zhang X, Yang Z, Alfieri JG, Kustas WP, Mueller R, Johnson DM, Prueger JH. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens Environ*. 2017;188:9–25.

23. Walker JJ, de Beurs KM, Wynne RH, Gao F. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sens Environ*. 2012;117: 381–393.

24. Walker JJ, de Beurs KM, Wynne RH. Dryland vegetation phenology across an elevation gradient in Arizona, USA, investigated with fused MODIS and Landsat data. *Remote Sens Environ*. 2014;144:85–97.

25. Zhu L, Radeloff VC, Ives AR. Improving the mapping of crop types in the Midwestern U.S. by fusing Landsat and MODIS satellite data. *Int. J. Appl. Earth Obs. Geoinfo*. 2017;58:1–11.

26. Park S-J, Jeong SG, Park Y, Kim SH, Lee DK, Mo YW, Jang DS, Park KM. Phenological analysis of sub-Alpine Forest on Jeju Island, South Korea, using data fusion of Landsat and MODIS products. *Forests*. 2021;12(3):286.

27. Yin Q, Liu M, Cheng J, Ke Y, Chen X. Mapping Paddy Rice planting area in northeastern China using spatiotemporal data fusion and phenology-based method. *Remote Sens*. 2019;11(14):1699.

28. Emelyanova IV, McVicar TR, Van Niel TG, Li LT, van Dijk AIJM. Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sens Environ*. 2013;133:193–209.

29. Zhu X, Zhan W, Zhou J, Chen X, Liang Z, Xu S, Chen J. A novel framework to assess all-round performances of spatiotemporal fusion models. *Remote Sens Environ*. 2022;274:113002.

30. Wilson AM, Jetz W. Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS Biol*. 2016;14(3):e1002415.

31. Chen J, Chen J, Liao A, Cao X, Chen L, Chen X, He C, Han G, Peng S, Lu M, et al. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS J Photogramm Remote Sens*. 2015;103:7–27.

32. Claverie M, Ju J, Masek JG, Dungan JL, Vermote EF, Roger JC, Skakun SV, Justice C. The harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens Environ*. 2018;219:145–161.

33. Li J, Roy D. A global analysis of sentinel-2A, sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens*. 2017;9(9):902.

34. Bolton DK, Gray JM, Melaas EK, Moon M, Eklundh L, Friedl MA. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sens Environ*. 2020;240:111685.

35. Zhang Y, Parazoo NC, Williams AP, Zhou S, Gentine P. Large and projected strengthening moisture limitation on end-of-season photosynthesis. *Proc Natl Acad Sci USA*. 2020;117(17):9216–9222.

36. Li X, Zhu W, Xie Z, Zhan P, Huang X, Sun L, Duan Z. Assessing the effects of time interpolation of NDVI composites on phenology trend estimation. *Remote Sens*. 2021;13(24):5018.

37. Zhu W, Pan Y, He H, Wang L, Mou M, Liu J. A changing-weight filter method for reconstructing a high-quality NDVI time series to preserve the integrity of vegetation phenology. *IEEE Trans Geosci Remote Sens*. 2012;50(4):1085–1094.

38. Zhu X, Helmer EH. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. *Remote Sens Environ*. 2018;214:135–153.

39. Wang X-Y, Wang J, Jiang Z-Y, Li H-Y, Hao X-H. An effective method for snow-cover mapping of dense coniferous forests in the upper Heihe River basin using Landsat operational land imager data. *Remote Sens*. 2015;7(12):17246–17257.

40. Holben BN. Characteristics of maximum-value composite images from temporal AVHRR data. *Int J Remote Sens*. 1986;7(11):1417–1434.

41. Tian J, Zhu X, Chen J, Wang C, Shen M, Yang W, Tan X, Xu S, Li Z. Improving the accuracy of spring phenology detection by optimally smoothing satellite vegetation index time series based on local cloud frequency. *ISPRS J Photogramm Remote Sens*. 2021;180:29–44.

42. Zhou J, Chen J, Chen X, Zhu X, Qiu Y, Song H, Rao Y, Zhang C, Cao X, Cui X. Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction. *Remote Sens Environ*. 2021;252:112130.

43. Li J, Chen B. Global revisit interval analysis of Landsat-8 -9 and sentinel-2A -2B data for terrestrial monitoring. *Sensors*. 2020;20(22):6631.

44. Xie D, Gao F, Sun L, Anderson M. Improving spatial-temporal data fusion by choosing optimal input image pairs. *Remote Sens*. 2018;10(7):1142.

45. Shen M, Jiang N, Peng D, Rao Y, Huang Y, Fu YH, Yang W, Zhu X, Cao R, Chen X, et al. Can changes in autumn phenology facilitate earlier green-up date of northern vegetation? *Agric For Meteorol*. 2020;291:108077.

46. Wang C, Zhu K. Misestimation of growing season length due to inaccurate construction of satellite vegetation index time series. *IEEE Geosci Remote Sens Lett*. 2019;16(8):1185–1189.

47. Shen M, Zhang G, Cong N, Wang S, Kong W, Piao S. Increasing altitudinal gradient of spring vegetation phenology during the last decade on the Qinghai–Tibetan plateau. *Agric For Meteorol*. 2014;189–190:71–80.

48. Ma X, Zhu X, Xie Q, Jin J, Zhou Y, Luo Y, Liu Y, Tian J, Zhao Y. Monitoring nature's calendar from space: Emerging topics in land surface phenology and associated opportunities for science applications. *Glob Change Biol*. 2022;28(24):7186–7204.

49. Bulla L. An index of evenness and its associated diversity measure. *Oikos*. 1994;70(1):167–171.

50. Zhang X, Liu L, Liu Y, Jayavelu S, Wang J, Moon M, Henebry GM, Friedl MA, Schaaf CB. Generation and evaluation of the VIIRS land surface phenology product. *Remote Sens Environ*. 2018;216:212–229.

51. Zhang X, Friedl MA, Schaaf CB, Strahler AH, Hodges JCF, Gao F, Reed BC, Huete A. Monitoring vegetation phenology using MODIS. *Remote Sens Environ*. 2003;84(3):471–475.