Remote Sensing of Environment 276 (2022) 113046

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Spatial-aware SAR-optical time-series deep integration for crop phenology tracking

Wenzhi Zhao [a,b,*], Yang Qu [c], Liqiang Zhang [a,b,*], Kaiyuan Li [a,b]

[a] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, PR China
[b] Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, PR China
[c] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, PR China

ARTICLE INFO

Editor: Marie Weiss

Keywords:
Optical-SAR time-series
Deep learning
Crop phenology

ABSTRACT

Accurate crop phenology information is essential for precision farming and agricultural productivity improvement. In recent years, in-situ equipment on crop phenology observation has been boosted, which generates high-quality real-time pictures on capturing vegetation phenological changes. However, due to the limited number of ground sites, it is impossible to measure large-scale crop phenology with local observations. Complementary, the freely available Sentinel satellites with high revisit frequency provide an opportunity to map accurate crop phenology at an unprecedented fine spatial scale. Because of the differences in viewing angle and range, the consistency of crop phenological stages varies between satellite and ground observations. To fill the gap between satellite and ground observations, we developed a spatial-aware scheme to integrate SAR and optical time-series data for accurate crop phenology tracking. To be specific, we propose a new deep learning model called Deep-CroP framework to improve the alignment between satellite and ground observations on crop phenology. The experiment results on selected ground sites demonstrate that the proposed Deep-CroP is able to accurately identify crops phenology and narrow the discrepancies from 30+ days to as high as several days. In addition, we applied the Deep-CroP to large-scale Sentinel time-series to map spatial patterns of phenology at fine resolution imagery on two study areas (i.e., TA1 and TA2). In general, the potential of satellites time-series for ground-level crop phenology observation is verified. Also, the consistency between satellite and PhenoCam observations is expected to be further improved.

1. Introduction

Crop phenology, also known as crop growth stages, refers to the biophysical development of crop plants from planting to harvest. Phenological parameters are the key indicators for dynamic crop monitoring (Richardson et al., 2013). Accurate crop phenology information retrieval is important for precision farming (Gao et al., 2017; Jentsch et al., 2009), crop yields estimation (Yuan et al., 2016), and agricultural productivity improvement (Jung et al., 2021; Thenkabail et al., 2010; Weiss et al., 2020). In this context, the efficient crop phenology extraction method is urgently needed.

The conventional approach for crop phenology information extraction is to record the timing of specific changing events through frequent observations by well-trained personnel (Schnelle and Volkert, 1964). An alternative approach is to use fixed-position digital cameras with a high temporal resolution that repeatedly captures phenology information over a given area by multiple times per day. This type of digital camera has been widely implemented in phenology observation projects, including the European Phenology Network (Wingate et al., 2015), the Phenological Eyes Network (Nasahara and Nagai, 2015), and PhenoCam Network (Klosterman et al., 2014). Among them, the PhenoCam Network, with hundreds of web-enabled digital cameras, is designed to capture time-lapse photography on vegetation phenological changes (such as green-up and senescence) with various landscapes across North America and Europe since 2000 (Richardson et al., 2009). The dense time-series Green Chromatic Coordinate (GCC) data derived from daily

* Corresponding authors at: State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, PR China.
E-mail address: wenzhi.zhao@bnu.edu.cn (W. Zhao).

https://doi.org/10.1016/j.rse.2022.113046
Received 29 December 2021; Received in revised form 7 April 2022; Accepted 11 April 2022
Available online 23 April 2022

PhenoCam images is one of the most reliable indicators for local crop phenology modeling and phenological parameter validation (Richardson et al., 2018). Currently, several software applications and packages have been developed to facilitate the extraction and processing of data from PhenoCam imagery, including xROI and Phenocamr (Hufkens et al., 2018; Seyednasrollah et al., 2019). Based on PhenoCam observations, it is now possible to establish site-based crop phenology models with localized threshold and key parameter settings. Still, the potential of large-scale crop phenology prediction with scattered ground sites remains unexploited.

Remote sensing technology has the advantages of large-scale monitoring, low cost, with short revisit periods. It provides a cost-effective and reliable approach for crop dynamic monitoring (Jiao et al., 2014; Qu et al., 2020). Multi-temporal remote sensing data can be used to derive phenology descriptors of vegetation growth dynamics termed 'land surface phenology' (LSP) (De Beurs and Henebry, 2005; Ganguly et al., 2010). The LSP parameters are typically associated with Spatio-temporal changes from satellite images on vegetated land surfaces, such as the start of greening/season (SOS) and the onset of senescence or end of the season (EOS). Traditionally, remote sensing data-based phenology studies primarily rely on medium- to coarse-resolution (250 m to 8 km) optical imagery (Justice et al., 2002; Justice et al., 1985), including the Advanced Very High-Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectrometer (MODIS). These medium-resolution sensors can effectively capture large-scale and even global LSP parameters, but low spatial resolution mixed crop fields with complex backgrounds thus may not properly represent the actual crop phenology. Meanwhile, the new generation earth observation satellites provide higher spatial resolution imagery with shorter revisit intervals such as Sentinel-2 (S2), which ignited further study on fine-scale crop phenology mapping. Now, it is possible to resolve individual farmland parcels from finer-scale satellite data and estimate the crop-specific phenology.

Recently, satellite-based LSP researches are based on the dense time-series data acquired by the S2 satellite (Marzialetti et al., 2019; Misra et al., 2020) or the multi-optical harmonized Landsat 8 and S2 time-series (Bolton et al., 2020; Burke and Rundquist, 2021; Zhou et al., 2019), which obtained satisfying results on large-scale phenology tracking. However, some studies demonstrated that the phenological parameters estimates from the digital camera are quite different from those observed by satellite. For example, Vrieling et al. (2018) found that EOS estimates from camera GCC series are on average almost two months ahead of NDVI-based estimates. The main reason is due to differences in viewing frequency and viewing coverage area between the satellite and in-situ cameras. To tackle the above challenges, for one thing, how to combine multiple satellite observations with different viewing dimensions to boost crop observation frequency is urgently needed; for another, how to alleviate the mismatch of spatial coverage between satellite and ground observations (i.e., in-situ camera with few hundreds detection range while remote sensing image with 10 m individual pixels).

For the purpose of boosting the viewing dimension of satellite data, a possible solution is to provide synthetic aperture radar (SAR) imagery as additional information for crop phenology monitoring. Since SAR backscatter is significantly related to the vegetation biomass, thus it is sensitive to vegetation structure and ground conditions which also correlated with crop phenological indicators (McNairn and Brisco, 2004). The recent freely accessible Sentinel-1 (S1) provides both high revisit frequency and fine-scale spatial resolution imagery, which boosts the research of dynamic agricultural monitoring with SAR time-series data, such as crop type mapping (Ajadi et al., 2021; Inglada et al., 2016). The integrated backscatter signal in S1 imagery is the combination of the backscatter directly reflected from the canopy, the ground backscatter attenuated by the canopy layer, and the vegetation-ground interaction induced backscatters (Ulaby, 1982). These three scattering components are related to vegetation structure, water content, the physical properties of the surface (such as soil moisture, surface roughness, and terrain topography) (Ferrazzoli et al., 1992), respectively. In the case of crop monitoring, the relative importance of these three scattering components depends on the phenological stages. To be specific, the backscattering from the ground dominates at the early and late crop phenology stages, and the significance of vegetation backscatter varies during crop development. Based on the evolving characteristics of backscatter variation, S1 has been used to detect crop phenology stages. In this scope, intensive studies have focused on fusing optical and SAR data to monitor crop phenology. Veloso et al. (2017) show the potential of using S1 data in crop growth monitoring and verified the temporal correlation between S1 and S2 by a thorough analysis of VV and VH, CR, and NDVI temporal profile. Although this study highlighted the correlation between S1 and S2 in the temporal domain, the question of integrating SAR and optical data to predict crop phenology stages remains unsolved. Stendardi et al. (2019) analyzed the correlation of S1 VH backscatter with respect to S2 NDVI in alpine meadows area and compared the retrieved LSP parameters extracted from both sources. Jin et al. (2015) focus on the joint utilization of vegetation indices derived from the Huanjing-1A/B optical satellite and polarimetric indicators derived from RADARSAT-2 to estimate the LAI and biomass of winter wheat. The highest correlations can be found when jointly utilizing the optical and SAR data. Mercier et al. (2020) evaluate the potential S1 and S2 data to retrieve wheat and rapeseed phenological stages with a classification approach over a limited set of fields in northern France. The study reveals that the joint use of S1 and S2 extracted features improved the accuracy in identifying crop phenology compared with the single utilization of S2 data. Meanwhile, Harfenmeister et al. (2021) also proven that correlation between dual-polarimetric parameters and crop biophysical parameters can be formulated. In general, most studies aim to stack SAR and optical data for crop phenology identification directly. Although the overall correlation of S1 and S2 time-series features in crop growth trend and turning point can be recognized, still, the extracted features (such as VV, NDVI) demonstrated a weak correlation in the temporal domain. In other words, instead of the simple linear correlations, the SAR-optical time-series share a complex non-linear response on crop phenology in the temporal domain. Thus, the direct feature stacking failed to explore the non-linear complementary relationship between the two data (Ienco et al., 2019). The precise formulation of such a non-linear relationship is becoming the bottleneck of SAR-optical satellite data utilization and crop phenology tracking.

Over the past few years, deep learning models have achieved significant progress in a variety of remote sensing tasks. Among them, the most commonly used deep learning models are Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which can be applied to extract spatial and temporal representative features from satellite imagery, respectively (Huang et al., 2017; Mou et al., 2018; Zhao et al., 2016). Meanwhile, several deep learning methods have been proposed to establish the non-linear relationships between multi-source satellite data fusion. For example, Shao et al. (2019) developed an extended super-resolution convolutional neural network (ESRCNN) that integrates the deep spatial features of Landsat-8 and S2 imagery to improve image quality. Although CNNs are efficient in extracting robust spatial features, it also requires accurate extraction of contextual features in the temporal dimension for the task of phenological parameter retrieval. In contrast, RNN is more suitable for temporal feature extraction, and it has been proven (Lyu et al., 2016; Rußwurm and Korner, 2017) effectively change point identification with respect to remote sensing sequence imagery. Nevertheless, crop phenological patterns often present with long-range correlations with a similar trend can be observed from both SAR and optical time-series data. But RNNs are usually with weak ability in long-range feature formulation. Recently, the transformer demonstrates its superiority in the long-range contextual feature capturing with the help of a multi-head attention mechanism (Dosovitskiy et al., 2020). Compared with the conventional

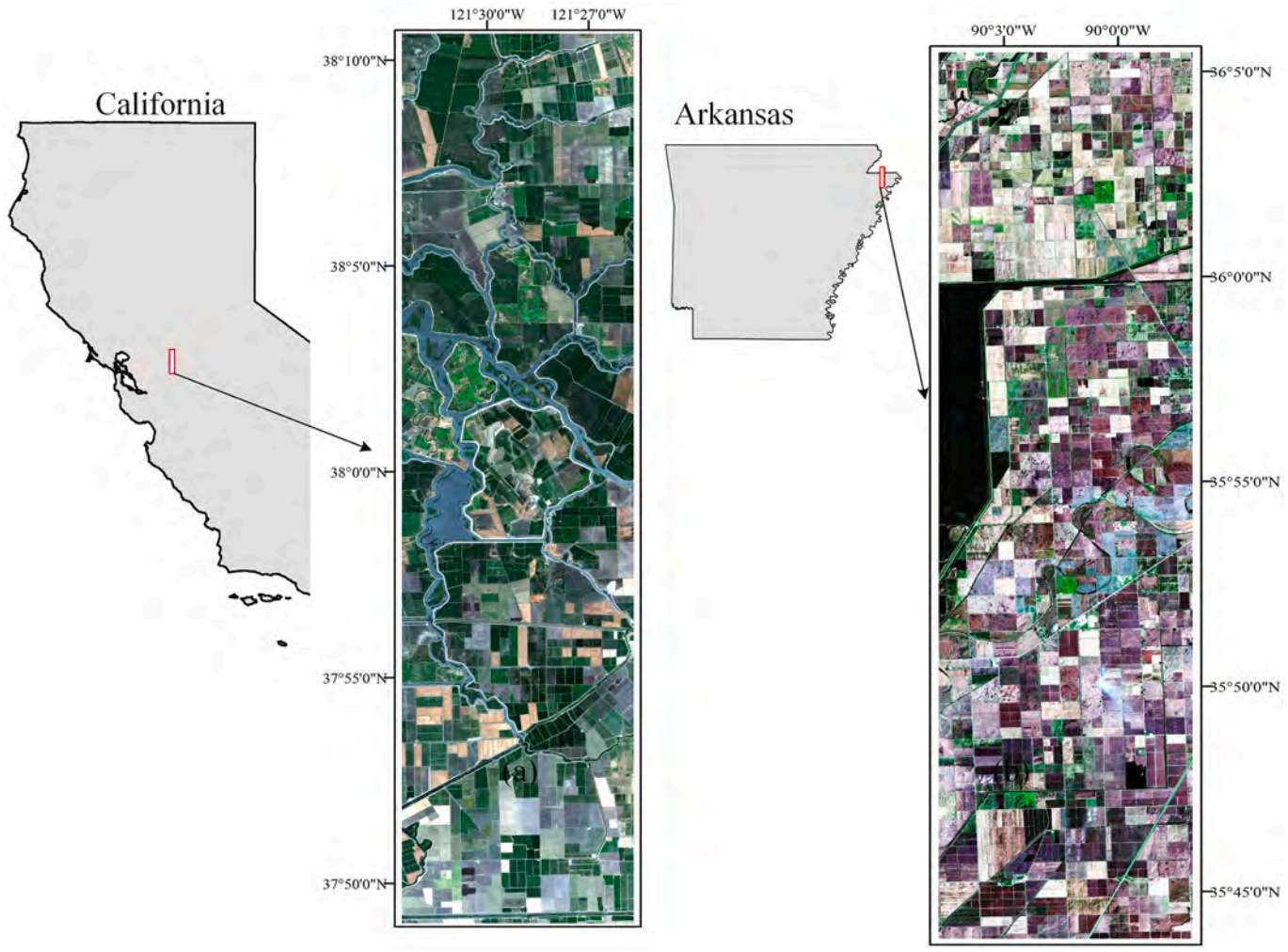

Fig. 1. Illustration of PhenoCam sites and the selected study areas. (b) TA1 S2 satellite image dated on 2018/06/24 with true-color composite, and (c) TA2 satellite image dated on 2018/10/18 with true-color composite.

RNNs, it can process the whole sequence data in parallel and focus on long-range dependencies of critical information for sequential pattern identification (Vaswani et al., 2017). This could be a promising solution for crop phenology identification by formulating SAR and optical long-range contextual dependencies.

Still, the discrepancies between satellite and ground observations remain unsolved. Currently, most crop phenology studies focus solely on temporal feature formulation (Zeng et al., 2020), while neglecting the importance of spatial features. Moreover, there are still significant differences between satellite observations and ground cameras (i.e., the mismatch between satellite and ground observations) (Tian et al., 2021). To be specific, the accurate phenological information acquired by digital cameras represent a certain interest of regions (e.g., 50-100 m) in a specific direction, while individual pixel of S2 or S1 only covers a small area (e.g., 10 m*10 m) which inevitably reduces the consistency between satellite and ground observations. The integration of crop features from both the spatial and temporal dimensions is also needed.

The purpose of this research is to integrate optical and SAR data to improve the accuracy of crop phenological parameter extraction. Specifically, we combine SAR and optical satellite observations to create spatial-aware deep learning models for crop phenology identifying with the help of ground observations. To serve this purpose, we first extract multi-source satellite images on small areas around the ground sites (e.g, PhenoCam) to capture the growth information on the crop that is consistent with the camera observation range. Then, the spatial-temporal aware deep learning-based crop phenology model (DeepCroP) is trained to extract crop spatial-temporal evolving features with ground phenological labels acquired by PhenoCam. Finally, the timing of phenological changes is extracted from the long time-series features over a large-scale area with a previous well-trained deep learning model. Therefore, this paper offers a new approach to combine the advantages of multi-source satellites and ground observations for crop phenological retrieval, which is potentially valuable for future phenology studies. In general, the contributions of this work are:

(1) it explores a new crop phenology tracking mechanism by combining the advantages of optical and SAR time-series.
(2) it proposes an effective crop phenological parameter extraction strategy with spatial-aware deep features and long-range time-series dependency.

The paper is organized as follows: Section 2 starts with the introduction of study areas and data collection for both PhenoCam time-series and satellite time-series. Section 3 presents the proposed DeepCroP model. Then, Section 4 analyzes the accuracies on phenology retrieval. Limitations of the model are discussed in Section 5. And, the conclusion is illustrated in Section 6.

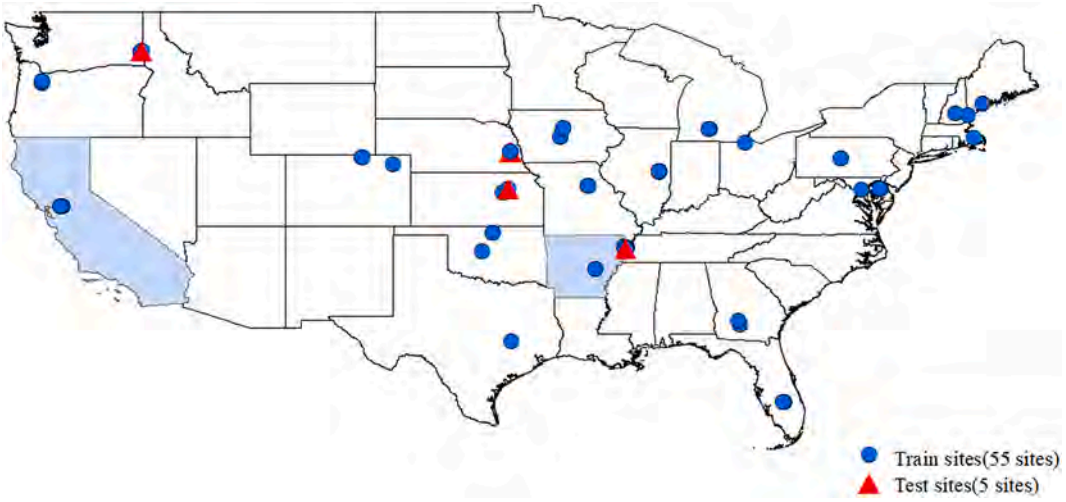

Fig. 2. The illustration of 60 PhenoCam sites with crop phenology observations.

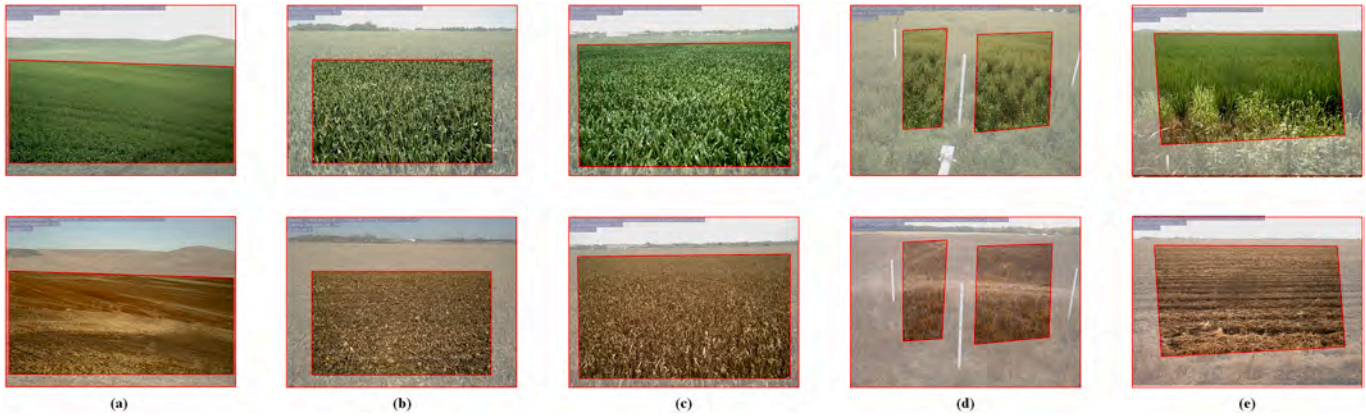

Fig. 3. Images acquired by the PhenoCam. The red polygon indicates the area of interest used for the extraction of the GCC series. (a) cafcookwestlar01 site, (b) mead2 site, (c)mead3 site, (d) NEOND06DP100042 site, (e)ufs6 site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Study area and data

2.1. Study areas

The first test area (TA1) is situated in an agricultural district stretching over Solano and Yolo counties of California, Northern California, around 38°N and 121°30′W, covering a region about 40 km × 10 km, as shown in Fig. 1 (left). The area has a Mediterranean climate with dry, hot summers and wet, cool winters. This region has plain terrain, and it is one of the most productive agricultural areas with a complex agricultural system in the United States. The main crop types for this area include corn, rice, winter wheat, and tomatoes with significant phenology variations. Annual precipitation of about 750 mm, concentrated in the spring and winter seasons.

The second test area (TA2) is also located in the agricultural area of northeast Arkansas, around 35°55′ N and 90°W, covering an area of about 40 km × 10 km, as shown in Fig. 1 (right). Arkansas has a humid subtropical climate, generally with humid summers and mild, slightly drier winters. In this area, most vegetations reach full bloom by early April. The area is one of the largest sources for rice production as well as soybeans, corn, cotton, wheat, and grain sorghum.

2.2. Data collection and preparation

2.2.1. In-situ phenology observation sites

The technique of near-surface remote sensing for vegetation phenology monitoring has advanced greatly over the past decade. In this context, the PhenoCam Network has deployed more than 500 web-enabled cameras across the globe (Brown et al., 2016). The PhenoCam repeatedly acquires photography to capture color information and measure variations in vegetation phenology across diverse ecosystems.

The indicator GCC_c (i.e., the GCC data calculated by the camera photography) was acquired by calculating the ratio of the green channel digital numbers to the total brightness in RGB camera images as described in Richardson et al. (2018). Due to the constant degradation of PhenoCams, the quality of GCC_c time-series signal varies (e.g., without significant seasonal patterns, large temporal gaps, and limited available observations) and thus requires intensive manual screening (Liu and Wu, 2020). Given such conditions, we selected 60 PhenoCam agricultural sites to carry out our studies (Fig. 2). Among them, 5 test sites were selected to quantitatively analyze the performance of the model. And, Fig. 3 provides representative photos captured by the PhenoCam camera.

2.2.2. Optical-SAR satellite time-series

The available Sentinel-1 satellite images in GRD mode and Sentinel-2 satellite images were obtained from Google Earth Engine (GEE)

Table 1
Vegetation indices calculated from Sentinel-2 images. B=Blue, G = Green, R = Red, NIR = Near-infrared.

Index	Equation	S2 bands used
NDVI	$NDVI = (NIR - R)/(NIR + R)$	$NDVI = (B7 - B4)/(B7 + B4)$
GCC_s	$GCC\_s = G/(B + G + R)$	$GCC\_s = B3/(B2 + B3 + B4)$

platform. The processing steps of Sentinel-1 include: (1) thermal noise removal, (2) applying orbit file, (3) radiometrically calibrated to sigma0, (4) geocoding, and (5) backscatter images (δ) transform to the logarithmic dB scale. And we calculated NDVI and GCC_s (subscript s means satellite) with the spectral bands 2, 3, 4, and 8 of Sentinel-2 imagery, as shown in Table 1.

Fig. 4a. and Fig. 4b. demonstrate the temporal distribution of Sentinel-1 and Sentinel-2 images collected in the two test areas. In total, we downloaded all available (partially) cloud-free Sentinel-2 imagery over two test areas from Jan. 2018 to Jan. 2019, with 42 and 32 scenes, respectively. And 28 scenes of Sentinel-1A images from Jan. 2018 to Jan. 2019 were collected for both TA1 and TA2.

In order to build the relationship between in-situ- and satellite-derived LSP, we also downloaded corresponding S2 and S1 data for all 60-study sites. As we mentioned, the PhenoCam is designed to observe the local region of interest (ROI) given the orientation direction, and the spatial coverage of each ROI is larger than individual pixels in high-resolution imagery of the satellite. To tackle the mismatch between the ROI coverage (i.e., improve the correlation of PhenoCam GCC_c and satellite VI time-series curves), we segmented a fixed-size (18×18) square patch from S1 and S2 imagery where the center pixel is the location of the local camera site, as shown in Fig. 5.

2.3. SOS/EOS extraction

Currently, a number of LSP extraction methods were proposed to identify the timing of SOS and EOS, which could be broadly divided into two categories: the inflection point category and the threshold category (Beck et al., 2006; Meroni et al., 2021; White et al., 2009; Zhang et al., 2003). Among them, the inflection point category uses a fixed definition to determine SOS and EOS timing, and the threshold category select a threshold of seasonal amplitude to define the SOS and EOS timing, and the latter has been widely used in studies of vegetation phenology tracking. In this study, a local polynomial function (i.e., adaptive Savitzky-Golay) is introduced for time-series curve fitting with the TIMESAT (Jönsson and Eklundh, 2004) software package to reconstruct the continuous seasonal trajectories from irregular and noisy GCC_c time series. To construct the reference labels of PhenoCam time-series induced SOS/EOS, we chose 20% and 50% fractions of seasonal amplitude to derive SOS and EOS, respectively, as shown in Fig. 6.

3. Methodology

The overall framework of the proposed crop phenology retrieval scheme is shown in Fig. 7. For the complete phenology detection, there are four steps before crop phenological parameters can be determined, that are, 1) Ground and optical-SAR satellite data acquiring; 2) Time-series index calculation, including NDVI, GCC_s, VV, VH, GCC_c; 3) Training spatial-aware deep learning (Deep-CroP) model for satellite phenology extraction with the reference phenological parameters (such as SOS, EOS) extracted from PhenoCam. Finally, the fine-scale crop phenology tracking and mapping with the well-trained Deep-CroP model. In addition, we quantitatively validate the Deep-CroP model derived EOS and SOS timing to PhenoCam observations over different test sites.

Fig. 8 shows a visual representation of the proposed Deep-Crop architecture. The model takes S1 and S2 satellite images as input that to be fed into two streams. Then, the proposed architecture extracts feature from both SAR and optical time-series imagery with separate streams and spatio-temporal features concatenated by an elaborated multi-level fusion module. Finally, phenological parameters are identified with the help of multi-head attention modular. In the following, we will provide a detailed description of the spatial-aware time-series feature extraction and multi-modular deep feature fusion.

3.1. Spatial-aware optical-SAR time-series feature fusion

In this scope, we designed the integrated model based on two-stream architecture to extract features from SAR and optical images with different branches. Specifically, each branch contains a three convolutional layer CNN framework with the input image sizes of 18 × 18. The first Conv2D layer has 32 features, the second Conv2D layer has 64

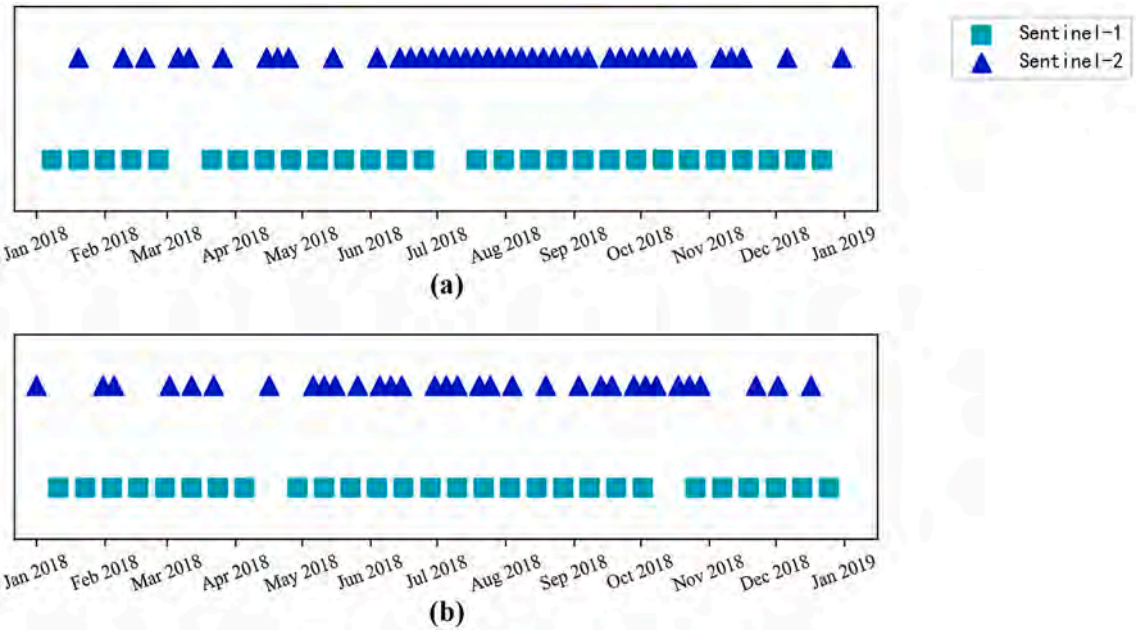

Fig. 4. The temporal distribution of Sentinel 1 and 2 over the TA1 (a) and TA2(b).

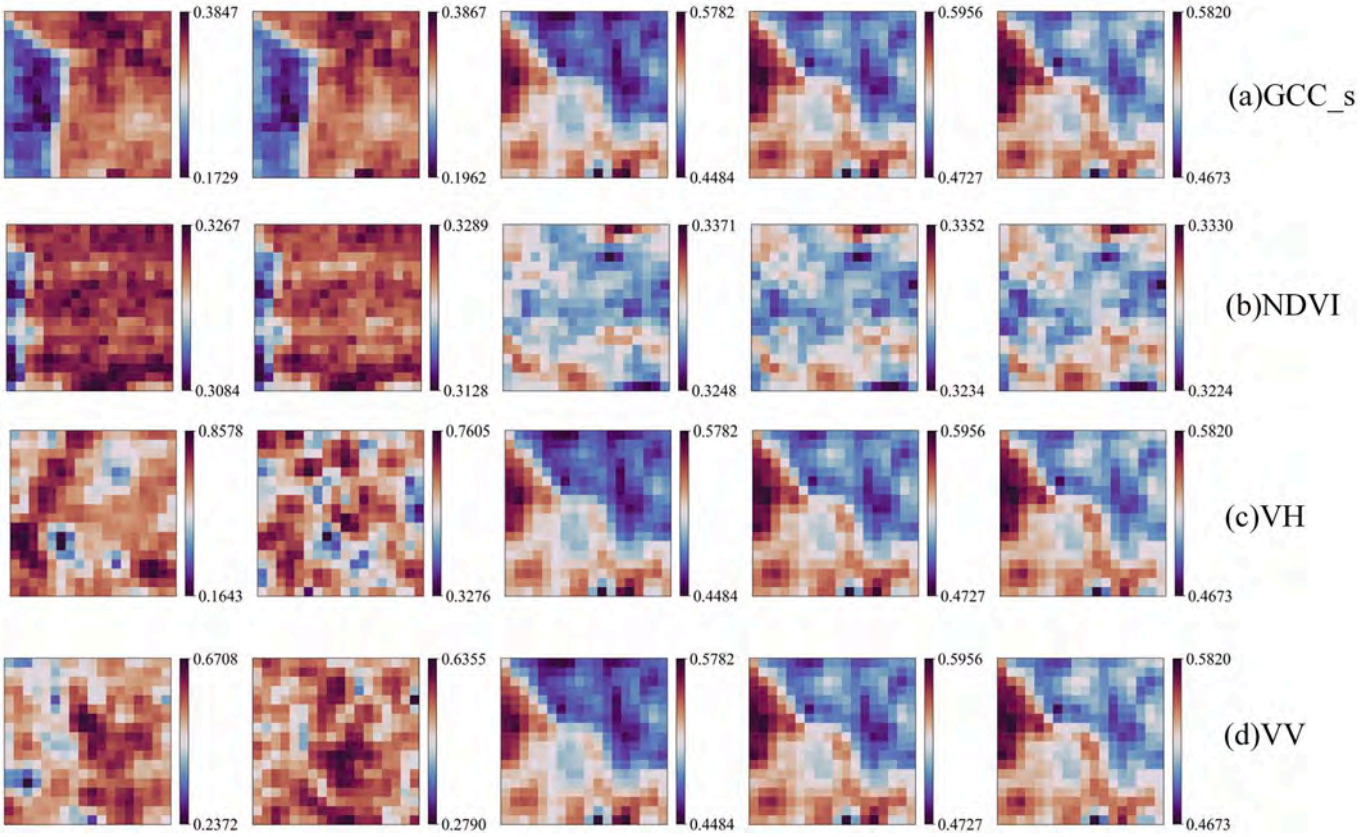

Fig. 5. The illustration of the spatial-patches from satellite time-series. (a) NDVI; (b) GCC_s; (c) VH; (d) VV images around PhenoCam location.

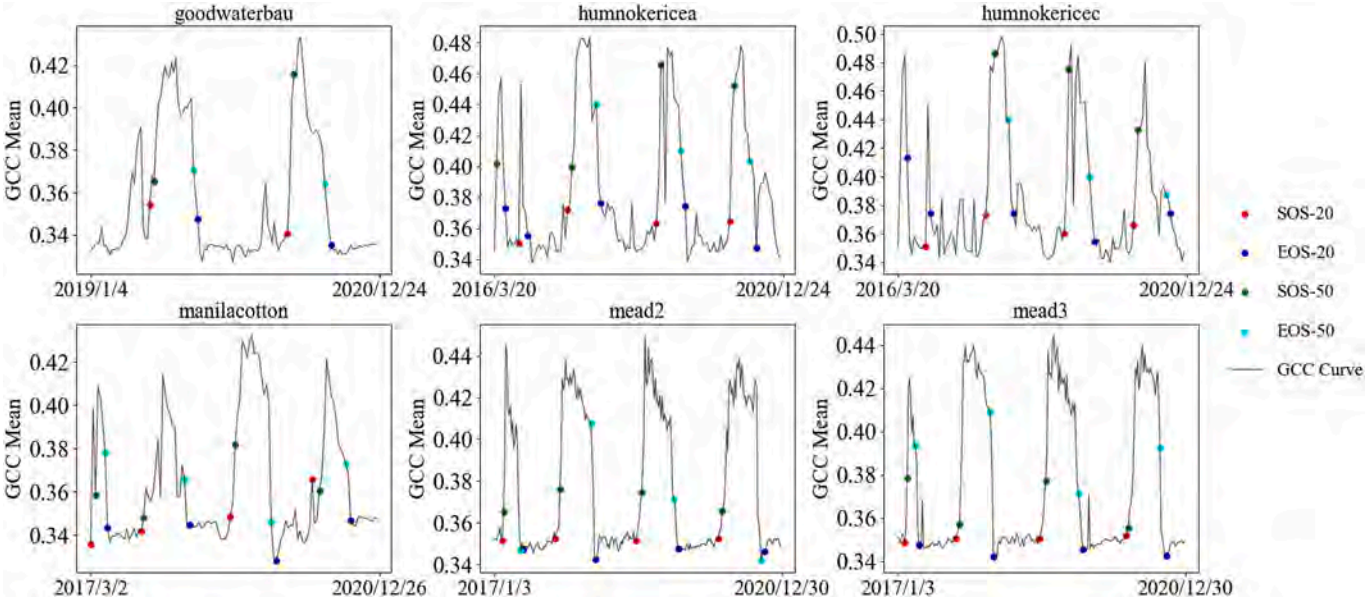

Fig. 6. Crop phenology (i.e., SOS/EOS) timing point derived from GCC_c time-series on 6 representative sites.

features with a max-pooling operation, and the last Conv2D layer contains 128 features, where both Conv2D layers share the filter sizes of 5 × 5. In the convolutional phase, each convolution operation is delivered to a nonlinear activation function, such as the rectified linear activation function (ReLU). This phase is as follows:

$$y^{(j)} = ReLU\left(bisa^{(j)} + \sum_i W^{(i)(j)} \circledast x^{(i)}\right), \quad (1)$$

where $x^{(i)}$ and $y^{(j)}$ are the input feature map and output feature map of the i-th and j-th convolution layers, and $j = i + 1$, respectively. W and $bisa$ are parameters learned by the neural network model, and $\circledast$ is the convolution operator. In the pooling phase, a pooling function

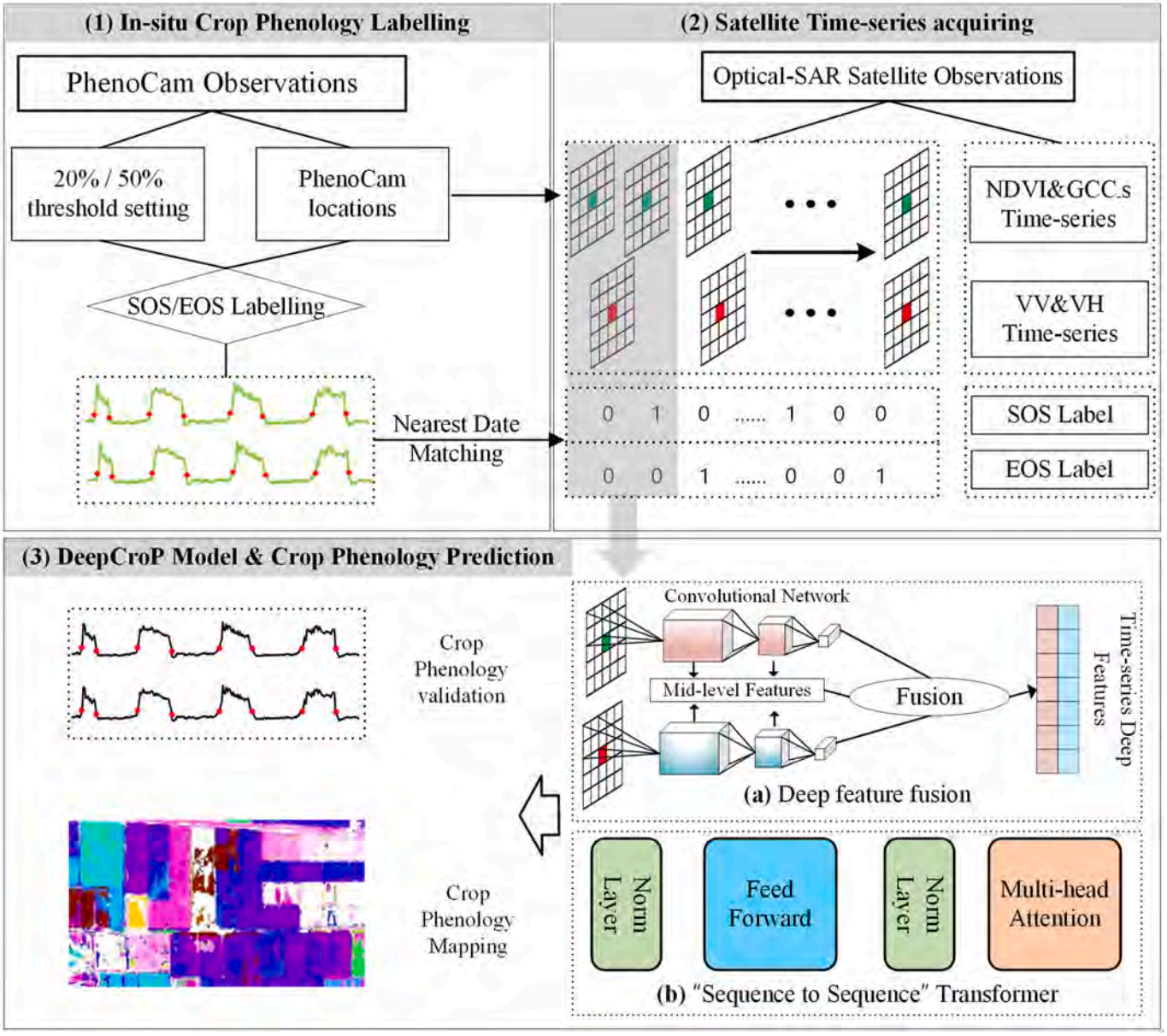

Fig. 7. The overall framework of the calibration process. (1) PhenoCam time-series SOS/EOS labeling; (2) Satellite time-series imagery acquiring with spatial patch extraction; (3) DeepCroP model training and crop phenology mapping.

configured is adopted to perform local averaging, and sub-sampling, contribute to local translation invariance, and reducing the number of parameters and computational complexity of the model, as follows:

$$y_{p,q} = \max_{m,n\in\{-R,\cdots,R\}} \left(x_{p+m,q+n}\right) \tag{2}$$

where $y_{p,\,q}$ is the neuron value at (p,q) in the output layer, and m, n indicate the pixel location around the center neuron, and the window half-length as R.

Convolution operation has a strong non-linear fitting capacity, still it cannot maintain the integrity of its multi-source input data (Huang et al., 2017). Therefore, the final output feature of two-streams may lose some complementary information between the two data. Based on this, we construct a fusion block in the process of feature extraction of two-stream. The fusion block can maintain the feature maps obtained at each CNN branch of the two-stream framework and concatenate the output to an integrated feature vector. The concatenation result of two-streams is a feature sequence that summarizes the extracted feature from optical and SAR time-series imagery, including $F_{s1} \in \{y_1^{S1}, y_2^{S1}, \cdots, y_{T1}^{S1}\}$ and $F_{s2} \in \{y_1^{S2}, y_2^{S2}, \cdots, y_{T2}^{S2}\}$ where single output feature has the dimension of $F_{s1} \in \mathbb{R}^{T_1,\ d_1,H,\ W}$, for each y has the dimension of Conv2D output feature d_1, and y_{T1}^{S1} is used to indicate the output feature map of the last convolution layer in $T1$-th temporal step of S1 branch. Also, for the S2 branch, y_{T2}^{S2} has the dimension of $d_2 \times H \times W$ in terms of Conv2D output features. Meanwhile, the fusion operator accepts the mid-level features from both S1 and S2 CNN branches and generates F_{fusion} features.

$$F_{fusion} = sigmoid\left(f^{conv}\left(concat\left[F_{s1}^{mid}, F_{s2}^{mid}\right]\right)\right), \tag{3}$$

Where *concat* is the concatenation operation that stacks deep features from multi-sources. F_{s1}^{mid} and F_{s2}^{mid} are the mid-level features acquired by CNN branches. f^{conv} represents the 1D-CNN framework that deeply integrates mid-level features. Lastly, we concatenate all the extracted feature sequences (including F_{s1}, F_{s2}, F_{fusion}) in the temporal and spatial dimensions, as $X \in \mathbb{R}^{T_{1+2},\ d_1\times(H\times W)+d_2}$, shown in Fig. 8 (a). In this way, for multi-source satellite imagery, the seasonal variation information in both the temporal and spatial domain can be efficiently represented.

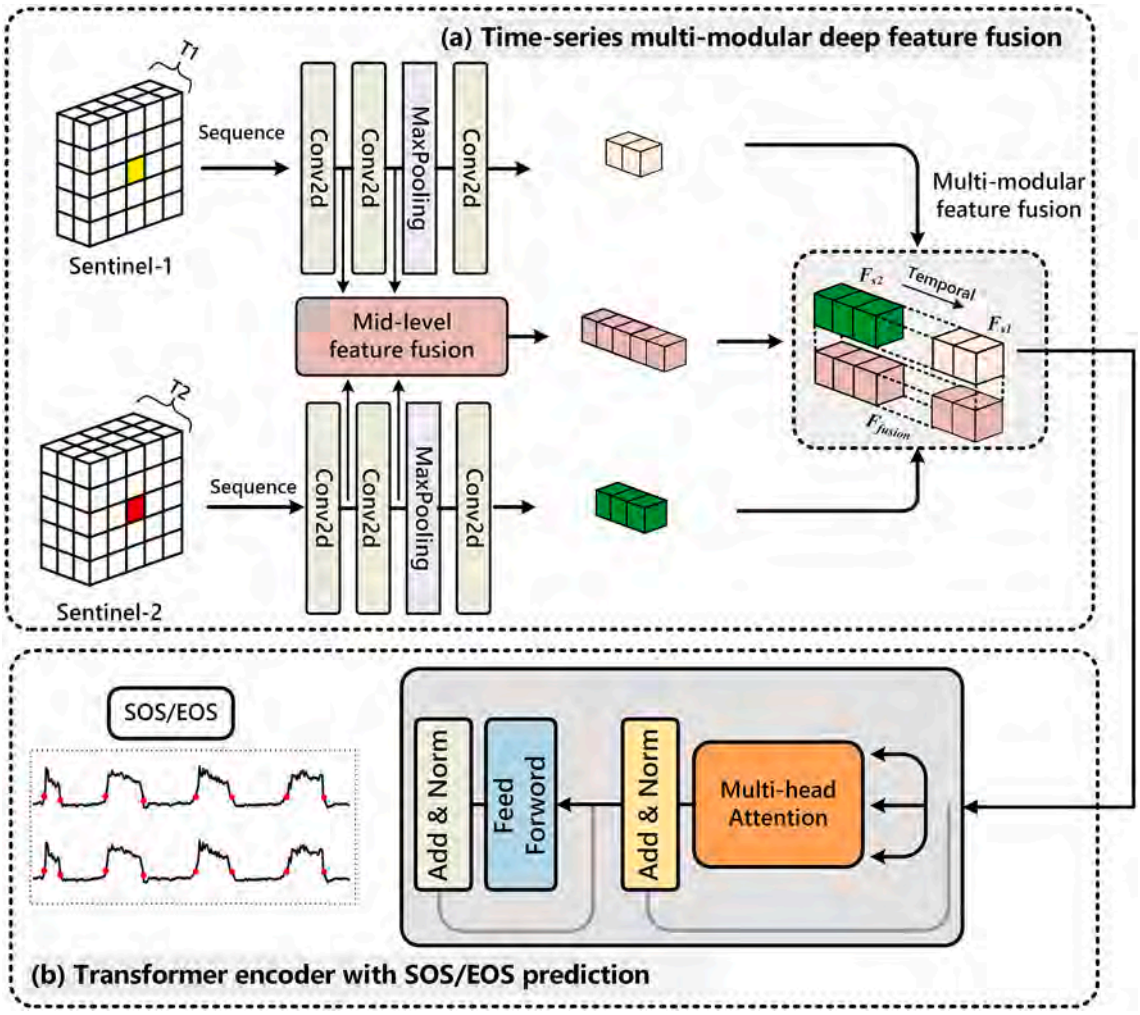

Fig. 8. The flowchart of the Deep-CroP model. (a) Two branches are designated to extract spatial-aware time-series feature and fuse multi-modular deep feature; (b) Transformer encoder with SOS/EOS prediction.

Table 2
The Mean Absolute Error (MAE) for the satellite-based SOS/EOS timing based on NDVI and Deep-CroP model derived crop phenological parameters compared to the PhenoCam observations at 5 test sites.

Site	NDVI				Deep-CroP			
	SOS20	EOS20	SOS50	EOS50	SOS20	EOS20	SOS50	EOS50
cafcookwestltar01	66.3	1.7	16.3	8.3	11.0	5.6	8.3	11.0
mead2	33.3	9.7	14.3	17.0	5.0	3.3	10.0	7.7
mead3	37.0	8.8	15.3	17.5	4.0	7.5	10.0	2.0
NEOND06DP100042	35.0	3.3	15.0	23.3	6.7	15.0	15.0	8.3
usof6	47.5	30.0	25.0	25.0	12.5	2.0	10.0	5.0

3.2. Multi-modular deep feature for phenology identification

To capture the phenology changes for crop fields, the transformer that exclusively relies on the self-attention mechanism is able to capture global dependencies has become the model of choice in natural language processing (NLP). Therefore, for multi-source phenological feature extraction, we apply transformer encoder, including Multi-Head Self-Attention module (MHSA), Feed-forward module (FFN), and Multi-Layer Perceptron (MLP) to perform feature formulation and phenological parameter extraction. Initially, the input size of MHSA is 30 with 30 multi-heads. Then, FFN with 900 units is designed to compact the multi-branch features. Lastly, a 4-layer MLP with the configuration of 10,000, 5000, 1000, and 500 units is applied to the final phenological parameter estimation. Suppose the combined feature $x = [F_{s1}, F_{s2}, F_{fusion}]$ in sequence is input into the encoder of the transformer. The overall process can be summarized as:

$$x_l' = MHSA(LN(x_{l-1}) + x_{l-1}), \tag{4}$$

$$x_l = LN\left(FFN\left(x_l'\right) + x_l', \right. \tag{5}$$

$$y = LN(x_L), \tag{6}$$

$$MHSA = Concat(head_1, \cdots, head_h)W^0, \tag{7}$$

$$head_1 = softmax\left(QK^T \Big/ \sqrt{d_k}\right)V, \tag{8}$$

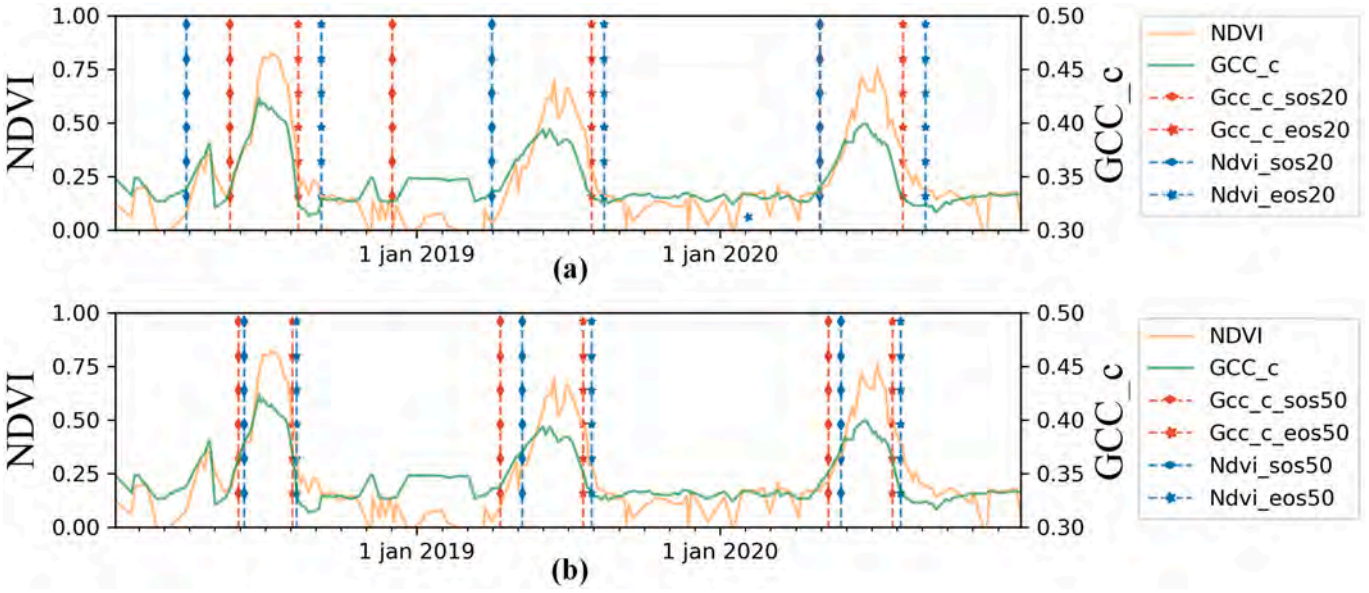

Fig. 9. Temporal profiles of NDVI, GCC_c, and the (a)SOS20/EOS20, (b)SOS50/EOS50 date derived from the fitted curves with TIMESAT at Camera_Site1.

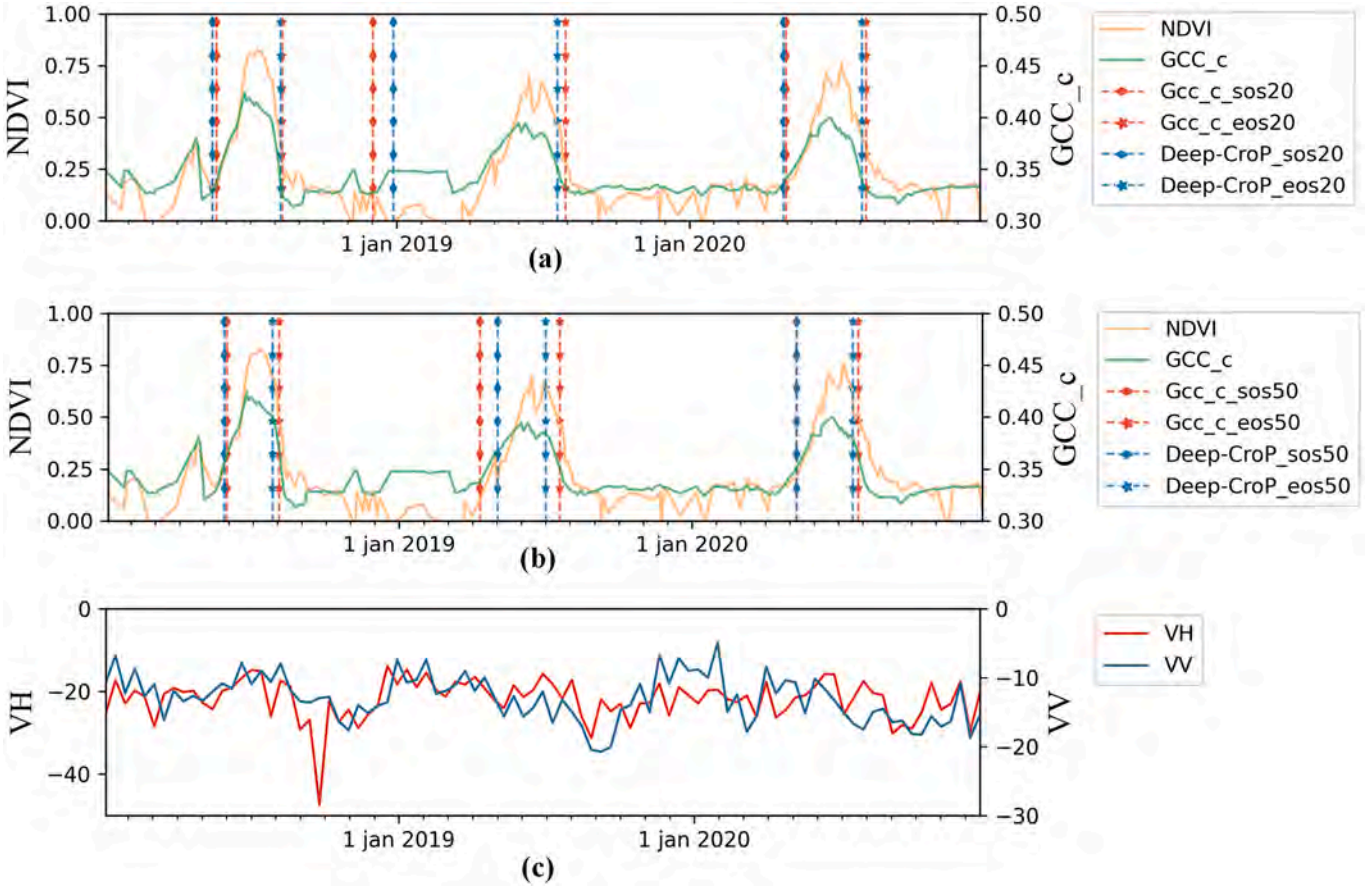

Fig. 10. Temporal profiles of NDVI, GCC_c, VV, and VH. The (a)SOS20/EOS20, (b)SOS50/EOS50 dates derived from the fitted curves with TIMESAT and Deep-CroP model at Camera_Site1.

Where x_{l-1}, x_l are the input and output of the l-th layers, respectively. x_l' is the output of the l-th multi-head self-attention sub-layers, and y represent the output of the transformer encoder, $l \in (1, L)$.

$$P(Y = i|y) = \frac{e^{y_i}}{1 + e^{y_i}}, \tag{9}$$

With the deep feature descriptor y, crop phenological stages can be further assigned using the logistic regression. The final probability P denotes the potential location of crop phenological stages at i-th temporal point. In this way, crop phenological parameters can be identified.

Table 3
The discrepancies between phenological parameters retrieved from the satellite NDVI-derived, Deep-CroP predictions with PhenoCam observations at Camera_Site1. Best results are shown in bold.*The crop type information comes from the NASS Cropland Data Layer.

Year	Crop type	SOS20		SOS50		EOS20		EOS50	
		NDVI	Deep-CroP	NDVI	Deep-CroP	NDVI	Deep-CroP	NDVI	Deep-CroP
2018	Dry beans	17	**−5**	7	**−3**	−2	−2	**5**	−8
2019	Winter wheat	157	**25**	27	**22**	**0**	−10	**10**	−18
2020	Spring wheat	25	**3**	15	**0**	**−3**	−5	10	**−7**

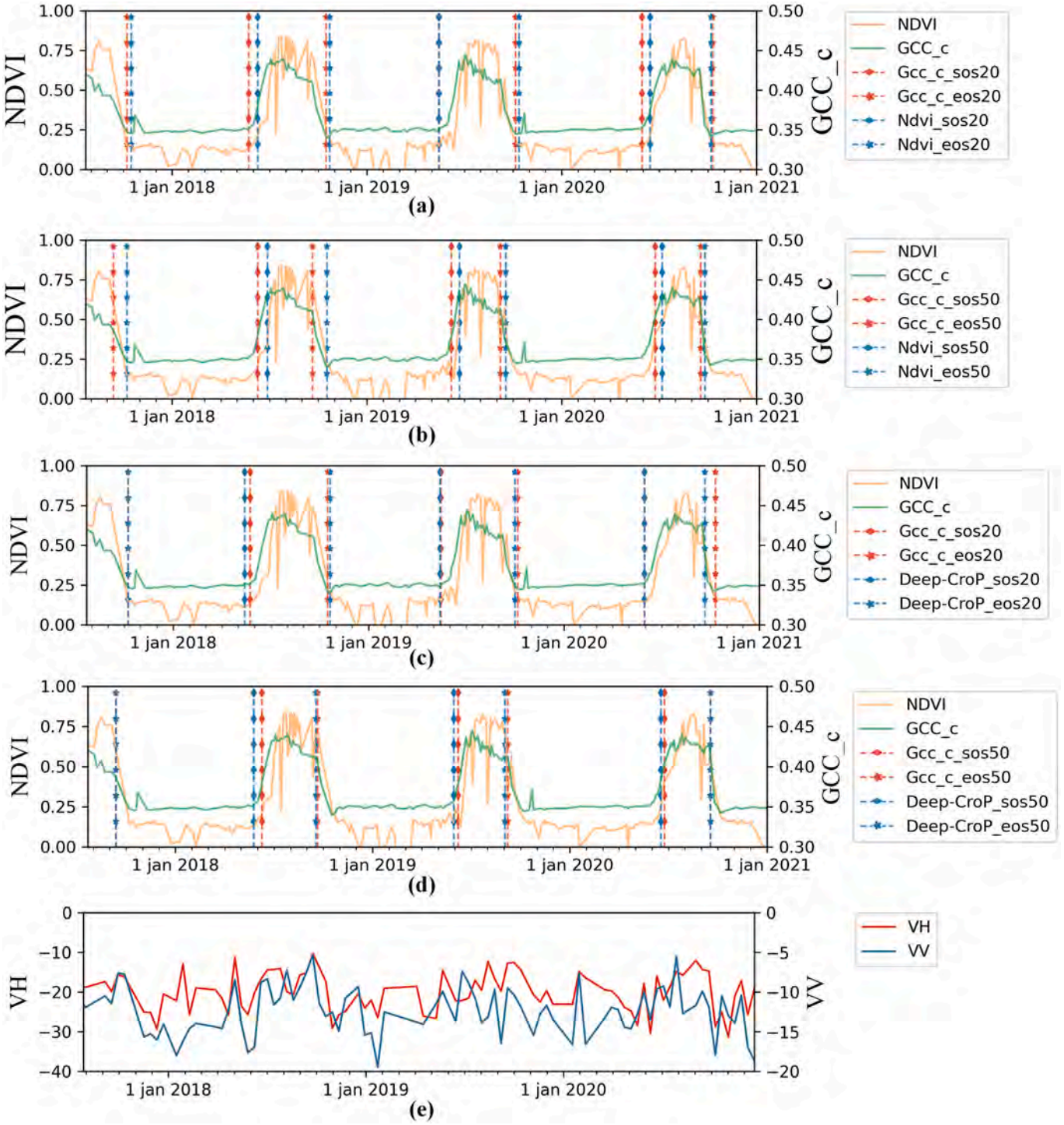

Fig. 11. Temporal profiles of NDVI, GCC_c, VV, and VH. The EOS/SOS date derived from the curve fitting with TIMESAT and Deep-CroP model at Camera_Site2.

Table 4
The discrepancies between phenological parameters retrieved from satellite NDVI-derived, Deep-CroP predictions with PhenoCam observations at Camera_Site2. Best results are shown in bold. *The crop type information comes from the NASS Cropland Data Layer.

Year	Crop type	SOS20		SOS50		EOS20		EOS50	
		NDVI	Deep-CroP	NDVI	Deep-CroP	NDVI	Deep-CroP	NDVI	Deep-CroP
2017	Corn	/	/	/	/	0	0	25	**0**
2018	Soybeans	35	**−10**	18	**−15**	**2**	5	27	**−3**
2019	Corn	38	**−2**	15	**−8**	−18	**−5**	10	**−5**
2020	Soybeans	38	**0**	13	**−7**	**−15**	−20	8	**0**

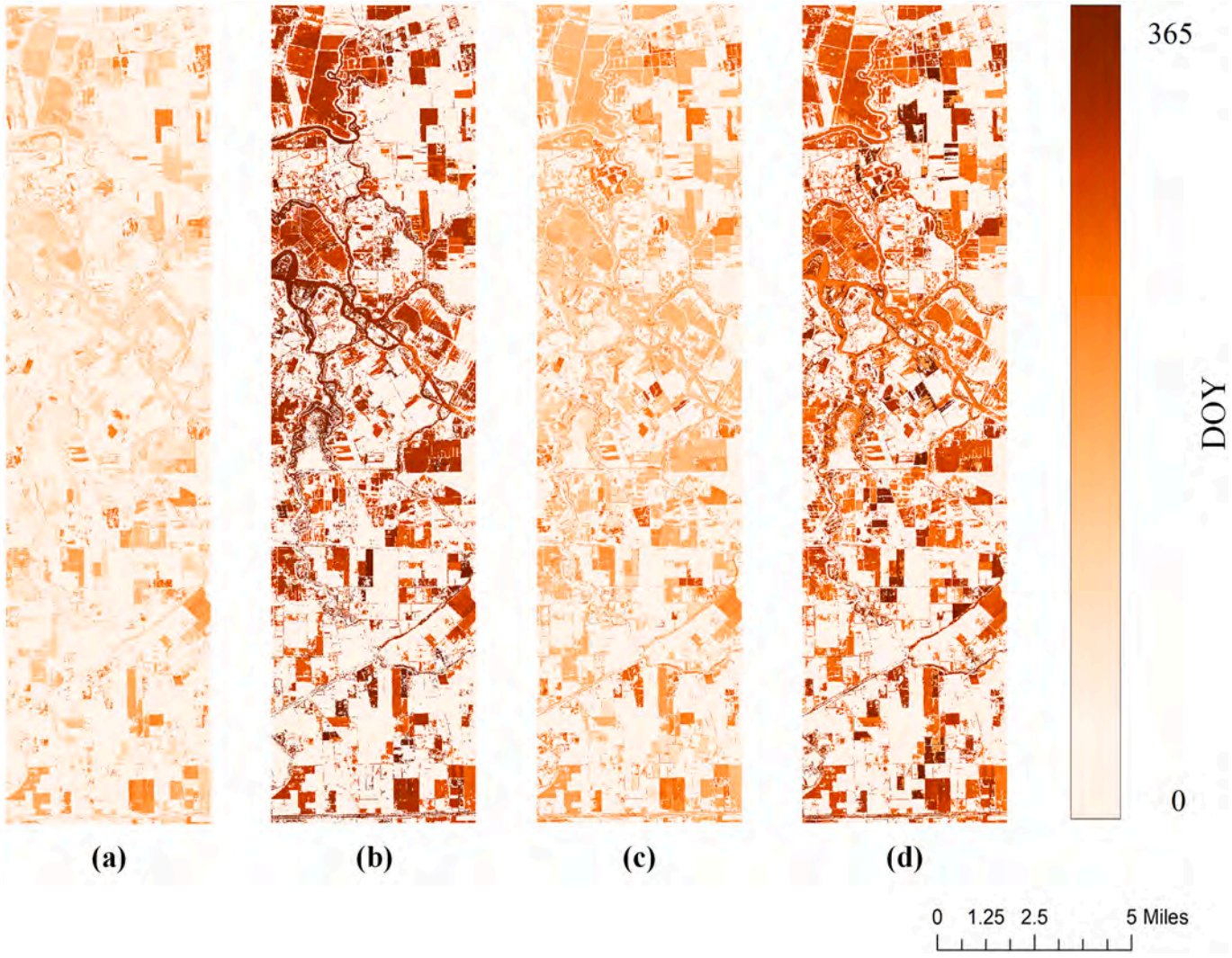

Fig. 12. Crop phenology mapping on (a)SOS20, (b)EOS20, (c)SOS50, and (d)EOS50 from Sentinel-2 NDVI time-series (based on the fitted curves with TIMESAT) for the TA1.

4. Experiments and results

4.1. Experiments designs

In this section, the detailed experiment design is explained. For the two-stream Conv2D framework, a 3-layer CNN framework was trained separately based on SAR and optical time series data. The sizes of input samples were set to $30 \times 18 \times 18 \times 2$ and $10 \times 18 \times 18 \times 2$ for Sentinel-2 and Sentinel-1 branches, respectively. In order to get a better grasp of the complementary information of the two data, the output of each convolution layer for the two branches was combined and input into the fusion block. Then, the output of the two CNN branches, the output of the fusion block, and the time-series signal of the site location were concatenated with global and local awareness. Finally, the resulting feature was fed into the transformer to detect the crop phenological parameters in the temporal domain.

To generate crop phenological reference labels, we applied the time-series curve fitting technique to extract phenology timing points in the PhenoCam time-series with the TIMESAT software. To be specific, we set the threshold to 20% and 50% for phenology parameter extraction under different standards. For the Deep-CroP-based phenology tracking, we set the reference label as the nearest date of available satellite imagery (i.e., cloud-free optical image or SAR image), due to the limited temporal resolution of satellite imagery. Once the reference label was

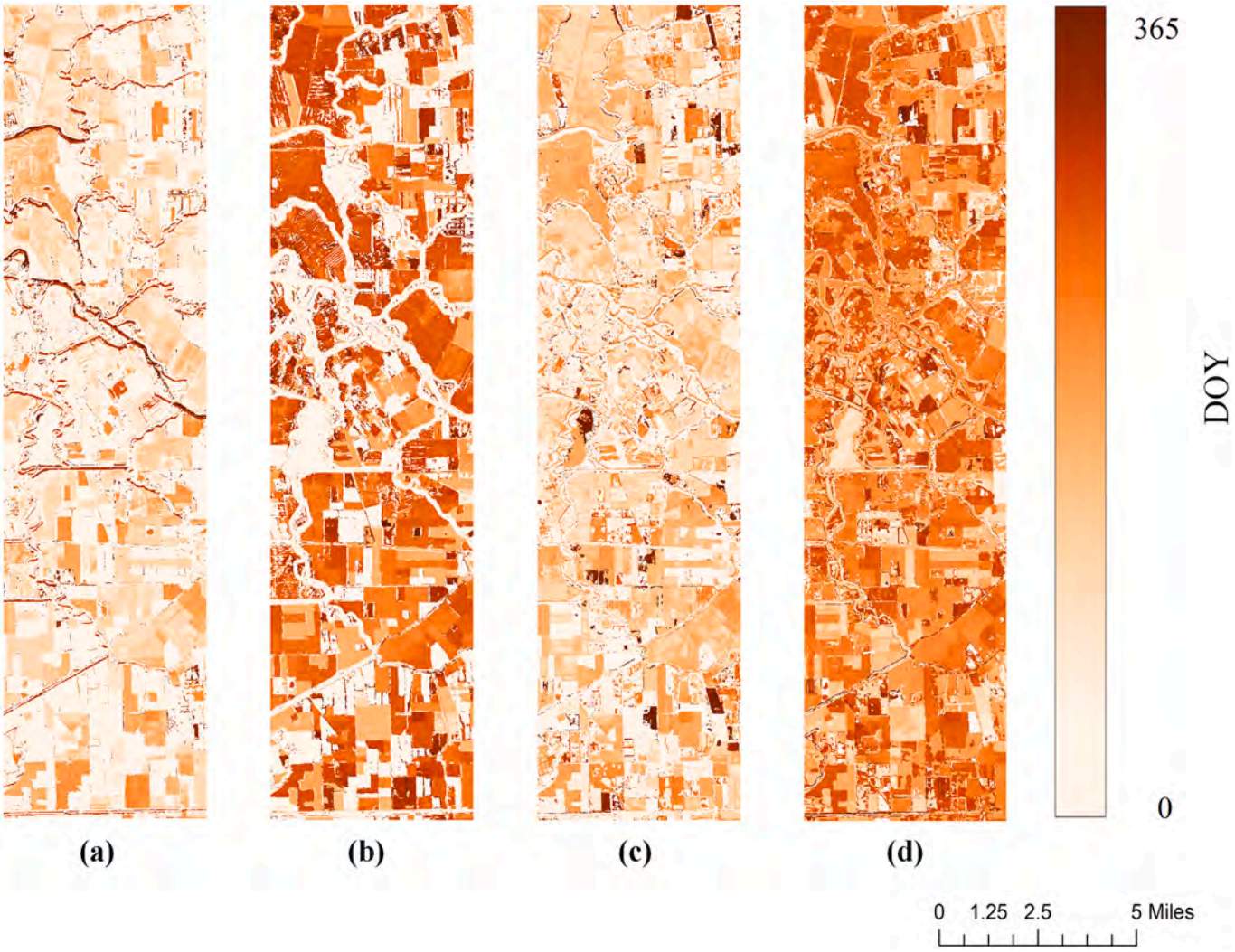

Fig. 13. Crop phenology mapping on (a)SOS20, (b)EOS20, (c)SOS50, and (d)EOS50 from Sentinel-1 and Sentinel-2 series with Deep-CroP for the TA1.

constructed, we can train and validate the Deep-CroP model efficiently. During the training stage, the learning rate was set to 0.0001, and the mini-batch size was set to 50. The spatial size of extracted image patches is fixed to 18. And the temporal window widths of optical and SAR were set into 30 and 10, respectively. For the competitive NDVI-based phenology calculation (without nearest date modification), we also utilized TIMESAT software to predict SOS/EOS from optical satellite time-series data with 20% and 50% threshold settings.

4.2. *Phenological retrievals from camera vs. satellite*

To evaluate the consistency of phenological parameters retrieved with satellites by the NDVI-based method and the proposed multi-source Deep-CroP model with standard ground site observations (PhenoCam). We calculated the Mean Absolute Error (MAE) for each crop phenological parameter compared to the PhenoCam phenology baselines. In this experimental design, we aimed to demonstrate the phenology timing consistency between the PhenoCam observations and the Deep-CroP model predictions compared to the conventional satellite-based models. To serve this purpose, we analyzed five test stations in terms of quantitative evaluation on crop phenological parameters, as shown in Table 2. Based on the PhenoCam-derived crop phenological SOS/EOS parameters, we compared them with the NDVI and Deep-CroP-derived results. For the conventional NDVI-based phenology identification, under the threshold of 20%, large discrepancies can be found in SOS with a minimum absolute value of 33.3 for five tested sites. However, smaller differences can be observed in EOS20 for the first four sites, while significant variations worsen the prediction results. Similarly, for the threshold of 50%, both SOS and EOS witnessed large discrepancies compared to PhenoCam phenological parameters. For the Deep-CroP model trained with PhenoCam phenology labels, under the threshold of 20, the predicted SOS and EOS are much more accurate compared to the NDVI-based method with MAE values ranging from 4 to 15. Also, given the reference labels with threshold 50, the Deep-CroP model predicts crop phenological parameters with similar accuracies for five test sites. To validate the spatial consistency of the Deep-CroP model, both the mead2 and mead3 share similar performances and are adjacent PhenoCam sites. In the following study, we selected two representative sites, cafcookwestltar01, and mead3 as Camera_site1 and Camera_site2 for the detailed analysis on phenological parameter extraction.

4.2.1. *Camera_Site1*

Temporal profiles of PhenoCam-derived GCC_c (GCC with cameras) and optical satellite-derived NDVI time-series curves are displayed in Fig. 9a. Together with the extracted phenological parameters (i.e., SOS and EOS) for GCC_c and NDVI are plotted against each other. The GCC_c is displayed with the secondary y-axis as it has smaller values than NDVI.

The Fig. 9. shows that the GCC_c and NDVI time-series data both demonstrated distinct seasonality and exhibit similar temporal behavior. It is worth noting that GCC_c series generally display an earlier

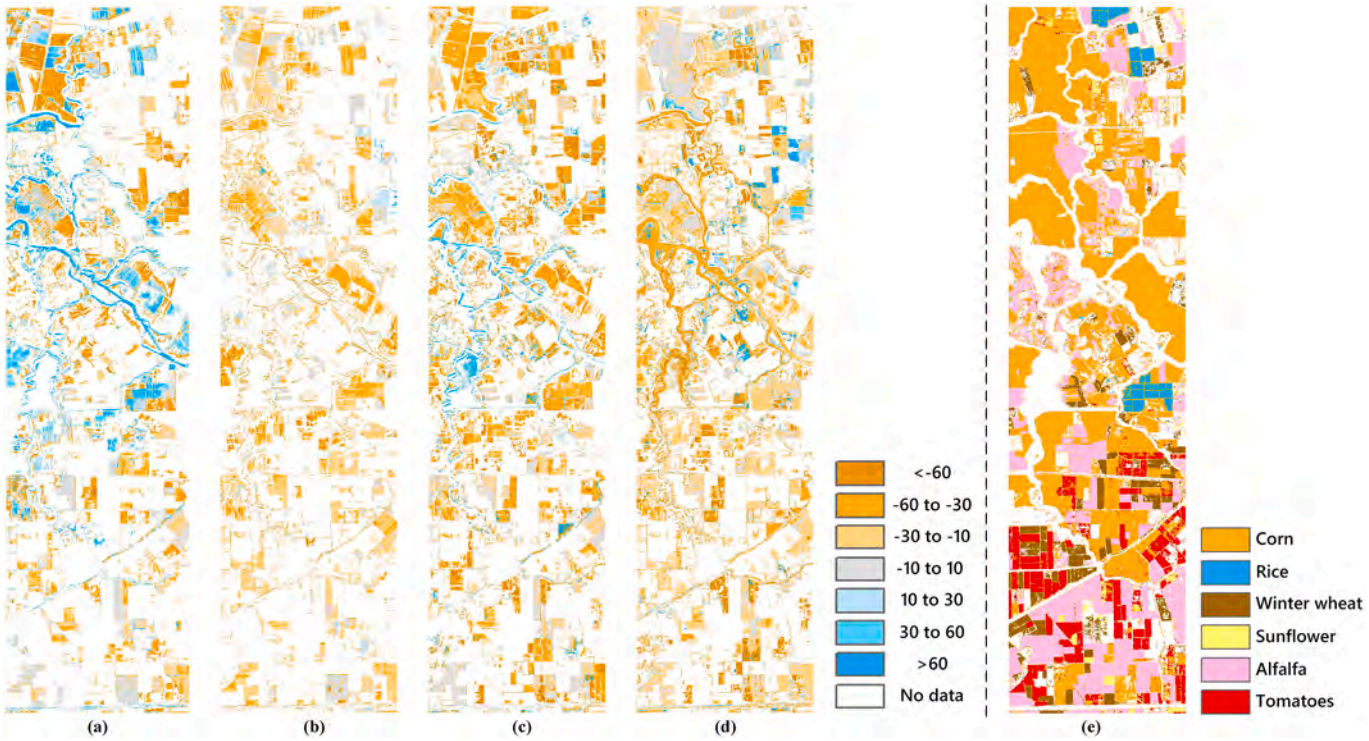

Fig. 14. Difference (in days) between phenology stages derived from curve fitting (Sentinel-2 NDVI series) and Deep-CroP (Sentinel-2 and Sentinel-1 series) for (a) SOS20, (b)EOS20, (c)SOS50, (d)EOS50, (e)2018 CDL data in TA1. Negative values mean that the Deep-CroP-derived resulting in earlier dates than the NDVI-based method.

and faster decrease in greenness than the satellite NDVI time-series. One possible reason is that non-photosynthetic elements dominate (such as flowers, grass stems, and seed heads) the camera view but are less sensitive to the satellite observations (e.g., vertical elongation with lower leaf indexes)(Vrieling et al., 2018). Therefore, it is easy to find that the EOS retrieved from GCC_c precedes that retrieved from the NDVI, either with the threshold of 20% or 50%. For SOS, NDVI-based predictions are generally earlier than GCC_c-based retrievals, around 10 days. Specifically, in 2019, we found that GCC_c-based SOS20 is significantly earlier than NDVI based retrieval almost 66 days prior. As the GCC_c-based phenology observation, the winter wheat enters the growing season at the end of 2018, which is difficult to detect when using optical satellite data alone.

Fig. 10 shows Deep-CroP derived crop growth parameters in terms of SOS and EOS. Compared to the NDVI-derived phenological parameters, it is highly consistent with the phenological timing derived by the PhenoCam within 10 days of discrepancies. Therefore, it indicates that the information of S2 and S1 are efficiently combined by the Deep-CroP model, and the timing of the key growth stages is correctly determined. According to Cropland Data Layer (CDL), this site planted dry beans, winter wheat, and spring wheat in 2018, 2019, and 2020, respectively. However, the Deep-CroP model identified the beginning of crop growth for winter wheat (i.e., SOS20) in late 2018, but it was significantly later than the camera-based approach, as shown in Table 3. This is probably because the initial crop growth is quite sparse, with a much shorter plant height than the usual year. Thus, the crop growth has less impact on the SAR backscatter variation, which makes it difficult for the Deep-CroP model to use additional information for EOS timing prediction. On average, the optical-SAR combined framework may be earlier than PhenoCam observations, which is possibly due to strong response to crop structural changes of SAR time-series.

4.2.2. Camera_Site2

Fig. 11 illustrates the vegetation index trajectories of crop development at the mead3 site. It can be observed that the temporal profiles of GCC_c and NDVI are highly consistent with each other, which demonstrated the growing and senescence period for crops. As a result, the derived crop phenology (including EOS and SOS) is very close. Among them, PhenoCam retrievals were earlier than satellite retrievals for EOS50 in 2017 (25 days) and 2018 (27 days), as shown in Table 4. It can be seen that there are certain differences in their temporal profiles over these two periods. To be specific, GCC responds quickly to crop harvest with a sudden drop in the curve, while NDVI remains stable and begins to decline after a curve fitting window.

Complementary, the key crop phenological parameters derived from S1 and S2 data are shown in Fig. 12. Compared with the phenological timing derived from S2 alone, the Deep-CroP model combined the merits of S1 and S2 data to obtain phenological parameters that are close to GCC_c retrieval. For the EOS50 retrieved in 2017 and 2018, we found that phenology differences are greatly corrected by adding S1 data with the Deep-Crop model, and a more accurate timing was obtained (0 days and − 5 days, respectively). This indicates the ability of the S1 satellite with the C band is able to penetrate crops and acquire structural information. However, it can be observed that the discrepancy in EOS20 retrieved by the proposed model in 2020 is increased compared to the NDVI-based method. As Veloso et al. (2017) mentioned, the reason for this phenomenon is that the height of soybean and the number of soybean stems per surface unit are low, which resulted in a significant surface scattering from the soil and produced a poor attenuation of the backscattered signal. This makes the Deep-CroP model more sensitive to crop senescence and harvest, leading to the early determination of EOS.

4.3. Mapping large-scale crop phenology on EOS and SOS

Fig. 12 and Fig. 13 illustrate the results of applying the conventional phenological retrieval approach with optical S2 NDVI time-series for the test areas TA1 and TA2. Due to the limited seasonal variations on alfalfa, there are 50% of the total pixels with detected SOS and EOS. While, for

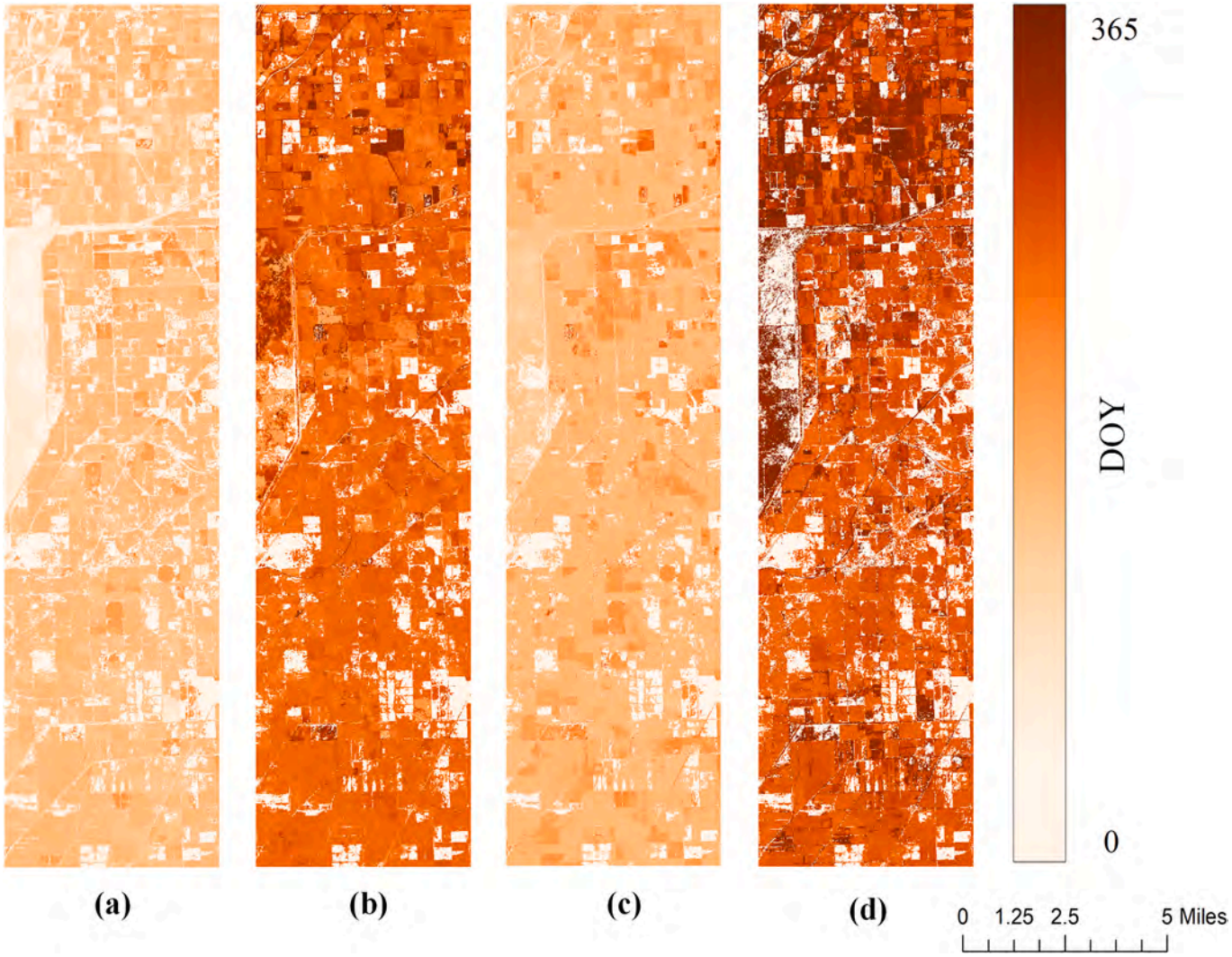

Fig. 15. Crop phenology mapping on (a)SOS20, (b)EOS20, (c)SOS50, and (d)EOS50 from Sentinel-2 NDVI time-series (based on the fitted curves with TIMESAT) for the TA2.

rice and corn-dominated area TA2, there are 80% of the available pixels are detected with SOS and EOS predictions.

Specifically, for the TA1, there is a large spatial variability in both within and between crop fields. In general, most pixels reach SOS20 before the end of April, and the senescence process begins in October. However, for the threshold of 50% setting, the SOS50 starts in June, and the EOS50 steps in September. The average dates for SOS in TA1 gradually increase from the high latitude to the low latitude. In contrast, the average dates for EOS are slowly decreased from higher latitudes to lower latitudes. Cornfields have relative uniform dates for SOS and EOS, due to the ripening and harvest in early October and green-up in mid-May. As a consequence, cornfields are clearly visible in the upper left corner of Fig. 13.

Deep-CroP model was also applied in crop phenology mapping over the TA1 and TA2, and the phenological retrieval results were shown in Fig. 17 and Fig. 20. From Fig. 13, it can be seen that the Deep-CroP model successfully retrieves more available pixels (over 75% for both SOS and EOS) than the NDVI-based curve fitting method. For the rest of the unidentified pixels, there are mainly non-agricultural land (such as water) and the confusing alfalfa. In addition, Deep-CroP derived EOS is earlier than that of the NDVI-based method, while SOS timing is similar. This proves that the Deep-CroP derived results to be consistent with GCC_c (i.e., the EOS timing of GCC_c retrieval is usually earlier than the NDVI-based retrieval).

The discrepancies between NDVI and Deep-CroP (combined with S1 and S2 data) derived phenology retrievals in the TA 1 are shown in Fig. 14. The difference maps confirm earlier findings that SOS, EOS is a few days earlier for Deep-CroP retrievals. From the discrepancy map of SOS20 and SOS50, it can be seen that the date difference between NDVI and Deep-CroP method decreased from high latitude to low latitude (The differences range from 10 to 30) for the same type of crop (e.g., corn). Specifically, the difference is near a month at high latitudes, while at low latitudes, it is only within 10 days. In contrast, the differences in EOS20 and EOS50 increase with higher latitudes. It is worth mentioning that the Deep-CroP model is able to predict fine-scale phenological maps over agricultural areas, regardless of various crop types.

Compared with TA1, the study area of TA2 has fewer crop types with limited spatial variations in SOS and EOS. On average, for most crop fields, the SOS20 starts on 5th June, EOS20 begins with 28th September, also SOS50, EOS50 on 10th June on and 13th September, respectively. As can be seen from Fig. 19, the timing of SOS crops at high-altitude is earlier than that of low latitude crops, while EOS with the opposite phenomenon.

As for the TA2, the results of phenological retrieval based on Deep-CroP are significantly different from the NDVI-based curve fitting method, which can be seen in Figs. 15 and 16. Among them, the NDVI-based SOS20 begins on 16 May, and SOS50 starts on 5 June, while EOS20 and EOS50 are initials in September. In contrast, for Deep-CroP-

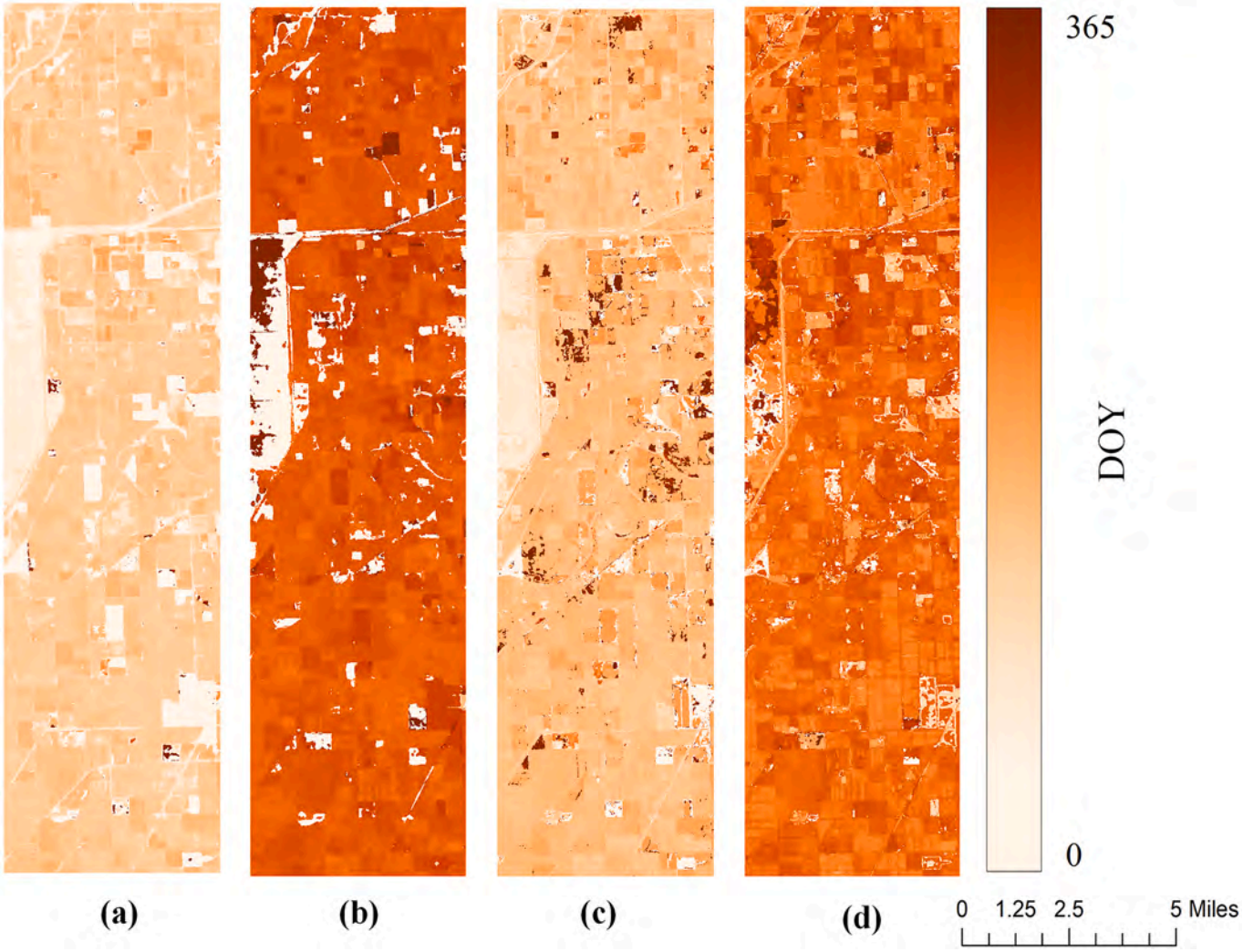

Fig. 16. Crop phenology mapping on (a)SOS20, (b)EOS20, (c)SOS50, and (d)EOS50 from Sentinel-1 and Sentinel-2 time-series with Deep-CroP for the TA2.

based retrievals, the timing of SOS20 ahead of NDVIs-based retrievals by almost 20 days.

Fig. 17 shows the differences between NDVI-based and Deep-CroP phenology retrievals in TA2. It can be seen that the differences in phenology date (the timing of SOS and EOS) between the two methods are mainly concentrated at high latitudes with −10 to −30 days. While the differences at low latitudes are generally within 10 days. The exception is that the date differences of SOS20 range from-10 to 30 in both high and low dimensions. From the CDL data, it is obvious that rice is the main crop around the high latitudes of the TA2, while soybean is the dominant crop in other regions, which may trigger unevenly distributed variations.

5. Discussion

5.1. Agreement between GCC_c- and Deep-CroP derived phenology

The feasibility of phenological retrieval with satellite time-series and PhenoCam observations has been proved with previous experimental results. However, due to the differences between PhenoCam and satellites in terms of viewing angle and range, there are certain differences between these two data sets (Vrieling et al., 2018). Therefore, the GCC calculated by the camera to be more sensitive to biomass and able to stabilize earlier in the decay phase, as compared to the satellite-based NDVI time-series filtering curves. The discrepancies of NDVI-based retrievals are on average within 17 days with in-situ GCC_c retrievals on SOS50, and EOS50, while early green-up (SOS20) and senescence (EOS20) had a mean difference value of 11.0 and 45.0 days, respectively. Two factors weigh heavily for inaccurate crop phenology tracking of traditional NDVI-based method, 1) limited observation frequency; 2) curve fitting introduces additional uncertainty, as shown in Fig. 18.

Despite differences between NDVI and PhenoCam-based phenological retrievals, we introduced S1 SAR data to add additional descriptions of crop information. The experiment results show that the accurate retrieval of crop phenology at fine spatial resolution is possible by combining Sentinel-2 and S1 imagery without curve fitting, as shown in Fig. 19. The detected phenology timing is more consistent with the ground observation network, thanks to the capture of additional information about physiology from S2 and geometry from S1. This observation agrees with Mercier et al. (2020), who found that by combining the S1 and S2 series data, the phenological stages of some crops can be more accurately identified. Nonetheless, there are still some errors, which are related to contradictory crop ripening and decay stages on various crop types. For most cases, the GCC_c-derived EOS timing is earlier than that model-derived, because the decline of GCC_c stabilizes earlier in the decay phase than that of satellite signals. However, the GCC_c-derived EOS timing is later than satellite observations occasionally, the Deep-CroP model will mistakenly judge crop senescence in advance, resulting in the earlier detection. This is partially attributed to the backscatter variations of the S1 time series during crop ripening and

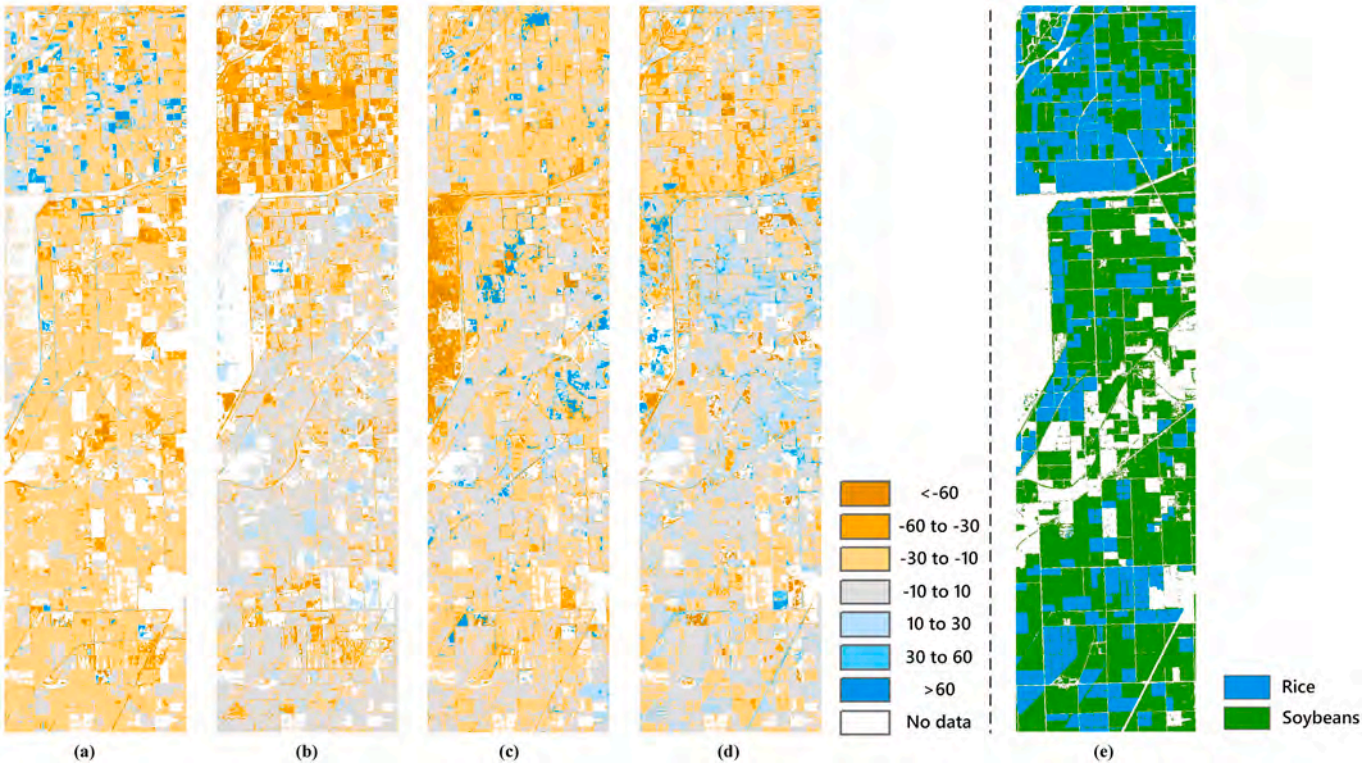

Fig. 17. Difference (in days) between phenology stages derived from curve fitting (Sentinel-2 NDVI series) and Deep-CroP (Sentinel-2 and Sentinel-1 series) for (a) SOS20, (b)EOS20, (c)SOS50, (d)EOS50, (e)2018 CDL data in TA2. Negatives mean that the Deep-CroP-derived resulting in earlier dates than NDVI-based method.

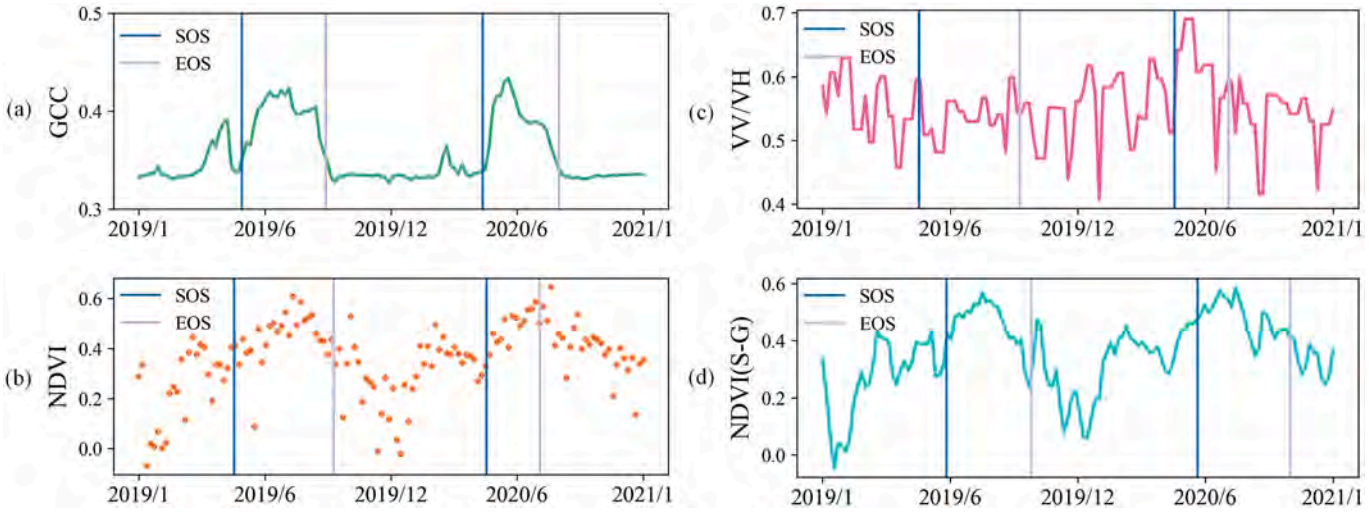

Fig. 18. The Discrepancies of crop phenology stages derived from the traditional curve fitting method (a) and Deep-CroP (b, and c) on SOS20, EOS20. (d) the SOS/EOS prediction with Savitzky-Golay (S-G) NDVI time-series filtering. Compare to the reference label (a), the Deep-CroP is more accurate than curve fitting method.

decay stages which confuse the Deep-CroP model inaccurate phenology identification.

5.2. Contribution of opitcal-SAR multi-modular scheme

In this study, a spatial-aware deep learning method was used to combine SAR and optical data for phenological retrieval. This reduces the differences in the phenological stages observed by the ground sites and satellites, making it possible to establish a physical co-relation between ground sites and satellite time series over crop areas. Compared to the optical-based crop phenology retrieval, the fusion of multi-source satellite achieved better consistency with site GCC_c data (as shown Fig. 22). This suggests that the additional S1 features provide an important contribution to the phenological stages retrieval. However, we found that in some cases, the model showed poorer performance than the conventional method. In order to further explore the contribution of SAR time-series to the Deep-CroP model, we designed additional experiments with a different configuration of input data (i.e., the single optical data v.s. Optical-SAR data) in this section. For the model with single optical data input, we modified the Deep-CroP model to remove the branch of SAR data processing, which it was simply denoted as Deep-CroP_optic, and the rest setting of the model keeps unchanged.

Fig. 20 and Fig. 21 demonstrate the differences between Deep-CroP and Deep-CroP_optic over the four test sites. It can be found that SAR

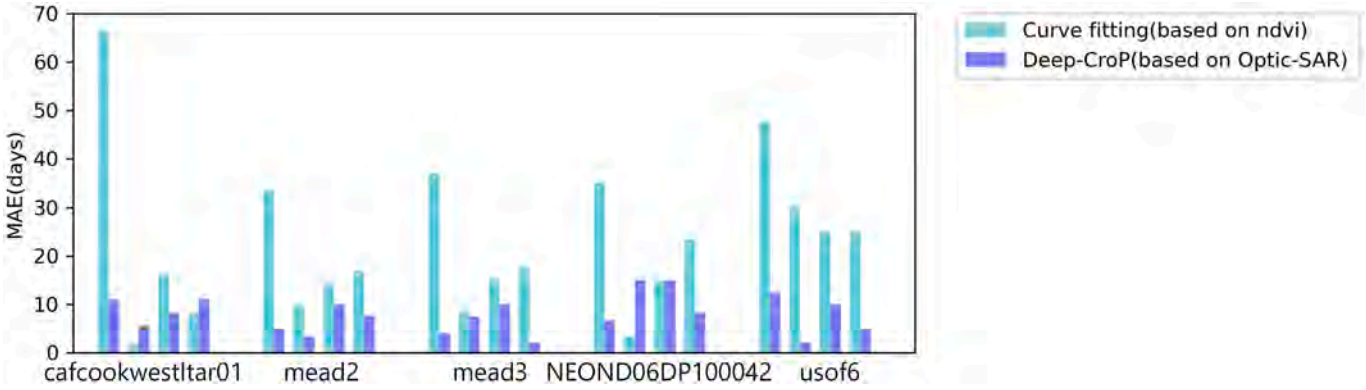

Fig. 19. MAE between GCC_c- and Deep-CroP/NDVI-derived LSP in sites, including cafcookwestlar01, mead3, NEOND06DP100042, and usof6. The bar for each site is listed from left to right in the order SOS20, EOS20, SOS50, and EOS50.

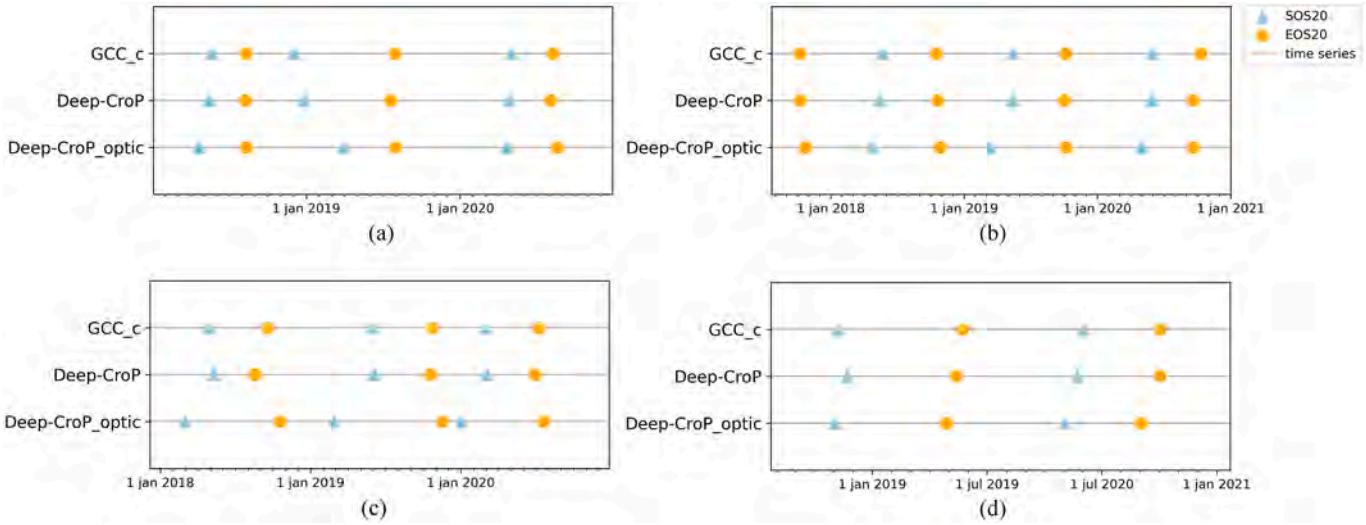

Fig. 20. Comparison of SOS20 and EOS20 derived from GCC_c, Deep-CroP(with SAR-optical fusion), and Deep-CroP _optic (with optical alone)on 4 test sites.

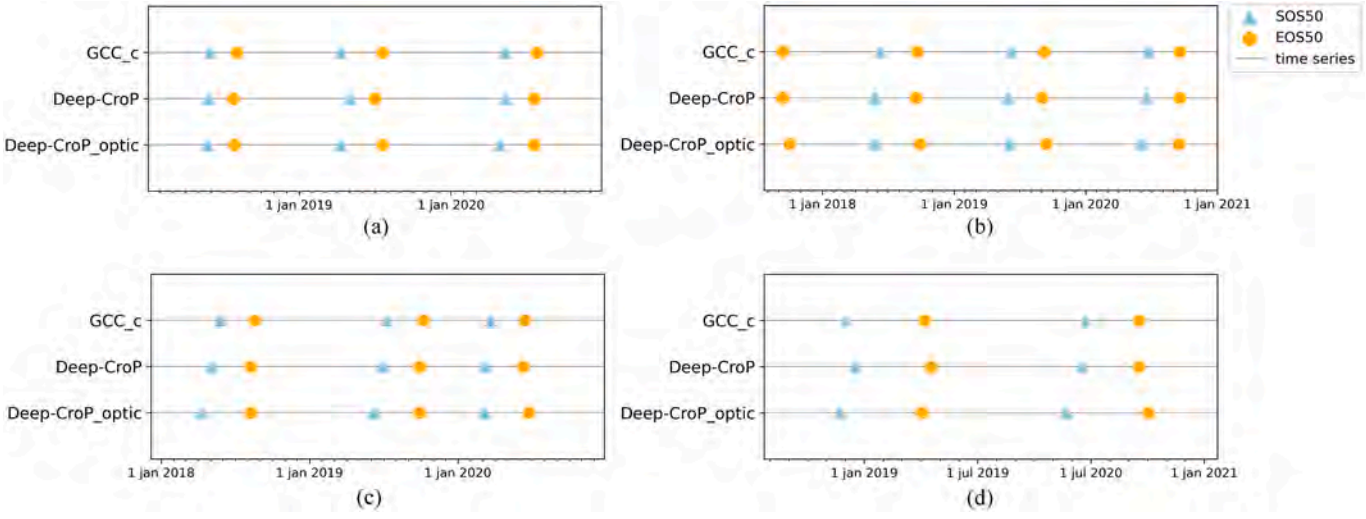

Fig. 21. Comparison of SOS50 and EOS50 derived from GCC_c, Deep-CroP (with SAR-optical fusion), and Deep-CroP _optic (with optical alone) on 4 test sites.

information makes an important contribution to the retrieval of EOS20 and SOS20, which helps the model accurately identify phenological stages over different crops. It's worth noting that the timing of SOS20 derived from single optical data is later than that of Deep-CroP retrieval in the second phenological cycle on site 1. This verified the previous assumption that the information provided by SAR data could help the model capture abrupt phenological events. In the results of EOS50 and SOS50, it is found that SAR can still help the model to identify

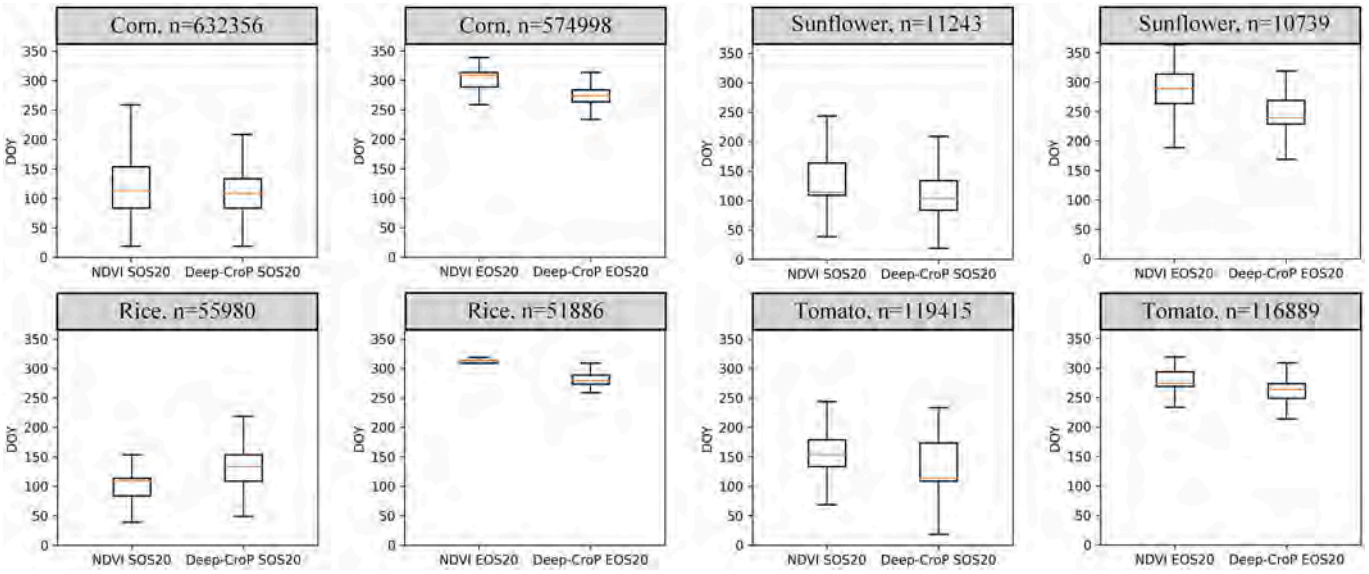

Fig. 22. Crop phenology (SOS20 and EOS20) variations acquired from curve fitting (based on NDVI) and Deep-CroP model for the TA1 with main crop types. The selection of pixels is based on the CDL data, and the pixels that are successfully retrieved by both curve fitting (based on NDVI) and Deep-CroP are selected.

Table 5
The Mean Absolute Error (MAE) for the satellite-based SOS/EOS timing based on CNN and Transformer model derived crop phenological parameters compared to the PhenoCam observations at cafcookwestltar01 test sites.

Methods	SOS20	EOS20	SOS50	EOS50
CNN	34	6.6	10	13.3
Transformer	45	9.3	11	10.6
Deep-CroP	11.0	5.6	8.3	11.0

phenological stages, but sometimes it may lead to a slight deviation. Also, to validate the impact of Deep-CroP integration architecture, the crop phenological parameter extraction with the CNN and transformer architecture were conducted, respectively. From Table 5, we can conclude that the spatial-aware CNN architecture outperforms the transformer architecture in terms of SOS and EOS extraction. The reason is mainly that spatial-aware features extract from image patches are much more robust than single pixel-values that used in transformer architecture. Although, the transformer has the ability to find recurrent patterns in time-series data, still, the random noises in both spectral-backscatter domains result in worse predictions. Therefore, the integration of the CNN and transformer architecture is essential to the precise prediction of crop phenological parameters.

In addition, we note that the performance of the proposed model deteriorates, confirming the earlier idea that SAR information may confuse crop phenology stages occasionally with complex backscatter information. Polarimetric decomposition can discriminate useful crop-related components from complex backscatter signals. In this scope, polarimetric parameters have proven to be effective in crop biophysical condition retrieving. Although, the correlation between polarimetric parameters and crop biophysical indicators is generally weak, still, it is worth noting that polarimetric parameters in time-series could be another breakthrough in future studies (Harfenmeister et al., 2021). Moreover, the performance of the Deep-CroP model relies on the availability of optical remote sensing, since the training labels are matched by finding adjacent observations.

5.3. Impact of crop types on phenology identification

For optical satellite data, it combines biophysical parameters including canopy cover, biomass amount, and leaf chlorophyll concentration, while SAR backscatter is affected by factors related to crop biomass, structure, and ground conditions. Additionally, Sentinel-1 with C-band is a combination of the ground backscatter attenuated by the canopy layer and the backscatter from the canopy. Thus, the contribution of SAR time-series is not only dependent on the complexity of background information, the crop types also significantly impact the performance of accurate crop phenology retrieval.

The differences in crop timing of greening, water content, and structure may result in poor temporal consistency even with Sentinel-1 and Sentinel-2 time-series. For example, NDVI is sensitive to initial green leaf development, whereas the height of crops is not significant at this time, resulting in the flat curves of the SAR backscatter time series signal. This means that the NDVI of some crops responds to crop growth earlier than that of the SAR. Still, SAR is beneficial to improve the consistency between satellite and camera observations, as past studies demonstrated that the EOS timing of NDVI based phenological retrieval is earlier than that of PhenoCam in most cases. The reason for this phenomenon is that the presence of a non-green top layer will inevitably reduce the visibility of green vegetation to the camera in the greening stage. The introduction of SAR helps to correct the premature response of NDVI. Similarly, at the end of the crop cycle, due to the decrease of chlorophyll content and water content, both the NDVI and the SAR are characterized by a steady decrease until harvest, as a consequence of the decreasing chlorophyll and water content. Combining physiology from S2 and additional geometry information from S1, it increases the accuracy of identifying the crop senescence. However, NDVI decreased rapidly during the senescence or harvest of some crops, while SAR would still respond to the standing structure of standing green residues remaining on the field and then decreased gradually until the green residues dried out. Nevertheless, the scattering mechanisms are in general much more complex, and experimental observations are needed to provide insights into the scattering behavior of each crop type.

To further dive into this issue, we analyzed the differences between the same crops in terms of the phenological retrieval results with Deep-CroP and NDVI in TA1 (Fig. 22). We noted that for most crop types, Deep-CroP derived EOS20 was earlier than NDVI, which confirmed the expected timing differences. In addition, differences in SOS timing are related to crop types. For example, for rice, the Deep-CroP derived SOS timing is later than that of NDVI, and the average time for corn is similar, while for other crops, it is earlier than that of NDVI.

6. Conclusion

In this work, we proposed the Deep-CroP model to retrieve crop phenology parameters providing new opportunities for fine-scale phenology research. Aiming to match satellites and ground observations, we developed the spatial-aware mechanism to overcome the discrepancy between the viewing angle and range of the two. To be specific, we integrate Sentinel SAR and optical time-series data by introducing additional vegetation structural information to improve the consistency between satellite and ground observations. The results on test sites show that the method we proposed can accurately identify crop phenology stages and achieve high consistency with GCC_c time-series. Then, we applied the Deep-CroP model on two large-scale study sites and found that the crop phenology obtained by the proposed model has satisfying results compared to ground observations. In general, the deep fusion of optics and SAR time-series greatly improves the observation consistency of satellites and phenological cameras, which are expected to help improve existing crop monitoring systems.

Also, some limitations of the proposed approach must be stressed, such as the prediction accuracy of the Deep-CroP model is limited due to the available size of PhenoCam samples. To be specific, both the time span and data quality of the PhenoCam site limited the performance of the model. Another weakness of the proposed method is its high computational cost, which may take more time when calculating long time series given a large-scale area.

Credit statement

Wenzhi Zhao: Conceptualization, Investigation, Methodology, Result Analysis, Funding acquisition;

Yang Qu: Data labeling, Experiment design, Result Analysis;

Liqiang Zhang: Methodology, Result Analysis, Writing- Reviewing and Editing.

Kaiyuan Li: Data acquiring, Data labeling.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was supported by the National Natural Science Foundation of China Major Program under Grant (42192580, 42192584), the GF Project (31-Y30B09-9001-20/22-04, 31-Y30F09-9001-20/22-13, 31-Y30F09-9001-20/22-14) and the Natural Science Foundation of Beijing Municipality under Grant 4214065. Also, we appreciate two anonymous reviewers that have substantially improved this paper in terms of scientific contributions and results analysis.

References

Ajadi, O.A., Barr, J., Liang, S.-Z., Ferreira, R., Kumpatla, S.P., Patel, R., Swatantran, A., 2021. Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery. Int. J. Appl. Earth Obs. Geoinf. 97, 102294.

Beck, P.S., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sens. Environ. 100, 321–334.

Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. Remote Sens. Environ. 240, 111685.

Brown, T.B., Hultine, K.R., Steltzer, H., Denny, E.G., Denslow, M.W., Granados, J., Henderson, S., Moore, D., Nagai, S., SanClements, M., 2016. Using phenocams to monitor our changing earth: toward a global phenocam network. Front. Ecol. Environ. 14, 84–93.

Burke, M.W., Rundquist, B.C., 2021. Scaling Phenocam GCC, NDVI, and EVI2 with harmonized Landsat-sentinel using Gaussian processes. Agric. For. Meteorol. 300, 108316.

De Beurs, K.M., Henebry, G.M., 2005. Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high-latitude transects. Glob. Chang. Biol. 11, 779–790.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale arXiv preprint arXiv: 2010.11929.

Ferrazzoli, P., Paloscia, S., Pampaloni, P., Schiavon, G., Solimini, D., Coppo, P., 1992. Sensitivity of microwave measurements to vegetation biomass and soil moisture content: a case study. IEEE Trans. Geosci. Remote Sens. 30, 750–756.

Ganguly, S., Friedl, M.A., Tan, B., Zhang, X., Verma, M., 2010. Land surface phenology from MODIS: characterization of the collection 5 global land cover dynamics product. Remote Sens. Environ. 114, 1805–1816.

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sens. Environ. 188, 9–25.

Harfenmeister, K., Itzerott, S., Weltzien, C., Spengler, D., 2021. Agricultural monitoring using polarimetric decomposition parameters of sentinel-1 data. Remote Sens. 13, 575.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Hufkens, K., Basler, D., Milliman, T., Melaas, E.K., Richardson, A.D., 2018. An integrated phenology modelling framework in R. Methods Ecol. Evol. 9, 1276–1285.

Ienco, D., Interdonato, R., Gaetano, R., Minh, D.H.T., 2019. Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogramm. Remote Sens. 158, 11–22.

Inglada, J., Vincent, A., Arias, M., Marais-Sicre, C., 2016. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. Remote Sens. 8, 362.

Jentsch, A., Kreyling, J., Boettcher-Treschkow, J., Beierkuhnlein, C., 2009. Beyond gradual warming: extreme weather events alter flower phenology of European grassland and heath species. Glob. Chang. Biol. 15, 837–849.

Jiao, X., Kovacs, J.M., Shang, J., McNairn, H., Walters, D., Ma, B., Geng, X., 2014. Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data. ISPRS J. Photogramm. Remote Sens. 96, 38–46.

Jin, X., Yang, G., Xu, X., Yang, H., Feng, H., Li, Z., Shen, J., Lan, Y., Zhao, C., 2015. Combined multi-temporal optical and radar parameters for estimating LAI and biomass in winter wheat using HJ and RADARSAR-2 data. Remote Sens. 7, 13251–13272.

Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845.

Jung, J., Maeda, M., Chang, A., Bhandari, M., Ashapure, A., Landivar-Bowles, J., 2021. The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. Curr. Opin. Biotechnol. 70, 15–22.

Justice, C.O., Townshend, J., Holben, B., Tucker, e.C., 1985. Analysis of the phenology of global vegetation using meteorological satellite data. Int. J. Remote Sens. 6, 1271–1318.

Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D., Morisette, J., 2002. An overview of MODIS land data processing and product status. Remote Sens. Environ. 83, 3–15.

Klosterman, S., Hufkens, K., Gray, J., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M., Richardson, A., 2014. Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. Biogeosciences 11, 4305–4320.

Liu, Y., Wu, C., 2020. Understanding the role of phenology and summer physiology in controlling net ecosystem production: a multiscale comparison of satellite, PhenoCam and eddy covariance data. Environ. Res. Lett. 15, 104086.

Lyu, H., Lu, H., Mou, L., 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sens. 8, 506.

Marzialetti, F., Giulio, S., Malavasi, M., Sperandii, M.G., Acosta, A.T.R., Carranza, M.L., 2019. Capturing coastal dune natural vegetation types using a phenology-based mapping approach: the potential of Sentinel-2. Remote Sens. 11, 1506.

McNairn, H., Brisco, B., 2004. The application of C-band polarimetric SAR for agriculture: a review. Can. J. Remote. Sens. 30, 525–542.

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D., Hubert-Moy, L., 2020. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. ISPRS J. Photogramm. Remote Sens. 163, 231–256.

Meroni, M., d'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., Seguini, L., Verhegghen, A., 2021. Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and-2. Remote Sens. Environ. 253, 112232.

Misra, G., Cawkwell, F., Wingler, A., 2020. Status of phenological research using Sentinel-2 data: a review. Remote Sens. 12, 2760.

Mou, L., Bruzzone, L., Zhu, X.X., 2018. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. IEEE Trans. Geosci. Remote Sens. 57, 924–935.

Nasahara, K.N., Nagai, S., 2015. Development of an in situ observation network for terrestrial ecological remote sensing: the Phenological Eyes Network (PEN). Ecol. Res. 30, 211–223.

Qu, Y., Zhao, W., Yuan, Z., Chen, J., 2020. Crop mapping from Sentinel-1 polarimetric time-series with a deep neural network. Remote Sens. 12, 2493.

Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Jenkins, J.P., Ollinger, S.V., 2009. Near-surface remote sensing of spatial and temporal variation in canopy phenology. Ecol. Appl. 19, 1417–1428.

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agric. For. Meteorol. 169, 156–173.
Richardson, A.D., Hufkens, K., Milliman, T., Aubrecht, D.M., Chen, M., Gray, J.M., Johnston, M.R., Keenan, T.F., Klosterman, S.T., Kosmala, M., 2018. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. Sci. Data 5, 1–24.
Rußwurm, M., Korner, M., 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 11–19.
Schnelle, F., Volkert, E., 1964. Internationale phänologische gärten Stationen eines grundnetzes für internationale phänologische beobachtungen. Agric. Meteorol. 1, 22–29.
Seyednasrollah, B., Milliman, T., Richardson, A.D., 2019. Data extraction from digital repeat photography using xROI: an interactive framework to facilitate the process. ISPRS J. Photogramm. Remote Sens. 152, 132–144.
Shao, Z., Cai, J., Fu, P., Hu, L., Liu, T., 2019. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. Remote Sens. Environ. 235, 111425.
Stendardi, L., Karlsen, S.R., Niedrist, G., Gerdol, R., Zebisch, M., Rossi, M., Notarnicola, C., 2019. Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. Remote Sens. 11, 542.
Thenkabail, P.S., Hanjra, M.A., Dheeravath, V., Gumma, M., 2010. A holistic view of global croplands and their water use for ensuring global food security in the 21st century through advanced remote sensing and non-remote sensing approaches. Remote Sens. 2, 211–261.
Tian, F., Cai, Z., Jin, H., Hufkens, K., Scheifinger, H., Tagesson, T., Smets, B., Van Hoolst, R., Bonte, K., Ivits, E., 2021. Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe. Remote Sens. Environ. 260, 112456.
Ulaby, F.T., 1982. Microwave remote sensing active and passive. In: Rader Remote Sensing and Surface Scattering and Emission Theory, pp. 848–902.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008.
Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. Remote Sens. Environ. 199, 415–426.
Vrieling, A., Meroni, M., Darvishzadeh, R., Skidmore, A.K., Wang, T., Zurita-Milla, R., Oosterbeek, K., O'Connor, B., Paganini, M., 2018. Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. Remote Sens. Environ. 215, 517–529.
Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. 236, 111402.
White, M.A., de Beurs, K.M., Didan, K., Inouye, D.W., Richardson, A.D., Jensen, O.P., O'keefe, J., Zhang, G., Nemani, R.R., van Leeuwen, W.J., 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. Glob. Chang. Biol. 15, 2335–2359.
Wingate, L., Ogée, J., Cremonese, E., Filippa, G., Mizunuma, T., Migliavacca, M., Moisy, C., Wilkinson, M., Moureaux, C., Wohlfahrt, G., 2015. Interpreting canopy development and physiology using a European phenology camera network at flux sites. Biogeosciences 12, 5995–6015.
Yuan, W., Chen, Y., Xia, J., Dong, W., Magliulo, V., Moors, E., Olesen, J.E., Zhang, H., 2016. Estimating crop yield using a satellite-based light use efficiency model. Ecol. Indic. 60, 702–709.
Zeng, L., Wardlow, B.D., Xiang, D., Hu, S., Li, D., 2020. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sens. Environ. 237, 111511.
Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475.
Zhao, B., Zhong, Y., Zhang, L., 2016. A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 116, 73–85.
Zhou, Q., Rover, J., Brown, J., Worstell, B., Howard, D., Wu, Z., Gallant, A.L., Rundquist, B., Burke, M., 2019. Monitoring landscape dynamics in central us grasslands with harmonized Landsat-8 and Sentinel-2 time series data. Remote Sens. 11, 328.