ISPRS Journal of Photogrammetry and Remote Sensing 215 (2024) 271–291

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

EMET: An emergence-based thermal phenological framework for near real-time crop type mapping

Zijun Yang [a], Chunyuan Diao [a,*], Feng Gao [b], Bo Li [c]

[a] Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[b] United States Department of Agriculture, Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705, USA
[c] Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Keywords:
Crop mapping
Crop phenology
Near real-time
Deep learning
Agriculture

ABSTRACT

Near real-time (NRT) crop type mapping plays a crucial role in modeling crop development, managing food supply chains, and supporting sustainable agriculture. The low-latency updates on crop type distribution also help assess the impacts of weather extremes and climate change on agricultural production in a timely fashion, aiding in identification of early risks in food insecurity as well as rapid assessments of the damage. Yet NRT crop type mapping is challenging due to the obstacle in acquiring timely crop type reference labels during the current season for crop mapping model building. Meanwhile, the crop mapping models constructed with historical crop type labels and corresponding satellite imagery may not be applicable to the current season in NRT due to spatiotemporal variability of crop phenology. The difficulty in characterizing crop phenology in NRT remains a significant hurdle in NRT crop type mapping. To tackle these issues, a novel emergence-based thermal phenological framework (EMET) is proposed in this study for field-level NRT crop type mapping. The EMET framework comprises three key components: hybrid deep learning spatiotemporal image fusion, NRT thermal-based crop phenology normalization, and NRT crop type characterization. The hybrid fusion model integrates super-resolution convolutional neural network (SRCNN) and long short-term memory (LSTM) to generate daily satellite observations with a high spatial resolution in NRT. The NRT thermal-based crop phenology normalization innovatively synthesizes within-season crop emergence (WISE) model and thermal time accumulation throughout the growing season, to timely normalize crop phenological progress derived from temporally dense fusion imagery. The NRT normalized fusion time series are then fed into an advanced deep learning classifier, the self-attention based LSTM (SAtLSTM) model, to identify crop types. Results in Illinois and Minnesota of the U.S. Corn Belt suggest that the EMET framework significantly enhances the model scalability with crop phenology normalized in NRT for timely crop mapping. A consistently higher overall accuracy is yielded by the EMET framework throughout the growing season compared to the calendar-based and WISE-based benchmark scenarios. When transferred to different study sites and testing years, EMET maintains an advantage of over 5% in overall accuracy during early- to mid-season. Moreover, EMET reaches an overall accuracy of 85% a month earlier than the benchmarks, and it can accurately characterize crop types with an overall accuracy of 90% as early as in late July. F1 scores for both corn and soybeans also achieve 90% around late July. The EMET framework paves the way for large-scale satellite-based NRT crop type mapping at the field level, which can largely help reduce food market volatility to enhance food security, as well as benefit a variety of agricultural applications to optimize crop management towards more sustainable agricultural production.

1. Introduction

Crop type information is essential to understand the global food supply and support sustainable agricultural development over space and time. In particular, near real-time (NRT) crop type mapping, through the provision of crop species distribution estimates with low latency (e.g., one to two days) and frequent updates (e.g., daily to weekly), can help predict potential food shortfalls or surpluses and provide valuable insights for supply chain management, insurance design, and food market volatility in a timely fashion (Cai et al., 2018; Yang et al., 2023). NRT

* Corresponding author.
E-mail address: chunyuan@illinois.edu (C. Diao).

https://doi.org/10.1016/j.isprsjprs.2024.07.007
Received 8 December 2023; Received in revised form 15 June 2024; Accepted 6 July 2024
Available online 16 July 2024

crop type mapping can also facilitate a wide range of applications, including crop acreage estimation, crop phenology characterization, and yield predictions, within the growing season (Bégué et al., 2018; Yang et al., 2023; Zhang and Diao, 2023). With altered patterns of temperature and precipitation, and increased frequencies of extreme weather events, global food security has been increasingly challenged by climate change events (Pörtner et al., 2022; Prosekov and Ivanova, 2018). Through early characterization of crop type distribution, NRT crop mapping can provide important data for assessing the impacts of climate change and weather extremes on agricultural production throughout the growing season. These insights may further aid in early warning systems and risk management by helping identify early signs of potential food insecurity risks and provide rapid assessments of the damage, enhancing food security under changing climate conditions (Cai et al., 2018; Gao et al., 2023; Gao and Zhang, 2021).

Crop type mapping models typically utilize machine learning techniques to analyze diverse types of remote sensing imagery for identifying and classifying crop species of agricultural fields. A range of metrics (e.g., reflectance bands, vegetation indices [VIs], gray level co-occurrence matrix [GLCM], seasonal amplitude of VIs, length of the growing season) extracted from satellite remote sensing have been employed to differentiate crop types. These metrics, spanning across spectral, structural, textual, and temporal dimensions, are used to characterize phenological, biophysical, and physiological properties of crop species (Feng et al., 2019; Nguyen et al., 2020; Wang et al., 2019; Zhong et al., 2019). Particularly, the temporal crop phenological patterns extracted from remote sensing time series have been deemed among the most critical to identifying crop types (Zhong et al., 2019). Affected by a combination of weather, environmental, and management conditions, crop species usually maintain unique temporal phenological development patterns throughout the growing season (e.g., timing of emergence and flowering). Such temporal phenological patterns can be represented via phenological metrics or be modeled with machine (or deep) learning techniques from time-series satellite images (Diao, 2020; Diao et al., 2021). The phenological metrics can be extracted using threshold-based (White et al., 2009), derivative-based (Tan et al., 2010), curvature-based (Klosterman et al., 2014; Zhang et al., 2003), Gu-based (Gu et al., 2009), pheno-network (Diao, 2019), and curve fitting parameter-based (Soudani et al., 2008; Zhong et al., 2014) methods. Besides phenological metrics, several deep learning architectures (e.g., long short-term memory [LSTM], one-dimensional convolutional neural network [1-DCNN], and transformer) have been increasingly utilized to learn temporal patterns of crop species from high-dimensional time-series remote sensing data (Pelletier et al., 2019; Xu et al., 2020; Zhong et al., 2019). For example, the crop-specific temporal patterns can be modeled using the memory cell and gating mechanism (i.e., input gate, forget gate, and output gate) by LSTM, convolutional operations across the temporal dimension by 1-DCNN, as well as the self-attention mechanism by transformer (Hochreiter and Schmidhuber, 1997; Pelletier et al., 2019; Rußwurm and Körner, 2020; Xu et al., 2021; Zhong et al., 2019).

Despite the progress made in detecting crop species with characteristic temporal phenological patterns, NRT crop type mapping remains challenging. Most phenological metrics rely on satellite observations throughout the growing season, which makes these metrics difficult to be retrieved in NRT (You et al., 2023). Deep learning models, with the flexibility to learn feature representations from varying lengths of satellite time series, largely facilitate NRT applications. However, the crop type ground reference labels for training the deep learning models can be difficult to acquire within season, particularly at large scales. The limited availability of crop labels significantly hampers the model development during the growing season (Wang et al., 2019; Xu et al., 2020). To address the aforementioned issues, a favored strategy adopted by existing studies is to train deep learning classifiers using satellite time series from the past growing seasons alongside corresponding historical ground reference data (e.g., Cropland Data Layer [CDL] in the US) (Johnson and Mueller, 2021; Xu et al., 2021; Xu et al., 2020; Yaramasu et al., 2020; Zhu et al., 2017). The pre-trained model can then be applied to satellite observations of the current growing season to obtain crop type predictions (Azar et al., 2016; Cai et al., 2018; Johnson and Mueller, 2021; Konduri et al., 2020; Xu et al., 2020). Most of these models employ the calendar-based modeling strategy which assumes that crop growth patterns are comparable in terms of calendar dates across years and locations. Yet such assumptions may not be met due to the spatiotemporal difference in climatic and environmental conditions, which may lead to varying farming practices (e.g., planting timing) and crop phenological progress over space and time (Kerner et al., 2022; Lin et al., 2022; Zhong et al., 2014; Zhong et al., 2016). Recognizing the challenge of model transferability from past to current growing seasons, recent efforts have been devoted to reducing the differences between the distributions of satellite data and/or crop type statistics acquired in different years and locations (e.g., through feature shift or domain adaptation) (Kluger et al., 2021; Wang et al., 2023; Wang et al., 2022). While these methods demonstrate improved model scalability, potential large differences between the source and target distributions may be difficult to accommodate (Xu et al., 2024). It remains a challenge to effectively communicate the mechanism of crop phenology variability with crop mapping models, which holds the potential for accommodating the spatiotemporal variation in phenology and enhancing model scalability.

Given that most phenological metrics need to be extracted retroactively with season-long satellite observations, NRT acquisition of crop phenology remains a significant challenge, particularly when crop type information is not available. The efforts towards satellite-based NRT phenology characterization are still limited (Gao and Zhang, 2021). Among the limited efforts, the curve-based and trend-based phenology approaches have been devised to detect crop phenology using satellite imagery in NRT. The curve-based phenology approach estimates the NRT crop phenology by associating the current-season satellite time series with historical satellite observations alongside their corresponding phenological metrics in past seasons. (Liao et al., 2023; Liu et al., 2018; Shen et al., 2023). Yet this approach usually requires crop type labels for building the relationship between the current and past season satellite observations, which makes it hardly applicable during the early season when crop type information may not be available. The trend-based phenology approach estimates the crop phenology through identifying trend changes of VI curves using only the satellite observations in the current growing season (Gao et al., 2020a; Gao et al., 2020b). The trend changes of VI curves associated with crop emergence can be identified in NRT using the trend-based within-season emergence (WISE) model. Compared to the curve-based phenology approach, the trend-based approach has lessened requirement of both crop type labels and historical satellite observations, making it promising in characterizing the phenological variation in NRT for corresponding crop type mapping. Yet the trend-based phenology approach is more tailored for detecting the NRT early-season phenology (i.e., crop emergence). As the crop phenological development throughout the growing season is subject to the influence of environmental and climatic conditions (e.g., temperature in particular), the limited ability of the trend-based approach in timely accommodating the whole-season phenological dynamics remains a significant hurdle for scalable NRT crop type mapping.

To achieve NRT characterization of crop phenology and crop types at the field level, the dense time-series remote sensing data of both high spatial and temporal resolutions are needed (Gao et al., 2017). Yet most satellite datasets (e.g., MODIS, Landsat, and Sentinel-2) still suffer from the trade-off between spatial and temporal resolutions, and the frequent cloudy and rainy weather events in many agricultural systems further limit the availability of satellite time series. Those datasets may not be able to capture the rapid phenological changes during the crop growing season. The limited data availability during critical growing windows has also been found to negatively affect crop type mapping models with degraded classification performance (Cai et al., 2018; Xu et al., 2020). A

feasible solution to this data availability issue is via the spatiotemporal satellite image fusion, which blends satellite datasets of different resolution characteristics to generate consecutive remote sensing observations with both high spatial and temporal resolutions (Zhu et al., 2018). The spatiotemporal image fusion models can be categorized into the following groups: weight function-based, unmixing-based, learning-based, Bayesian-based, and hybrid methods (Gao et al., 2006; Song et al., 2018; Zhukov et al., 1999). Despite a variety of fusion models being developed, rapid and drastic temporal changes among imagery remain challenging to be predicted in most spatiotemporal image fusion studies. Many fusion models also require satellite images both before and after the target prediction date, which may not match the purpose of NRT data acquisition (Yang et al., 2021). Thus, an adequate design of spatiotemporal image fusion models that can tackle these issues is critical in the context of NRT field-level crop type mapping.

The objective of this study is to develop a novel EMErgence-based Thermal phenological framework (EMET) for NRT crop type mapping at the farm field level. The EMET framework synthesizes spatiotemporal data fusion, crop phenology normalization, and advanced deep learning to achieve scalable NRT crop mapping. In particular, an innovative NRT thermal-based phenological approach is proposed to normalize and register the crop growth progress throughout the growing season in a timely fashion. Specifically, we aim at 1) integrating spatiotemporal data fusion into NRT crop type mapping with a hybrid deep learning spatiotemporal fusion model; 2) devising an innovative NRT thermal-based crop phenology normalization approach by synthesizing NRT crop emergence characterization with thermal time accumulation throughout the season; 3) evaluating the influence of different levels of phenology normalization on NRT crop mapping performance across years and locations.

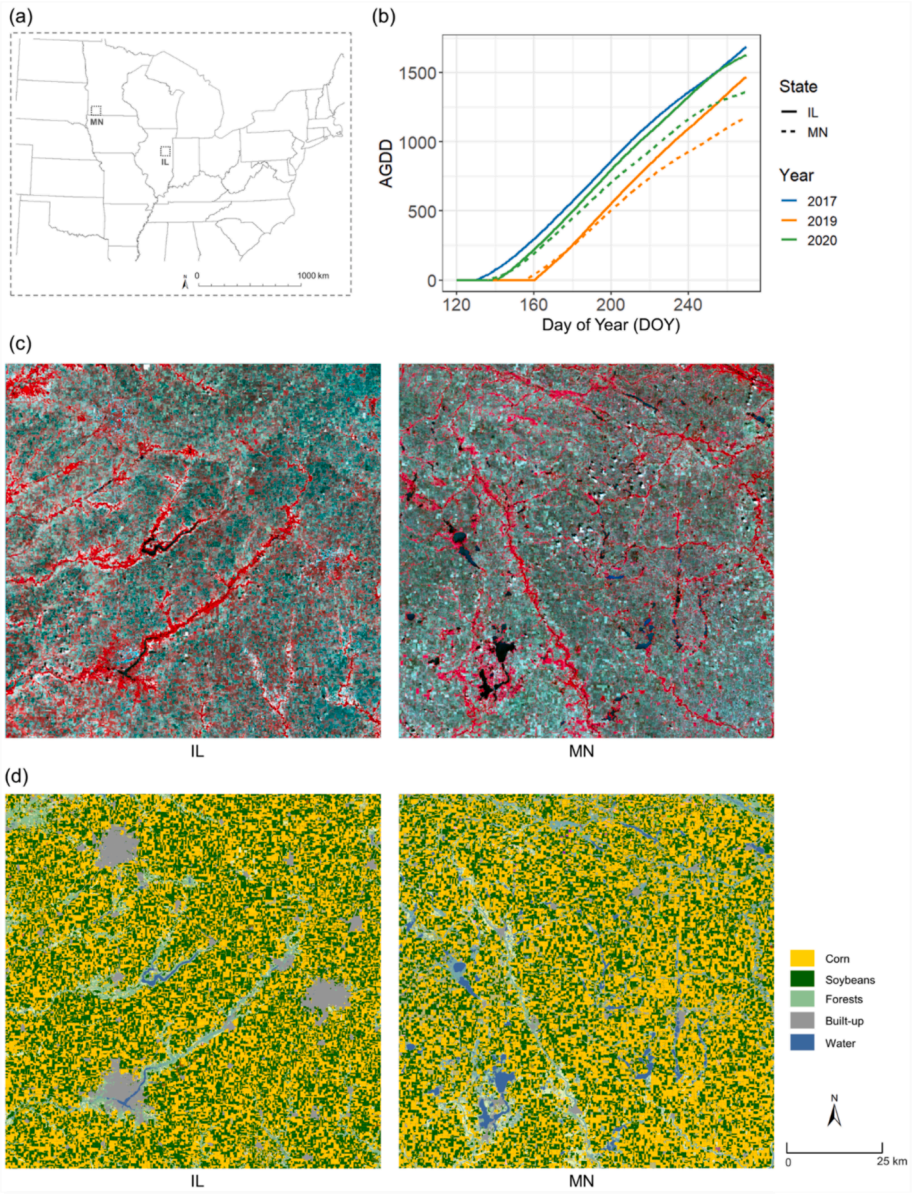

Fig. 1. (a) The geographic locations of the Illinois (IL) and Minnesota (MN) study sites; (b) The heat accumulation (measured by accumulated growing degree days [AGDD]) for the states of IL and MN in the training (2017) and testing (2019 and 2020) years from early- to mid-season; Heat accumulation approximates the amount of energy to be used for crop development. The variation in heat accumulation across space and time is substantial; (c) The early-season HLS tiles for the IL and MN study sites in false color composite, respectively; (d) The corresponding CDL data for the IL and MN study sites, respectively.

2. Data and materials

2.1. Study sites

This study is focused on the U.S. Corn Belt. The Corn Belt is an important global food production area, with corn and soybeans being the two predominant crop species. Specifically, we select two study sites located in the U.S. Corn Belt, namely eastern Illinois (IL) and southwestern Minnesota (MN) (Fig. 1a). Both study sites are mainly covered by agricultural fields (corn and soybeans), with a small fraction of other land cover classes such as forests, water bodies, and built-up areas. Crop phenological progress varies between the two spatially separate sites, with a discrepancy of about or over one week. Apart from the spatial phenological difference, substantial inter-annual variations in crop phenology progress are observed among different years (e.g., variation in crop emergence can be as large as four weeks across years) (USDA-NASS, 2022). To examine the influence of spatiotemporal crop phenological variations on NRT crop type mapping, the crop type mapping model is first built with the 2017 data in the IL study site, as 2017 is a relatively normal year in terms of meteorological conditions. The model's ability to accommodate inter-annual phenological variations will be tested in two other years, namely 2019 and 2020. These two years are selected because excessive precipitation events in spring 2019 caused considerable delay in crop planting and subsequent crop phenological progress, while 2020 is a year with relatively more similar meteorological conditions to the training year 2017 (Fig. S1). Both the IL and MN study sites are tested for assessing the transferability of the model across different locations (Fig. 1b).

2.2. Datasets

The high spatial- and temporal-resolution fusion images are employed as the primary source of satellite images for NRT crop type mapping in this study. The fusion data are generated from the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD43A4 nadir Bidirectional Reflectance Distribution Function (BRDF) Adjusted Reflectance (NBAR) dataset (Schaaf and Wang, 2015) and Harmonized Landsat Sentinel-2 (HLS) dataset (Claverie et al., 2018). The MODIS MCD43A4 products have a spatial resolution of 500 m with daily observations. The NBAR processing mitigates the effects of viewing angles on illumination and reflectance. The reflectance of MCD43A4 at each date is generated through modeling the satellite observations from both Terra and Aqua during its surrounding 16-day period with consideration of image quality, temporal distance, as well as observation coverage to reduce the sensor and atmospheric noise. The HLS product is generated through the integration of Landsat 8 and Sentinel-2 with a processing workflow incorporating atmospheric correction, geometric resampling and registration, and BRDF normalization. The HLS dataset utilizes satellite images acquired from both Landsat and Sentinel-2 to create consistent surface reflectance products with a spatial resolution of 30 m. The HLS dataset provides a more frequent revisit cycle (approximately three days) since 2017, as Sentinel-2 started operating at full capacity after that year, yet cloud cover may substantially reduce the availability of HLS images. In our study sites, the number of cloud-free images ranges from 6 to 9 for each growing season, necessitating spatiotemporal image fusion for NRT crop mapping. Before fusion, the MODIS and HLS images are spatially aligned. The MODIS MOD10A1 Snow Cover Daily Global 500 m Grid product is further utilized to mask the snow pixels within the study sites. Extreme values (i.e., out of three times of standard deviation of reflectance) in the MODIS images are also masked and excluded from analysis. The Quality Assessment layer of the HLS data is used to mask pixels contaminated by cloud, cloud shadow, haze, and other noises. Six shared bands in the MODIS and HLS datasets are included in the fusion process, including blue, green, red, near-infrared (NIR), short-wave infrared 1 (SWIR1), and SWIR2. The resulting fusion dataset provides images with a 30-m spatial resolution and daily observations, which is ideal for NRT agricultural applications (see details in 3.1). Normalized difference vegetation index (NDVI), the most widely used vegetation index, is derived from the fusion dataset for the characterization of NRT crop phenology and utilized as a variable for crop type identification.

The Cropland Data Layer (CDL) dataset is utilized to provide the ground truth label (Boryan et al., 2011). The CDL dataset is produced and published each year by the National Agricultural Statistics Service (NASS) of United States Department of Agriculture (USDA). The dataset includes annual crop type maps with a 30-m spatial resolution, with a coverage of the Continuous US (CONUS) since 2018. In both the IL and MN study sites, the major crop types included in CDL are corn and soybeans, and the producer's and user's accuracies for these two crop types in the CDL dataset are around 95%. Therefore, CDL is well-suited for providing reliable ground truth reference information for crop type mapping.

The crop progress report (CPR) is a dataset providing weekly reports on cumulative percentages of major crop types reaching a variety of phenological stages at the state level (USDA-NASS, 2022). Published by USDA NASS, the CPR dataset offers state-level information on the phenological status of crops across the U.S., including the IL and MN sites in this study. Specifically, the emerged percentages of corn and soybeans in CPRs for the two states are utilized to validate the results of the NRT characterization of crop emergence dates in EMET.

In order to obtain thermal time information, DAYMET is employed in this study for the provision of meteorological data. DAYMET is a dataset providing daily weather information across North America, including temperature, precipitation, and solar radiation (Thornton et al., 2022; Thornton et al., 2021). The data are generated based on a combination of ground meteorological observations and statistical modeling. The spatial resolution for this dataset is 1 km, and the temporal resolution is 1 day. Among all the near-surface meteorological variables provided by DAYMET, daily minimum and maximum temperatures are adopted to estimate the thermal time accumulation for the crop species in our study sites.

3. Methods

The EMET framework consists of three main components, namely hybrid deep learning spatiotemporal image fusion, NRT thermal-based crop phenology normalization, and deep learning-based crop type mapping (Fig. 2). The hybrid deep learning fusion model is first employed to blend the MODIS and HLS images to generate daily 30-m satellite images. The fused time-series satellite data are then used for characterizing crop emergence in NRT. The crop emergence dates are then utilized as the onset dates of thermal time accumulation. By integrating the NRT crop emergence characterization with thermal time accumulation, our proposed thermal-based phenology normalization approach can register the crop growth progress throughout the growing season, enhancing the model robustness and transferability when the model is applied across different years and/or locations. After phenology normalization, the phenology-normalized time-series satellite observations are fed into an advanced deep learning classifier to generate the predictions of their corresponding crop types. Moreover, the satellite datasets involved in this study (e.g., MODIS and HLS) are updated with a low latency of one to two days after the data acquisitions, further facilitating the NRT application of the proposed EMET framework.

3.1. Hybrid deep learning fusion model

NRT crop type mapping requires timely acquisition of satellite observations with high spatiotemporal resolutions in order to capture the rapid phenological changes during the growing season. In this study, the hybrid deep learning fusion model (Yang et al., 2021) is utilized to fuse the MODIS and HLS data to generate time-series satellite images with

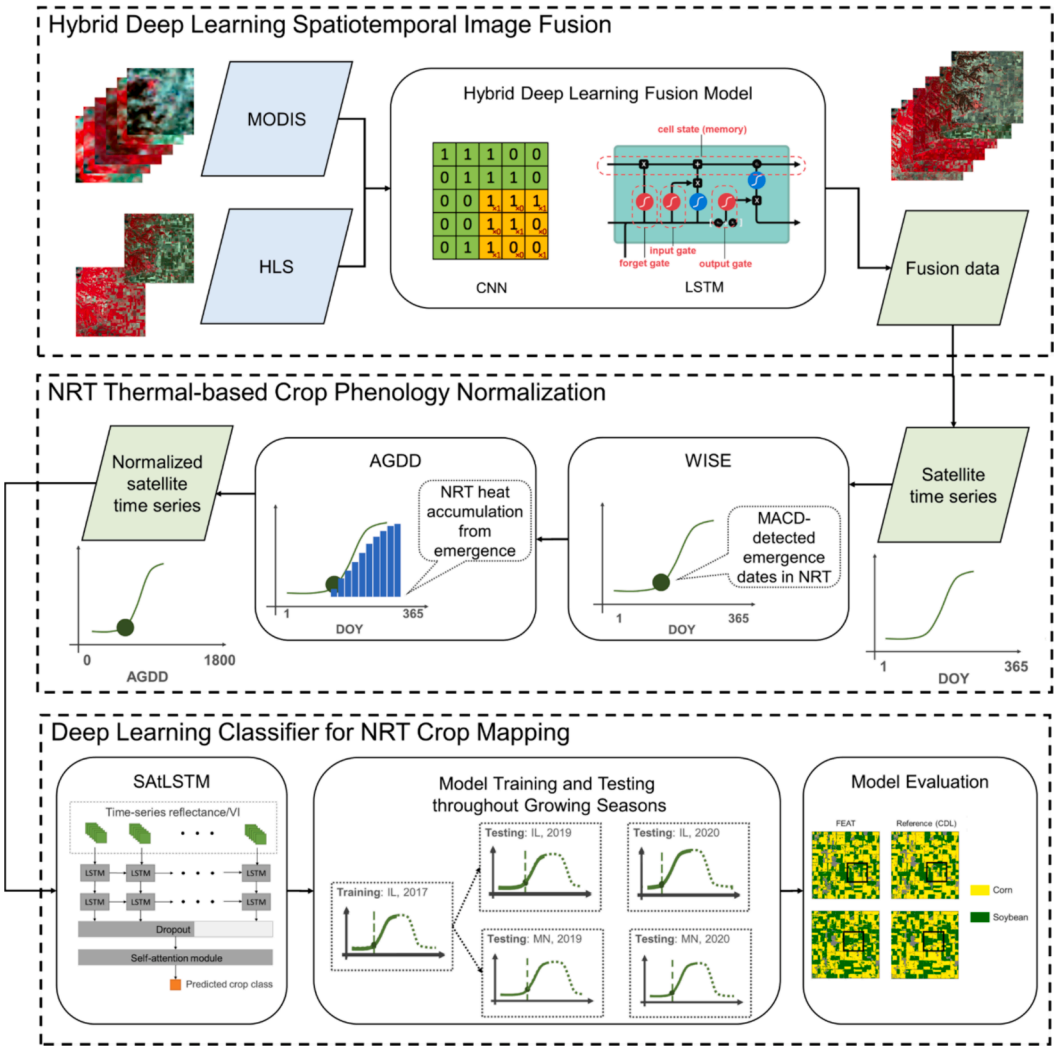

Fig. 2. The overall flowchart of the EMET framework.

the ability to predict temporal phenological changes among image acquisition dates. As Fig. 3 shows, the hybrid fusion model incorporates two types of deep learning architectures, namely super-resolution convolutional neural network (SRCNN) and long short-term memory (LSTM). SRCNN is capable of establishing the relationship between coarse features learned from the MODIS images and the fine features from corresponding HLS images, while LSTM is able to learn the temporal salient features in the sequential satellite images. Through integrating the two architectures, the hybrid fusion model can better capture spatial details and register the sensor difference of the two datasets due to the employment of SRCNN; the explicit modeling of temporal dependence through LSTM enables the model to capture rapid phenological changes among the time-series satellite observations. The hybrid deep learning fusion model is thus well-suited for NRT crop type mapping. On the one hand, the model ability to learn and predict drastic temporal changes is largely enhanced by integrating SRCNN and LSTM. On the other hand, the ability of LSTM to model complex temporal relationships allows the model to predict future dates without requiring satellite images both before and after the prediction dates.

The hybrid deep learning fusion model works as follows (Fig. 3). The SRCNN model consists of three convolutional layers. The first convolutional layer extracts critical spatial features from the MODIS images, followed by the non-linear mapping layer which maps the extracted features to the HLS scale. The third convolutional layer is to reconstruct

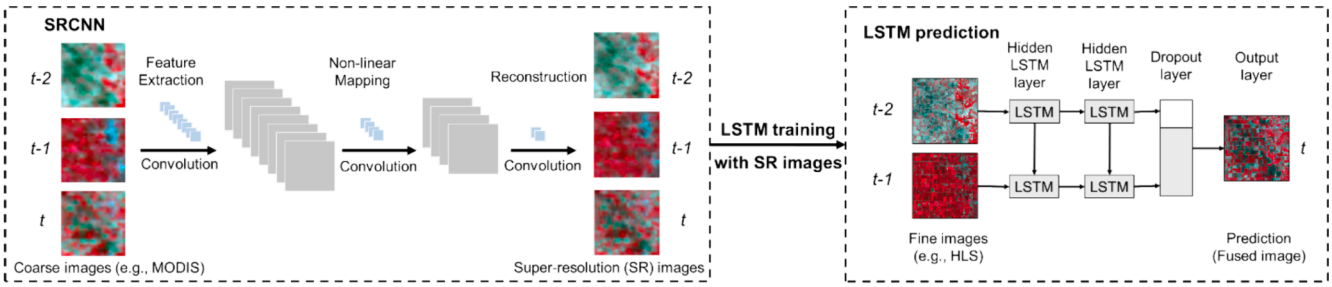

Fig. 3. The diagram of the hybrid deep learning fusion model.

the super-resolution (SR) images with the mapped features. The LSTM model consists of two LSTM layers, each with 100 neurons, followed by a dropout layer with a dropping rate of 25%. The LSTM model first learns the temporal features embedded in the sequential SR images. After learning the temporal changing patterns, the trained LSTM model is applied to the original HLS images to generate the final fusion image on the prediction date. Our previous study has shown that the hybrid deep learning fusion model can more robustly predict rapid phenological changes among images acquired during crop growing seasons, which ensures the high image quality that is appropriate for modeling dynamic agriculture systems, such as the U.S. Corn Belt. The detailed information about the model structure and parameters for the hybrid deep learning fusion model can be found in our previous study (Yang et al., 2021).

Before the fusion process, the MODIS images are first bilinearly resampled to the spatial resolution of HLS (30 m). To predict the fusion image on date *t*, the hybrid deep learning fusion model requires three MODIS images, namely M_{t-2}, M_{t-1}, and M_t, acquired on dates *t-2*, *t-1*, and *t*, respectively. The model also requires two HLS images H_{t-2} and H_{t-1}, collected on dates *t-2* and *t-1*, respectively. In practice, the MODIS and HLS image pairs that are closest to the prediction date (on dates *t-2* and *t-1*) are chosen to train the model. The SRCNN component learns the spatial relationships between features in the MODIS-HLS image pairs to generate HLS-wise SR images on dates *t-2*, *t-1*, and *t*. The sequence of SR images on dates *t-2*, *t-1*, and *t* are then utilized to train the LSTM model, which learns the temporal patterns across the image sequence. The LSTM model will then take the original HLS images on dates *t-2* and *t-1* and make the prediction of the fusion image on date *t*. We apply the fusion model on different predicted date *t* (i.e., from day 1 to 365), in order to generate a daily 30-m fusion dataset.

For each year and study site combination, 50,000 pixels (see Table S1 for the proportions of corn and soybeans) are randomly selected based on the corresponding CDL data. 50,000 time-series observations corresponding to the selected pixels are then generated with the daily 30-m fusion images. For each pixel, the time-series observations are from seven variables, namely six band reflectance (blue, green, red, NIR, SWIR1, and SWIR2) and NDVI. A moving Savitzky-Golay (SG) filter is then applied to the time-series observations to ensure that abnormal values are removed and that the time series are smoothed.

3.2. NRT thermal-based crop phenology normalization

3.2.1. NRT crop emergence characterization

In this study, an innovative within-season emergence (WISE) characterization model is incorporated in the EMET framework. WISE is designed as a trend-based NRT crop phenology characterization method, which is capable of identifying crop emergence timing using only the currently available satellite observations during the growing season. As Fig. 4 illustrates, the WISE algorithm utilizes the smoothed daily NDVI time series processed by the SG filter as mentioned in Section 3.1. The Moving Average Convergence Divergence (MACD) function, which is sensitive to early trend changes, is used to capture the signal of crop green-up. Specifically, the MACD is obtained through the difference between two Exponential Moving Average (EMA) functions with short and long moving windows, respectively. The EMA and MACD functions are as follows:

$$EMA(v(t), n) = v(t)*k + EMA(v(t-1), n)*(1-k) \quad (1)$$

$$k = 2/(n-1) \quad (2)$$

$$MACD(t) = EMA(v(t), a) - EMA(v(t), b) \quad (3)$$

where $v(t)$ represents the observation from the time-series NDVI data at time *t*; *n* is the size of moving window in EMA; *k* is regulated by *n* and represents the weight given to the $v(t)$. The MACD is obtained through two EMAs with window sizes *a* and *b*, which stand for a smaller size and a larger size of moving windows, respectively. Specifically, $a = 5$ and $b = 10$ in this study. The change of sign in MACD (i.e., from negative to positive) suggests a detected uptrend in the NDVI time series. Since MACD is a moving average-based indicator, the detected trend change in MACD can be delayed. Thus, the early trend changes are further characterized using MACD divergence (MACD_div), which is defined as follows:

$$MACD\_div(t) = MACD(t) - EMA(\mathrm{MACD}(t), c) \quad (4)$$

where the window size $c = 5$. The WISE model eventually identifies the emergence date *t* when the following conditions are satisfied:

$$\mathrm{MACD\_div}(t-1) < \mathrm{MACD\_div\_threshold}$$

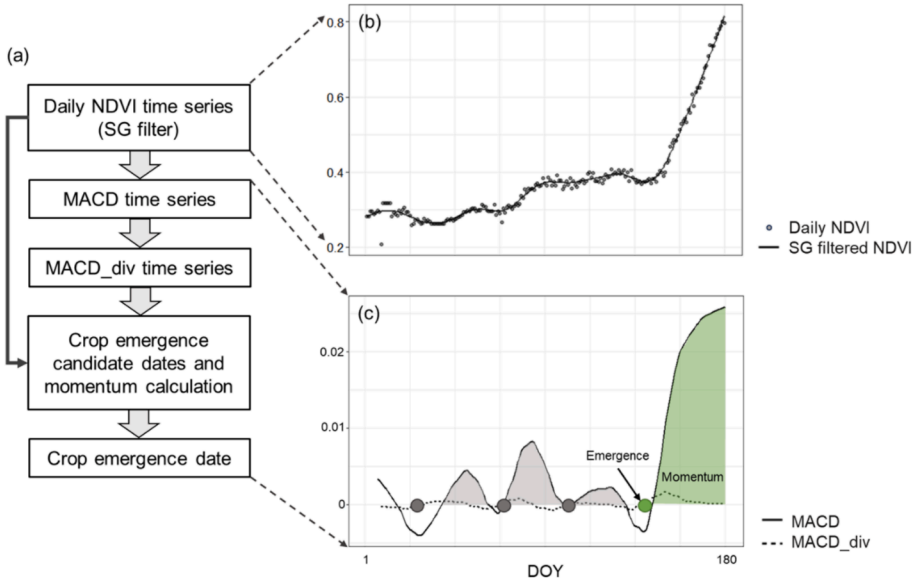

Fig. 4. (a) The flowchart of the WISE algorithm; (b) Daily NDVI time series from the fusion dataset smoothed with SG filter; (c) An illustrative diagram for the characterization of crop emergence based on the MACD and MACD_div curves generated from the smoothed NDVI time series. In this example, three dates meet the criteria of MACD and MACD_div, but only the third date (in green) meets the momentum requirement, while the other two dates (in gray) do not accumulate a momentum larger than 0.01.

and

$$\text{MACD\_div}(t) > \text{MACD\_div\_threshold}$$

and

$$MACD(t) < MACD\_threshold, \tag{5}$$

where the two thresholds MACD_threshold and MACD_div_threshold are empirically set to be 0.01 and 0, respectively.

Since WISE is a NRT crop phenology characterization model which may be applied during very early stages, it also computes the emergence momentum in order to ensure the significance of the detected uptrend in NDVI time series. The momentum is defined as an integral of positive MACD values after the detected emergence date divided by the number of days during the upward period. Through the incorporation of momentum, the WISE model can differentiate crop emergence which leads to significant uptrends in NDVI from other subtle uptrends caused by early-season weed growth or soil background information. The threshold for the momentum is set to be 0.01. All the parameters (i.e., *a*, *b*, *c*, MACD_threshold, and MACD_div_threshold) in WISE are tuned in consideration of a range of values with reference to previous studies (Gao et al., 2020a; Gao et al., 2021).

3.2.2. AGDD estimation

Growing degree day (GDD) is a widely used agricultural metric for measuring the heat accumulation of crop species throughout the growing season. The accumulated GDD (AGDD) approximates the energy available for crop growth and ultimately affects crop growth progress. This information of heat accumulation can also help determine the occurrence of specific crop phenological stages. In this study, GDD is calculated by subtracting the base temperature T_{base} from the adjusted daily average temperature $T_{adjusted\_avg}$. The difference between T_{base} and $T_{adjusted\_avg}$ represents the heat energy that can be utilized for crop growth and development. Specifically, GDD is defined as follows:

$$GDD = \max\left(0, T_{adjusted\_avg} - T_{base}\right) \tag{6}$$

$$T_{adjusted\_avg} = \frac{\max(T_{min}, T_{base}) + \min\left(T_{max}, T_{cap}\right)}{2} \tag{7}$$

where $T_{base} = 10\,^{\circ}\text{C}$, $T_{cap} = 30\,^{\circ}\text{C}$ (Akyuz et al., 2017; Angel et al., 2017). T_{min} and T_{max} stand for daily minimum and maximum temperatures, respectively.

AGDD is utilized to estimate the accumulated heat during the growing season. The AGDD estimation is obtained through the summation of daily GDD from the onset date (i.e., WISE-estimated crop emergence date) to the prediction date of the crop type information.

$$AGDD = \sum_{onsetdate}^{predictiondate} GDD \tag{8}$$

3.2.3. NRT thermal-based phenology normalization

In order to normalize the crop progress patterns and make crop time series comparable over space and time in NRT, the EMET framework employs a comprehensive thermal-based phenology normalization approach, integrating NRT crop emergence characterization and the thermal time metric AGDD. For a given pixel in the fusion dataset, the corresponding time-series NDVI is first extracted and smoothed. The WISE algorithm then detects the trend change points on the time series and estimates the emergence date with regard to the pixel. The WISE-estimated crop emergence date will serve as the onset date for heat accumulation. Starting from the WISE-estimated crop emergence date, the GDD value for each day is calculated and accumulated until the prediction date to obtain the AGDD for that pixel. All the calendar days will then be substituted with their corresponding AGDD values, and the time-series data, including the band reflectance and NDVI, will be transformed accordingly. Through this normalization process, the EMET framework can register the crop phenological progress across different years and regions, enhancing the model ability to accommodate the spatiotemporal variations in crop growth progress.

To evaluate the EMET phenological scenario, two benchmark scenarios are further designed and compared (Fig. 5). The first one is the calendar-based scenario, in which all the time-series data are corresponding to calendar dates, and no explicit phenology information is incorporated. All the time-series curves start at a certain date which is shared across regions and years. Since the growing season in the U.S. Corn Belt usually starts no earlier than late April, we set the shared start date as day of year (DOY) 120 for the first scenario. The second scenario is the WISE-based scenario, which incorporates the WISE-estimated emergence date to adjust the start dates of all the time-series data. Each curve will start at its corresponding emergence date derived by the WISE algorithm, so that the early-season phenology around crop emergence is registered over space and time.

3.3. Deep learning classifier

A self-attention based LSTM (SAtLSTM) model is devised in this study to model the time-series satellite data for crop type mapping. Time-series remote sensing observations are derived from the fusion dataset at the pixel level to serve as the input to the SAtLSTM model, with time steps in the model being 7 days or 50 GDDs. The predictions of crop types are determined by each pixel's predicted probabilities with regard to different crop types. The SAtLSTM model consists of two LSTM layers followed by a self-attention module (Fig. 6a).

The LSTM architecture is selected due to its ability to capture long-term dependencies in the time-series remote sensing images. During the growing season, the LSTM layers can identify different crop types using its learned features that are unique to their growth patterns and developmental stages over time. In each LSTM cell, the flow of information at a given time step is regulated by its input gate, forget gate, output gate, hidden state, and cell state. The gating mechanism can determine how much new information of satellite observation to be taken and how much of the previous states to be forgotten, which enables the model to learn long-term dependencies in time-series remote sensing data. For each time step, the LSTM cell state and hidden state are updated based on the cell state and hidden state from the previous time step, the satellite observation at the current time step, and the regulation mechanism of the three gates.

Self-attention is a deep learning technique that focuses on adaptively weighing the contributions of different inputs within a sequence to the final prediction results (Rußwurm and Körner, 2020; Vaswani et al., 2017). The self-attention mechanism allows deep learning models to assign higher weights to the periods that are most characteristic of each crop type. Moreover, the ability of the self-attention mechanism to model complex interactions between each time step and all the remaining time steps can help learn intricate temporal features that may differentiate crop types.

In the self-attention mechanism, the input time series (i.e., the hidden features from the LSTM layers) are transformed into three components, namely query, key, and value (Fig. 6b). The query and key are then utilized to calculate attention weights, which correspond to the importance of each hidden feature at different time steps to the final output. The attention weights are then used to weigh the values in the input sequence in order to obtain the output of the attention layer. Specifically, the self-attention mechanism works as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{9}$$

where Q, K, and V represent query, key, and value, respectively. d is the dimension of the key vector at a given time step. Q, K, and V are obtained through linear transformations of the input X:

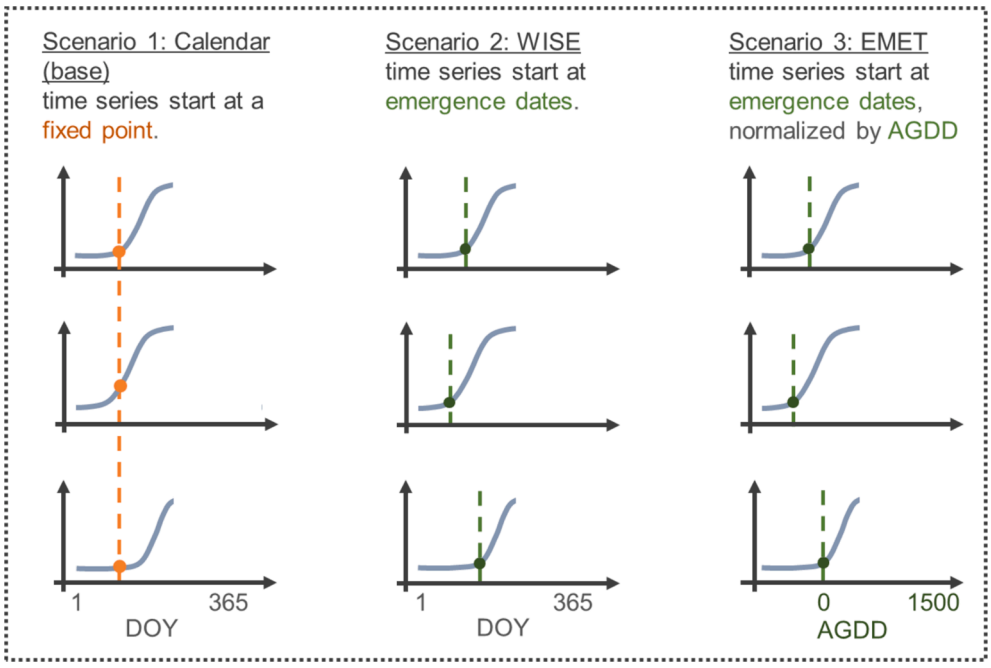

Fig. 5. Diagram of the three scenarios tested in this study. Scenario 1 is the calendar-based scenario. Scenario 2 utilizes the WISE-derived crop emergence dates to align the crop growth patterns. Scenario 3 illustrates how the proposed EMET framework integrates crop emergence information and thermal time together for phenology normalization throughout the growing season.

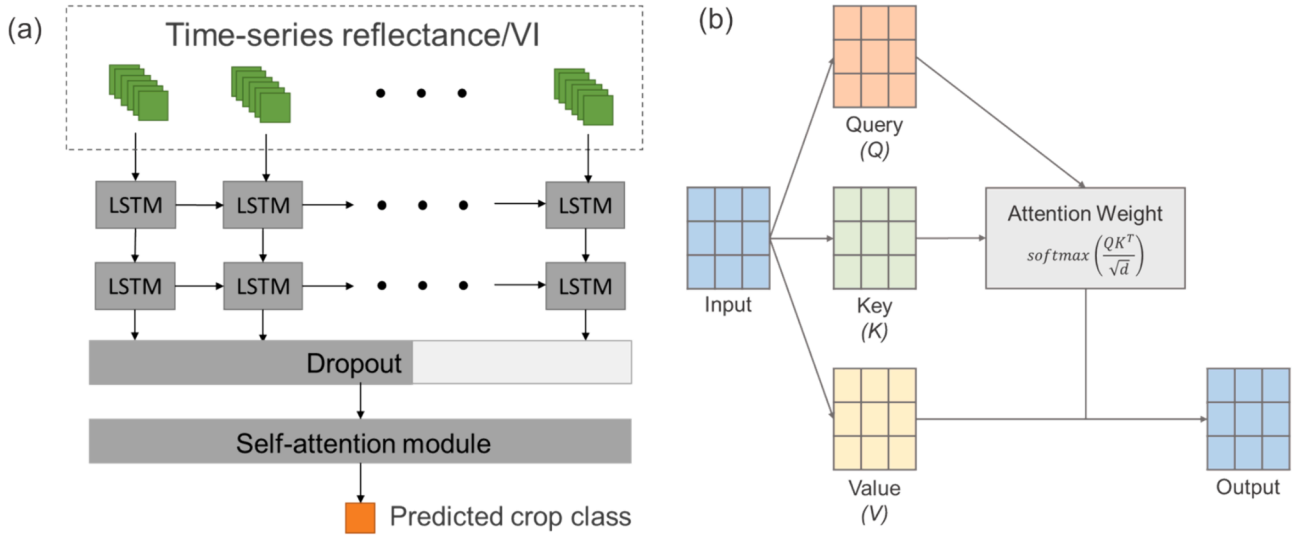

Fig. 6. (a) Model structure of the SAtLSTM model. Each green block represents one time step; the overlaid layers within each green block represent different features, including band reflectance and NDVI. (b) Schematic diagram of the self-attention mechanism.

$$\begin{aligned} Q &= W_Q X \\ K &= W_K X \\ V &= W_V X \end{aligned} \tag{10}$$

where X is the hidden state from the second LSTM layer, and W_Q, W_K, and W_V are the weights for linear transformations for query, key, and value, respectively.

Interpretation of the deep learning models can provide valuable insights in how the EMET framework can discern the differences between crop types. Here, we employ partial derivative for the SAtLSTM model to measure the importance of the input variables. In deep learning neural networks, partial derivative is utilized for updating the weights during backpropagation (Rußwurm and Körner, 2020). Variables that have larger effects on the model output can be identified through examining the partial derivative values of the model output with regard to the input variables. Specifically, the interpretation of partial derivative values is based on the magnitude and sign of the values. The magnitude of the partial derivative values represents the rate of change in the output caused by the change in the value of an input variable, corresponding to the variable importance. Partial derivatives can be positive or negative. Positive values stand for a positive relationship between the input variable and the model output, and vice versa.

The predicted probability of the SAtLSTM model for corn and soybeans is utilized to analyze how the prediction confidence evolves over time during the growing seasons. A dimensionality reduction technique, t-Distributed Stochastic Neighbor Embedding (t-SNE), is also employed to analyze the underlying structures in the high-dimensional hidden features learned by the SAtLSTM model (Van der Maaten and Hinton,

2008). t-SNE is able to project and visualize the high-dimensional data in low-dimensional space and preserve local structures. The t-SNE algorithm models probability distributions between data samples in both the high-dimensional and the low-dimensional spaces, and the resulting low-dimensional space is the one that minimizes the differences between those two probability distributions for all data samples. The t-SNE analysis of the hidden features from the LSTM layers can facilitate the evaluation of distribution and separability of the features learned from corn and soybeans pixels. The t-SNE results are further analyzed through the Silhouette score, a widely recognized metric for assessing the separability of different classes. The Silhouette score measures cluster separability through contrasting the intra-class distance and inter-class distance for the samples in each class. Ranging from −1 to 1, a higher Silhouette score indicates better separability between classes.

The SAtLSTM model is implemented with the Keras library and trained on an NVIDIA Tesla P100 GPU accelerator. Each LSTM layer includes 200 hidden units after a multitude of hyperparameter tuning. After the second LSTM layer, 50% of the neurons are randomly dropped out, in order to help the model avoid the overfitting issue. An Adam optimizer is adopted with its adaptive learning rates, and the initial learning rate for the optimizer is 0.001.

3.4. Experimental design

Time-series satellite observations are derived from the fusion data in 2017, 2019, and 2020 for the two study sites. Six band reflectance (blue, green, red, NIR, SWIR1, and SWIR2) and NDVI are included as the variables for phenology normalization and subsequent crop type mapping (Belgiu and Csillik, 2018; Cai et al., 2018; Xu et al., 2020). The deep learning classifier is first trained with the randomly sampled pixels of the IL study site in 2017, of which 80% are used for training and the remaining 20% are for validation. The model is then tested with the random samples in 2019 and 2020 for both the IL and MN sites. As mentioned in Section 3.1, 50,000 pixels are randomly sampled in each year and study site combination. By testing the EMET framework in a variety of years and regions, we aim to evaluate the robustness and scalability of the proposed EMET framework. To evaluate how the performance of the EMET framework evolves throughout the growing season, the SAtLSTM model is trained and tested every seven days, achieving a balance between regular updates and computational efforts. Model training and testing are from DOY 150 to 332 (approximately from late May to late November), with only the sequential satellite observations up to the corresponding DOY.

We evaluate the accuracy of the fusion images to ensure the reliability of the time-series fusion data for both study sites in the training and testing years. Specifically, three quantitative metrics are selected to evaluate the fusion results, namely root mean square error (RMSE), spectral angle mapper (SAM), and erreur relative global adimensionnelle de synthese (ERGAS) (Yang et al., 2023). RMSE is a single-band metric which calculates the average reflectance difference between the predicted fusion image and the reference HLS image. SAM is a cross-band metric evaluating the spectral distortion between the predicted and reference images. ERGAS is a cross-band metric which focuses on measuring the spectral difference between fusion images and reference HLS images by integrating normalized RMSE among all the bands. For all the three metrics, lower values suggest more favorable results. The results of NRT phenology characterization are also examined. The cumulative percentages of emerged corn and soybeans estimated by WISE are validated using corresponding CPRs, with RMSE as an accuracy metric.

The performance of EMET is compared with that of the calendar and WISE approaches. Overall accuracy and F1 scores are selected as the accuracy metrics to assess the model performance in classifying corn and soybeans. The overall accuracy measures the portion of the pixels that are correctly classified. The F1 score is the harmonic average of producer's accuracy and user's accuracy. It is calculated for corn and soybeans separately and provides insights into the model performance regarding different crop types. To understand how the SAtLSTM model leverages the satellite time series for crop type mapping, partial derivatives of the output predicted probabilities with respect to the model's input variables are also investigated (Xu et al., 2021). Partial derivatives can suggest the impact of each input feature at each time step on the final prediction of crop types. Euclidean distance is further utilized to measure the intraspecific and interspecific differences of corn and soybeans across years and study sites under different phenology normalization scenarios (Serra and Arcos, 2014). To investigate the benefits of the employment of the SAtLSTM model, random forest (RF), as a classic machine learning model, is adopted in this study as a benchmark model to be compared with SAtLSTM. As an ensemble model, RF is constructed with a diverse set of decision trees, which makes the model robust and effective in preventing overfitting. Since the RF model may not explicitly model sequential data, each band reflectance or NDVI at each time step is input as an individual variable to the RF model, and the prediction of crop types is determined by the voting results of the set of decision trees.

4. Results

4.1. Fusion performance

The mean RMSE, SAM, and ERGAS values of the hybrid deep learning fusion model for different year and study site combinations are calculated to assess the fusion accuracy (Table S2 and Fig. S2). The images used for validation for each year and study site combination can be found in Table S3. Fig. 7 shows the fusion results around DOY 200 in our study sites and years. The MODIS, HLS, and fusion images are in false color composite. The comparison between Fig. 7(b) and Fig. 7(c) suggests that the fusion images can well capture both the spatial and spectral features in the study sites. The spatial structures in the fusion images resemble those in the reference HLS images with clear field boundaries and similar within-field spatial patterns. For spectral information, the fusion and HLS images share similar color tones, and the fusion images can well capture the spectral differences among different crop fields. Two clusters of crop fields are observed in the HLS images, with their spatial distributions resembling those of the two crop species in the CDL dataset (Fig. 7(d)). The NDVI time series are extracted from the fusion images at randomly fields for corn (Fig. 7(e)) and soybeans (Fig. 7(f)) in the IL and MN study sites, overlayed with the corresponding NDVI values obtained from the cloud-free HLS images at the same locations. Compared to the HLS observations, the fusion images greatly improve the observation frequency, providing a much denser time-series dataset. The HLS data (orange dots) are well aligned with the fusion data (blue dots), which further confirms the performance of the hybrid deep learning fusion model. In terms of quantitative assessments, the mean RMSE values for visible bands remain lower than 0.03, and those for infrared bands range from 0.04 to 0.07. ERGAS also indicates the satisfactory performance of the fusion model. The mean ERGAS values spanning from 1.607 to 1.863, with standard deviation from 0.2 to 0.4; the mean SAM values are around 0.1. The standard deviations for these accuracy metrics remain low, suggesting a stable performance of the fusion model. All the quantitative metrics suggest the high accuracy of the fusion images produced by the hybrid deep learning fusion model.

4.2. WISE accuracy

The WISE algorithm is utilized to map crop emergence dates in NRT based on the NDVI time series derived from the fusion data. The emergence mapping results are presented for two example areas from the IL and MN study sites in 2019 and 2020, respectively (Fig. 8). The median emergence date of corn in 2019 is DOY 156 for the IL site and DOY 160 for the MN site, while their counterparts in 2020 are DOY 143 and 130, respectively. Within each map, the WISE algorithm can also retrieve the

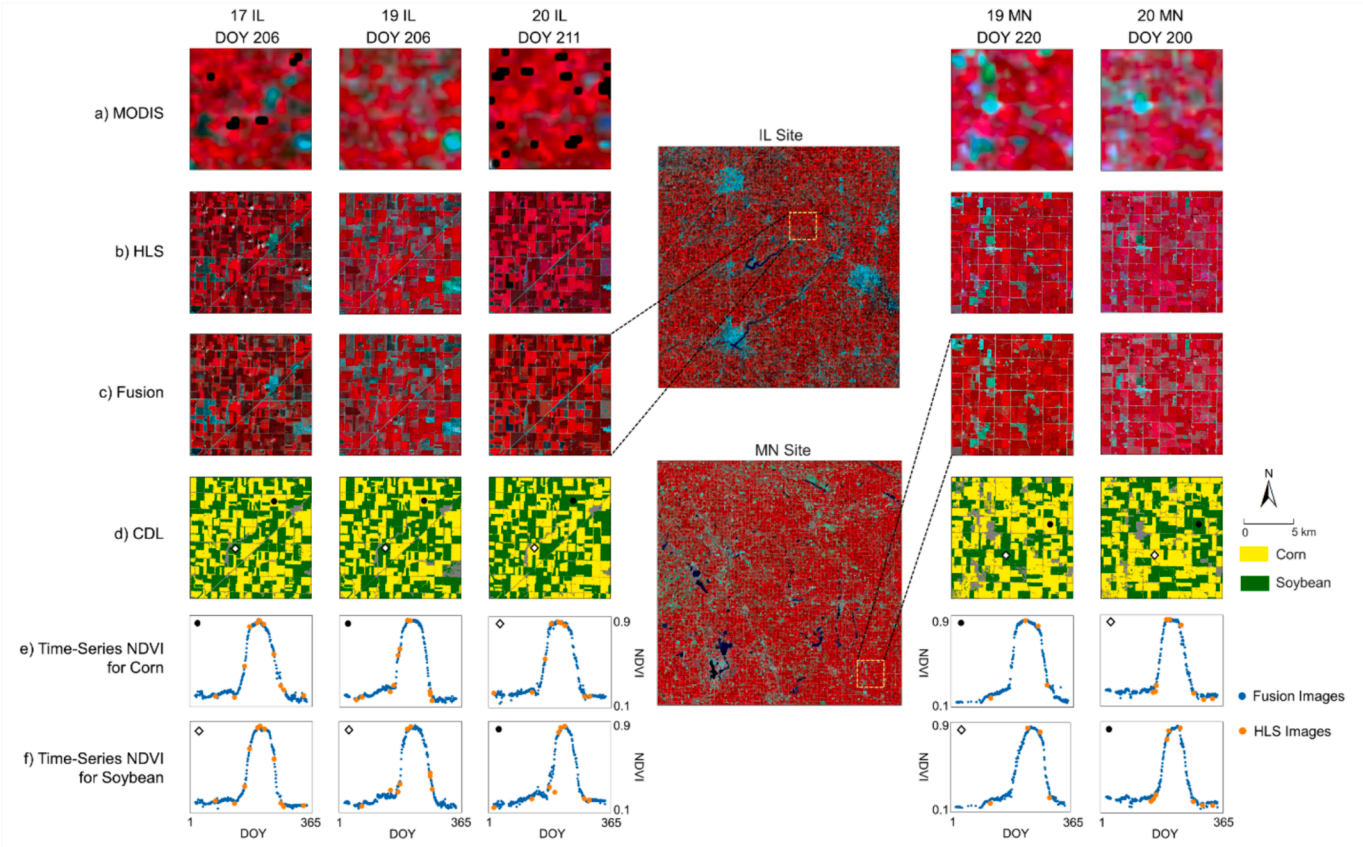

Fig. 7. Fusion results for the selected study sites and years. (a) MODIS images, (b) HLS images, and (c) fusion images are presented in false color composite. The fusion images show comparable spatial and spectral patterns to those of the reference HLS images, and the spatial structures resemble those in (d) the CDL data of the corresponding regions. Two randomly selected NDVI time series for (e) corn and (f) soybeans in the IL and MN study sites are plotted, respectively. Blue dots represent data derived from the fusion dataset, while orange dots are extracted from the original HLS images.

difference in crop emergence dates between corn and soybeans, as roughly two clusters of fields can be observed in all the maps, corresponding to the two main crop species in the study sites. Comparing the WISE mapping results with the CDL data, WISE is able to derive the general pattern that the emergence dates for corn are generally earlier than those for soybeans. Overall, the WISE algorithm gives promising results in NRT characterization of crop emergence, which is crucial in the normalization of the spatiotemporal variations in crop phenology.

To assess the accuracy of the WISE-derived emergence dates, results are aggregated for each year and study site combination and then compared with the state-level CPR dataset. Fig. 9 shows the cumulative percentage of crop emergence for corn and soybeans derived from both WISE (blue curve) and CPRs (red curve). As the extent of CPRs (state-level) is different from that of WISE results (HLS tile, a subset of the state), the CPR cumulative progress may exhibit a wider timespan compared to the WISE results. Yet the comparison between the WISE-derived emergence dates and CRPs indicates that the general emergence timings derived by the WISE algorithm match well with those of CPRs. The comparison also suggests that the WISE algorithm can capture the interannual variation of crop phenology, such as the delayed emergence of corn and soybeans in 2019. WISE can also well capture the interspecific phenological difference that corn tends to emerge earlier than soybeans. In most cases, the RMSE values for the various year and study site combinations are mostly around 10%. The difference between the WISE-derived median emergence dates and the CPR-derived median dates is within a week, which agrees with the previous findings (Gao et al., 2021).

4.3. Within-season crop mapping

4.3.1. Comparison across phenological scenarios

Fig. 10 shows the model performance of three phenological scenarios (i.e., calendar, WISE, and EMET) throughout the growing season for the IL and MN sites in both 2019 and 2020. In general, the EMET framework performs better than the calendar (base) and WISE scenarios. In the IL study site, EMET scores the highest end-of-season overall accuracy with 93.5% and 92.1% in 2019 and 2020, respectively (91.9% and 90.6% for WISE scenario; 90.3% and 89.8% for the base scenario). The EMET framework also achieves higher F1 scores for both corn and soybeans compared to the other two benchmark scenarios. Among the two benchmark scenarios, the WISE scenario tends to yield relatively more satisfactory results compared to the base scenario. The advantage of EMET is not only evident at the end of growing seasons but throughout the growing seasons. In most cases, EMET can reach the accuracy plateau around DOY 200, earlier than the benchmark scenarios. Taking 90% of overall accuracy as an example, the EMET framework reaches 90% accuracy around seven days earlier than the WISE scenario in both 2019 and 2020. In 2019, EMET achieves 90% accuracy more than 20 days earlier than the base scenario.

Compared to the IL study site, the advantage of the EMET framework is more substantial in the MN site, especially during the early- to mid-season. For both 2019 and 2020, EMET holds substantial advantage over the two benchmark scenarios until the end of September (around DOY 270). Similar patterns can also be observed for the F1 scores for corn and soybeans. In 2019, EMET maintains an advantage ranging from 6% to 10% in overall accuracy when compared to the base scenario from DOY 180 to 250, while the WISE scenario holds an advantage of around

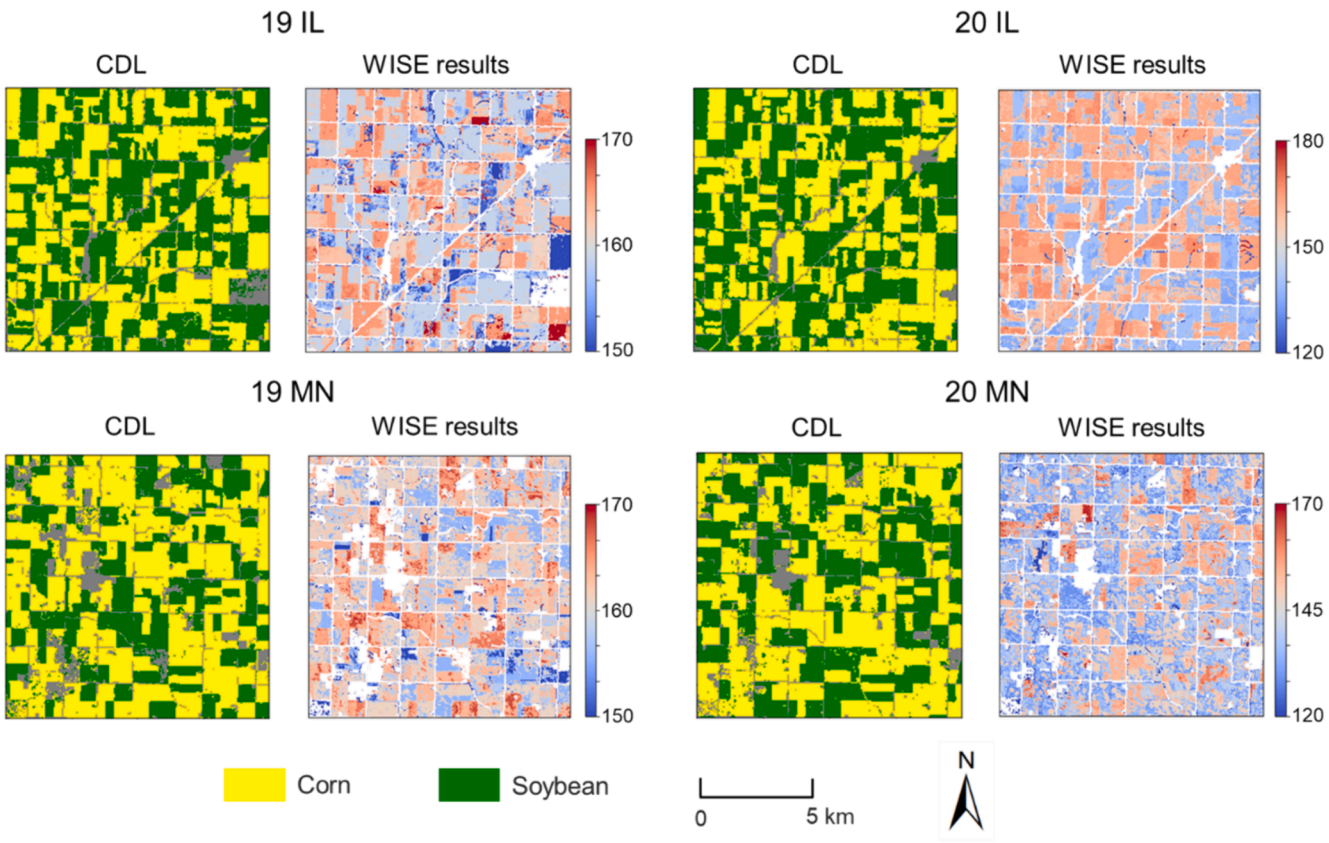

Fig. 8. Mapping results of the WISE-derived crop emergence dates, for example, subsets of the two study sites in 2019 and 2020.

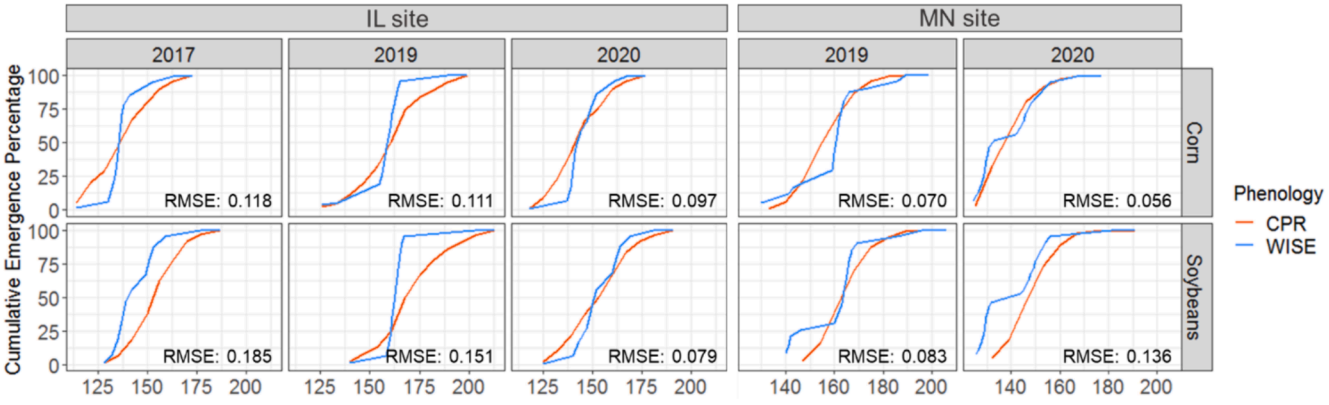

Fig. 9. Comparison of cumulative emergence percentages derived from WISE (blue) and CPRs (red) for corn and soybeans.

2% over the base scenario in the same period. In 2020, EMET also holds an evident advantage (between 2% and 5%) from DOY 200 to 260 compared to the two benchmark scenarios. In the MN site, EMET obtains an overall accuracy of 85% before DOY 200 in both 2019 and 2020, which is more than a month earlier than the two benchmarks. EMET is also the only one that reaches an overall accuracy of 90% in the MN site. Since the model is trained in IL and tested in MN, the end-of-season accuracy of the MN site is admittedly less favorable compared to that obtained in the IL site. Yet the substantial improvement brought by the EMET framework throughout the growing season suggests that the proposed model possesses the great ability to accommodate the spatio-temporal variations in crop phenological progresses and provide more reliable crop mapping results at an earlier timing.

As Fig. 10 indicates, it is more challenging to make an accurate map of crop types in the early season, considering the relatively limited availability of satellite observations. Thus, we further examine the mapping results at a relatively early timing in the growing seasons. Fig. 11 shows the mapping results of crop types on DOY 192 in two example subset areas of the IL and MN sites, respectively. For both the years 2019 and 2020, the EMET framework generates mapping results that are the most visually similar to the reference CDL data when compared to results from the benchmarks. The base scenario tends to generate maps with more salt-and-pepper effect. In 2019, the base scenario shows a tendency to overestimate the soybeans class in both study sites, while an overestimation of corn is observed in the MN site in 2020. Among the two benchmark scenarios, the WISE scenario can generate relatively more accurate maps compared to the base scenario. Overall, the better agreement between the EMET-generated crop type maps with the reference CDL data further demonstrates the enhanced model performance and scalability of the EMET framework.

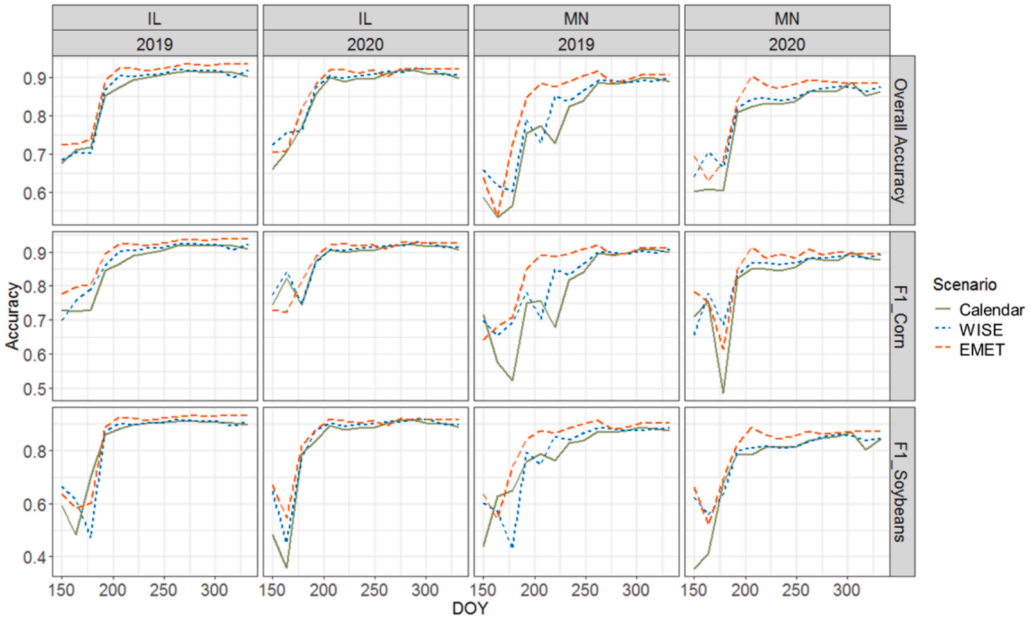

Fig. 10. Quantitative measurement of model accuracy throughout the growing seasons of 2019 and 2020 in the IL and MN study sites. Overall accuracy and F1 scores for corn and soybeans are employed as accuracy metrics. Green solid curves represent the calendar-based scenario, blue dotted curves stand for the WISE scenario, and the red dashed curves stand for the proposed EMET framework.

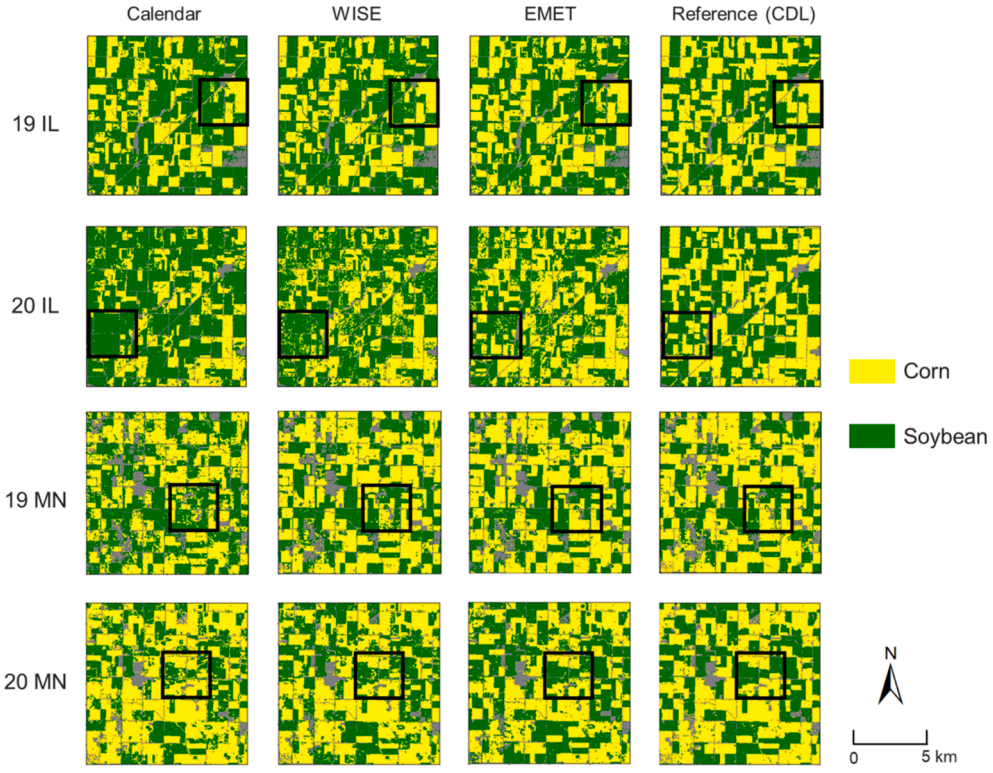

Fig. 11. Crop type mapping results of an example subset area in both IL and MN study sites on DOY 192 in 2019 and 2020. The EMET mapping results are most comparable to the reference CDL data.

With DOY 192 (July 11), 2019 in the MN site as an example, we further generate the t-SNE plots of the learned hidden features of SAtLSTM models under three phenological scenarios (Fig. 12). Overall, the features learned in the EMET scenario show better separability between the pixels of corn and soybeans. In the calendar-based scenario, the shapes of the clusters for corn and soybeans are more irregular, with substantial mixture between them. While the WISE scenario generally improves the separation between corn and soybeans, there exists a small cluster of data samples which are separate from the major clusters for both corn and soybeans and include mixed data samples from both classes. The SAtLSTM learned features in the EMET scenario show clearer separation between corn and soybeans, which suggests that the EMET framework is able to help the deep learning model extract the features that can better differentiate the two crop species. The Silhouette

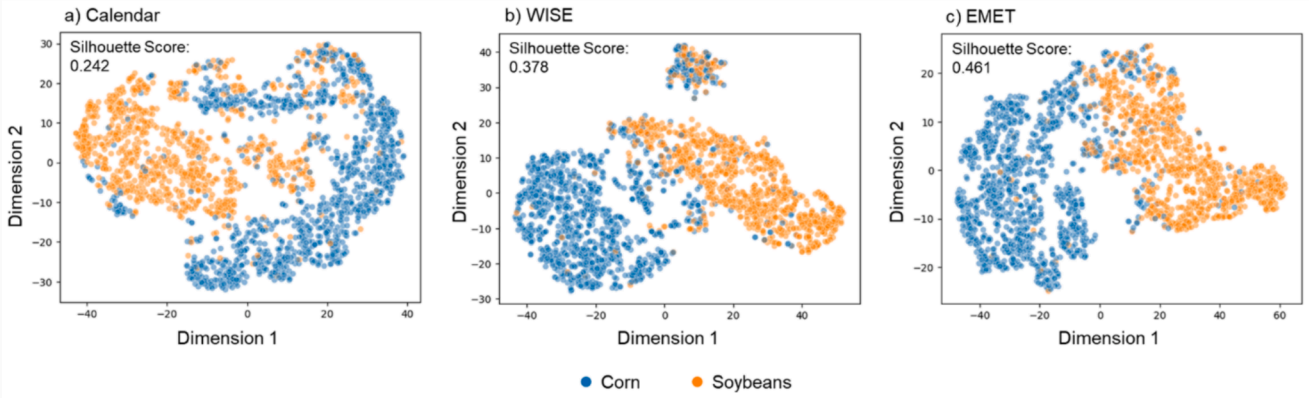

Fig. 12. t-SNE plots of the learned hidden features from the SAtLSTM model for (a) calendar-based scenario, (b) WISE scenario, and (c) EMET scenario. Plots are generated with the MN site in 2019 on DOY 192. The high-dimensional hidden features are reduced and projected to Dimensions 1 and 2.

score quantitatively confirms the findings. The Silhouette scores are 0.242 and 0.378 in the calendar and WISE scenarios, respectively, whereas the EMET scenario achieves a Silhouette score of 0.461.

Partial derivatives for all input variables throughout the growing season are derived from the SAtLSTM model during the training process under three scenarios (Fig. 13). In the calendar-based scenario, the sensitivity of the model output relative to the variables other than SWIR1 remains quite low. For the SWIR1 variable, the deep learning classifier does not respond much to the input until around DOY 170. The amplitude of partial derivatives then significantly increases from around DOY 170 to DOY 210, which coincides with the period when the accuracy of the model climbs fast (Fig. 10). After phenology normalization, SWIR1 still remains the most influential variable, with the model output being more responsive to other variables, such as Red and NIR. Comparing the magnitude of partial derivatives among scenarios, the calendar-based scenario yields much lower partial derivative values. The thermal-based EMET scenario generates overall the largest partial derivative values, indicating that the input variables in the EMET scenario can be better leveraged by the SAtLSTM model to accurately classify the crop types. The EMET scenario also generates partial derivatives with more peaks than the WISE scenario (e.g., NIR, SWIR1, and SWIR2).

We further examine the predicted probabilities of corn and soybeans pixels of the MN site under three scenarios from the early- to mid-season in the testing year 2019 (Fig. S3). The MN site in 2019 is selected as the accuracy results indicate that the spatial (IL vs. MN) and interannual (2017 vs. 2019) variations together in crop phenology pose greater challenges to crop mapping models. Results are shown every two weeks during the early- to mid-season. Pixels with a probability value larger than 0.5 are correctly classified by the model. In the early stage of the growing season, the calendar-based and WISE scenarios tend to make imbalanced predictions and largely overestimate one crop type over the other. For example, the WISE scenario tends to predict most pixels as corn in the early season. In contrast, the distribution of the predicted probability is more balanced between corn and soybeans under the EMET scenario, which leads to its better performance during the early season. After July, while all three scenarios can generally identify corn and soybeans with satisfactory accuracy, EMET still demonstrates better results compared to the benchmarks. For instance, on DOY 234 (mid-August), the median predicted probabilities of EMET for corn and soybeans are 0.936 and 0.964, respectively. Their counterparts for the calendar-based scenario are 0.822 and 0.894, respectively. For the WISE scenario, the median predicted probabilities for corn and soybeans are 0.825 and 0.918, respectively. Overall, the calendar-based and WISE scenarios yield more misclassified pixels compared to EMET. EMET also shows more left-skewed distributions and fewer pixels are with predicted probabilities lower than 0.5, which suggests its better performance over two benchmarks.

4.3.2. *Phenology normalization*

To examine how different phenology normalization approaches transform the crop progress patterns, we randomly select and average 1000 time series for each crop type in each year and study site combination. Fig. 14 presents the averaged time series of the SWIR1 band reflectance and NDVI time series for 120 days after the start dates of the time series in the three phenology normalization scenarios. As indicated in Fig. 13, this time period covers the time steps that are the most critical in differentiating the two crop species. SWIR1 is selected because this band is found to be the most important variable for accurate crop type mapping (Fig. 12); NDVI is selected as it is a widely used VI for assessing crop conditions and growth progress.

In the calendar-based scenario, there exists distinctive interspecific difference between the mean curves for corn and soybeans. Consistent with the CPRs, corn pixels are found with earlier emergence and growth progress compared to soybeans pixels in every year and study site combination. However, the intraspecific difference between the training and testing years is also substantial (i.e., the time-series patterns for the same species in the training and testing years are different). For example, the time series for corn in the testing year can sometimes be closer to that for soybeans in the training year, which likely leads to the confusion of these two crop species in the mapping results. Normalized by the crop emergence dates, the WISE scenario largely reduces intraspecific difference between the training and testing years. In general, the averaged time series for the two crop types in the testing years are close to their corresponding curves derived from the training year. Yet the interspecific difference also gets diluted after the WISE normalization, likely due to the fact that all the time series start at the emergence dates derived by the WISE algorithm. This observation is more obvious for the NDVI time series, which may stem from the fact that NDVI is the feature that WISE utilizes to identify crop emergence dates. Through the incorporation of the thermal unit AGDD, the EMET normalization scenario can reduce the intraspecific difference across years and study sites while still preserving the interspecific difference between corn and soybeans. While the WISE scenario can normalize the crop growth patterns with the detected crop emergence, it may not consider how different crop types respond to heat accumulation during the growing season. The employment of AGDD in place of calendar dates can reflect the differences in the growth rates across different phenological stages between corn and soybeans, which ultimately helps differentiate the growing patterns of one crop type from another.

Euclidean distance is employed to further quantify the intraspecific and interspecific differences (Tables 1 and 2). Intraspecific difference is measured by the Euclidean distance between the averaged time series of the six bands reflectance and NDVI for the same crop type in two

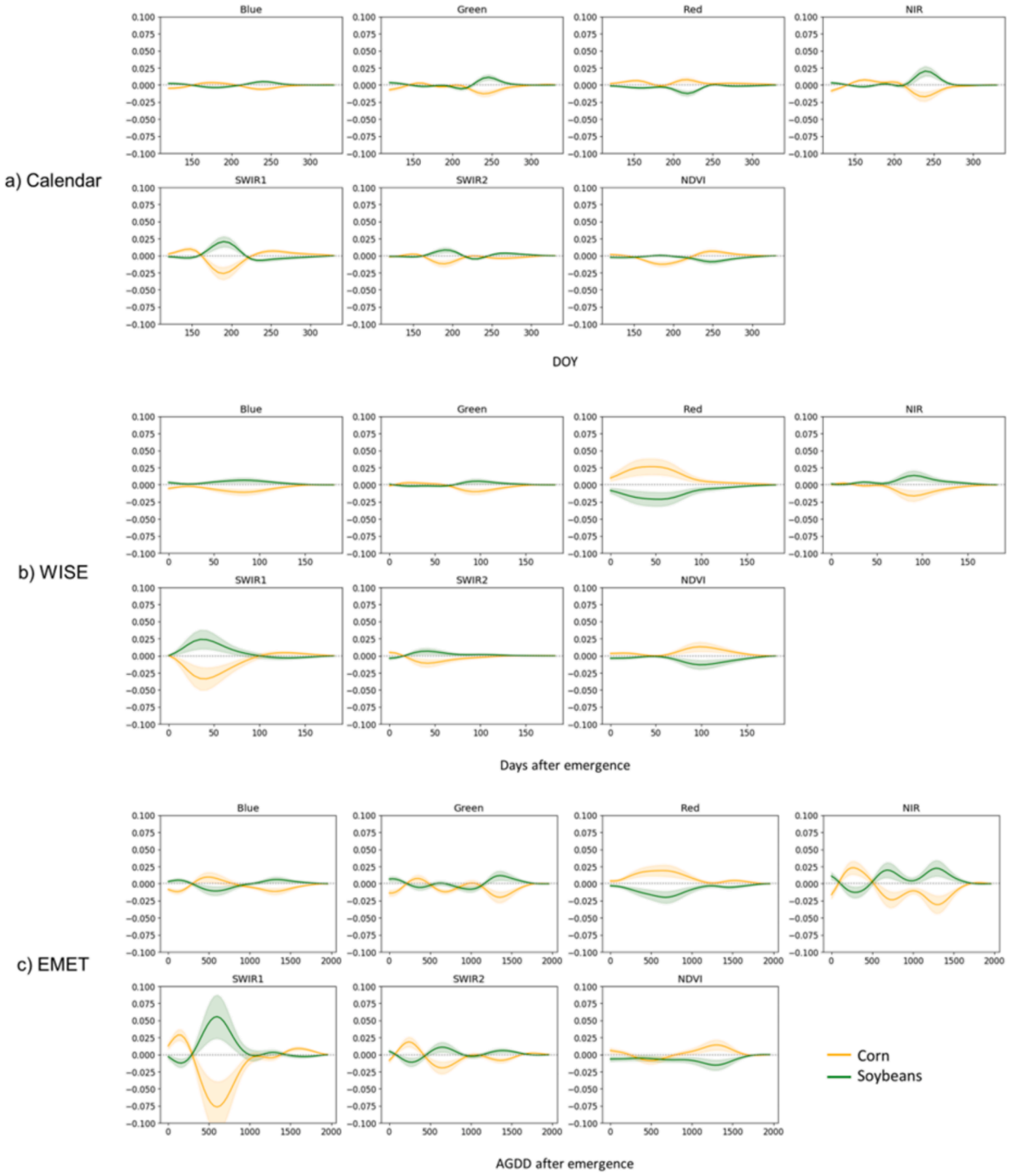

Fig. 13. Partial derivative values for the SAtLSTM model under the three phenology normalization scenarios: (a) calendar-based, (b) WISE, and (c) EMET. Partial derivative values are averaged for each crop type. Buffer areas represent 95% confidence intervals.

different years. Interspecific difference is obtained through the Euclidean distance between the averaged time series of the seven input features for the two crop types in the same year. Ideally, lower intraspecific differences together with higher interspecific differences are desired. Lower intraspecific differences suggest that the temporal patterns for the same crop type are more comparable across training and testing years, while higher interspecific differences indicate larger separability between corn and soybeans. For instance, with 2017 IL as the training case and 2019 IL as the testing one, the quantitative results suggest that the calendar-based scenario yields the larger distances for both intraspecific (14.739 for corn, 14.929 for soybeans) and interspecific differences (12.552 in 2017, 14.985 in 2019). While the WISE algorithm reduces intraspecific differences (7.037 for corn, 6.610 for soybeans), it also generates the lowest interspecific difference (5.335 in 2017, 11.313 in 2019). The proposed EMET normalization scenario not only yields substantially lower intraspecific difference compared to the calendar-based scenario (5.090 for corn, 7.044 for soybeans), but also retains a relatively large interspecific difference (8.126 in 2017, 12.449 in 2019). Similar patterns are observed in all the year and study site combinations – the EMET scenario can reduce the intraspecific

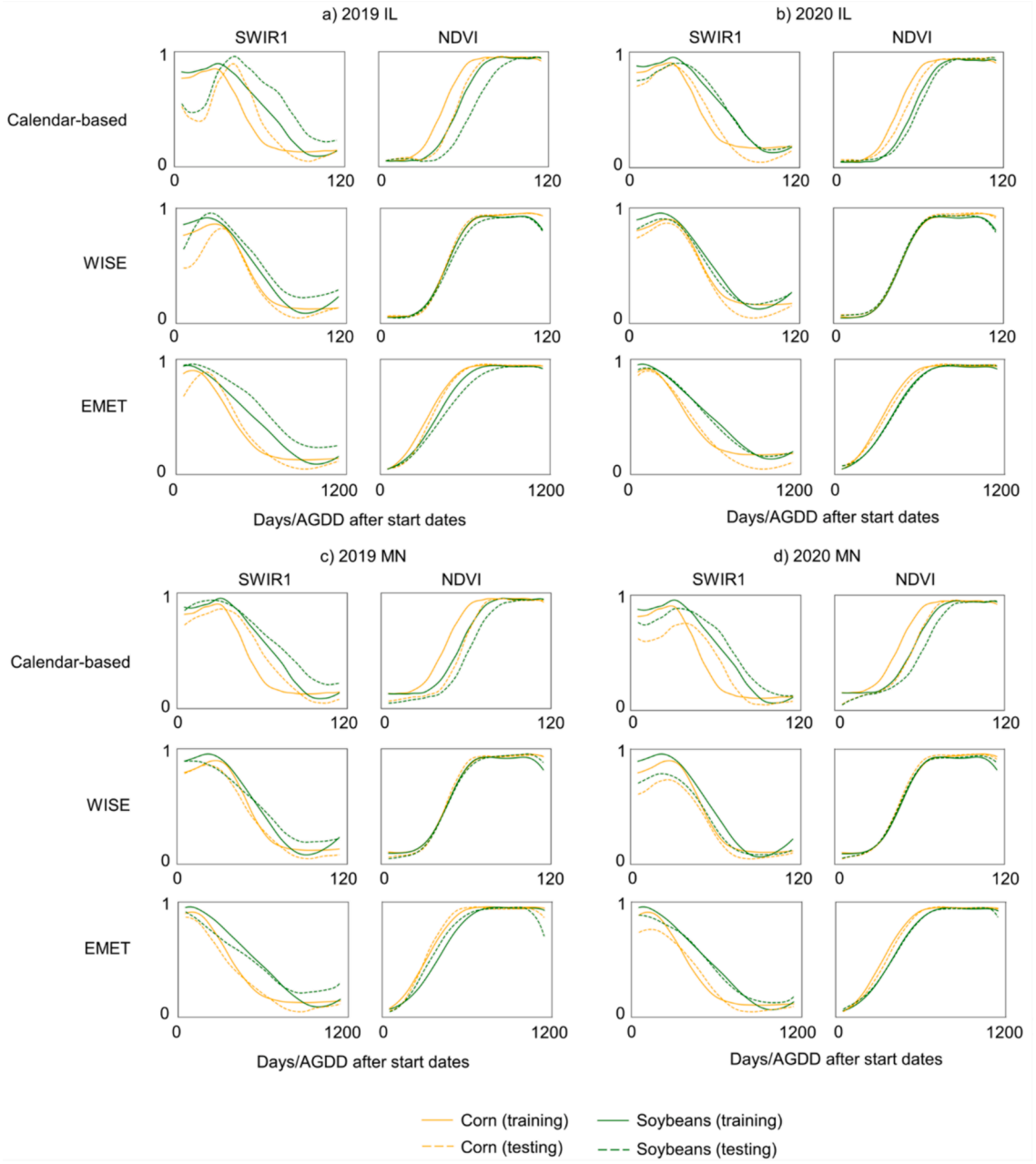

Fig. 14. Average time series of the normalized SWIR1 band reflectance and NDVI values generated from 1000 randomly selected pixels for both IL and MN sites in 2019 and 2020.

differences compared to the calendar-based scenario, while the inter-specific differences remain relatively large after the EMET phenology normalization compared to the WISE scenario. Overall, the results of Euclidean distances suggest the great potential of EMET in enhancing model scalability by normalizing the interannual varying phenological patterns while preserving the interspecific difference to help the model differentiate the crop types.

4.3.3. Comparison between SAtLSTM and RF

The overall accuracy and F1 scores of both classes are assessed throughout the growing seasons for the SAtLSTM and RF models, respectively (Fig. 15). The results suggest that the SAtLSTM model performs consistently better than the RF model for both the IL and MN study sites throughout the two testing growing seasons in 2019 and 2020. Among the four site-year combinations, the performances of the two models are relatively more similar for the IL site in 2020, while the advantage of SAtLSTM is more significant in the other three site-year combinations. Considering that the model is trained in the IL site, and that 2020 is a year that has more similar meteorological conditions to the training year 2017, it would presumably be less challenging to predict crop types in 2020 for the IL site. For the IL site in 2019, the SAtLSTM model possesses a consistent advantage in overall accuracy of around 5% throughout the growing season. The better performance of SAtLSTM is also evident in terms of F1 scores for both corn and soybeans. For the MN site in 2019 and 2020, the SAtLSTM model generates significantly higher overall accuracy and F1 scores for corn in the early season. After the accuracy plateau, the deep learning model can still hold an advantage of approximately 3%, accompanied by consistently higher F1 scores for soybeans. When the weather conditions and/or the locations differ from those of the IL site in 2017, the SAtLSTM model exhibits more superior capabilities of capturing the temporal-evolving features embedded in the time-series satellite observations, compared

Table 1
Intraspecific differences between the averaged time series in the training (2017 IL) and testing (2019 and 2020 for IL and MN) years, for corn and soybeans. Differences are measured by Euclidean distance.

Euclidean Distance		Intraspecific – Corn	Intraspecific – Soybeans
2017 IL & 2019 IL	Calendar	14.739	14.929
	WISE	7.037	6.61
	EMET	5.09	7.044
2017 IL & 2020 IL	Calendar	8.098	7.156
	WISE	5.81	5.333
	EMET	5.224	4.526
2017 IL & 2019 MN	Calendar	11.92	9.441
	WISE	6.643	6.769
	EMET	7.91	7.749
2017 IL & 2020 MN	Calendar	15.079	12.372
	WISE	10.809	8.676
	EMET	8.675	8.119

Table 2
Interspecific differences between the averaged time series of corn and soybeans in the training (2017 IL) and testing (2019 and 2020 for IL and MN) years. Differences are measured by Euclidean distance.

Euclidean Distance		Interspecific – Training	Interspecific – Testing
2017 IL & 2019 IL	Calendar	12.552	14.985
	WISE	5.335	11.313
	EMET	8.126	12.449
2017 IL & 2020 IL	Calendar	12.552	11.502
	WISE	5.335	7.218
	EMET	8.126	8.826
2017 IL & 2019 MN	Calendar	12.552	10.515
	WISE	5.335	7.283
	EMET	8.126	9.154
2017 IL & 2020 MN	Calendar	12.552	11.06
	WISE	5.335	7.113
	EMET	8.126	7.941

to the RF model.

4.3.4. *County-level crop type mapping*

To demonstrate the ability of the EMET framework to map crop types in large spatial areas, we present county-level crop mapping results in both the IL and MN site in 2019. The Piatt County, IL and the Watonwan County, MN are selected as two mapping areas. Fig. 16 shows the comparison of the county-level mapping results among the proposed EMET framework and the two benchmarks in the early season, using DOY 171 in mid-June for illustration. In both IL and MN sites, the calendar-based and WISE scenarios show limited ability to differentiate the crop types in the early season. In Piatt County, IL, the calendar-based and WISE scenarios yield overall accuracies of 0.614 and 0.620, respectively. They tend to overestimate corn distribution, and few soybeans fields are correctly identified. The EMET scenario achieves an overall accuracy of 0.774, with relatively more soybeans being correctly identified. In Watonwan County, MN, the calendar-based scenario largely overestimates soybeans pixels (overall accuracy: 0.631), whereas the WISE scenario substantially misclassifies soybeans as corn (overall accuracy: 0.676). On the contrary, the EMET framework demonstrates its improved capability of differentiating corn and soybeans (overall accuracy: 0.736). Though the EMET results are not perfect with blurred boundaries and salt-and-pepper effects in the early season, the proposed framework starts to capture the general spatial distribution patterns of the two crop species, suggesting the efficacy of our proposed phenology normalization strategy.

Fig. 17 further illustrates how the EMET crop mapping results evolve in the early- to mid-season in 2019 for the two selected counties in the IL and MN sites. Three maps generated for early July, early August, and early September are presented for the Piatt and Watonwan Counties, respectively. For the Piatt County, EMET can already produce a fairly accurate crop type map in early July in 2019 with an overall accuracy of 0.905, though there still exists an overestimation of soybeans, which leads to the slight salt-and-pepper effects in the map. As the growing season progresses, the Piatt crop type maps generated later in early August (overall accuracy: 0.937) and early September (overall accuracy: 0.940) are progressively more accurate when compared to the reference CDL data. Mapping crop types for Watonwan County in MN is more challenging due to the spatial difference between the IL and MN sites. As a result, maps of Watonwan County have relatively more salt-and-pepper effects when compared to those of Piatt County in IL. Yet the EMET framework can still well capture the general distribution of corn and soybeans fields in early July (overall accuracy: 0.857). The mapping accuracy in Watonwan County gradually increases as the season progresses, achieving an overall accuracy of 0.885 in early August and 0.934 in early September. Overall, the county-level mapping results demonstrate that the EMET framework is promising in accurate mapping of crop type over extended geographical regions in NRT.

5. Discussion

The novel EMET framework is proposed in this study to facilitate

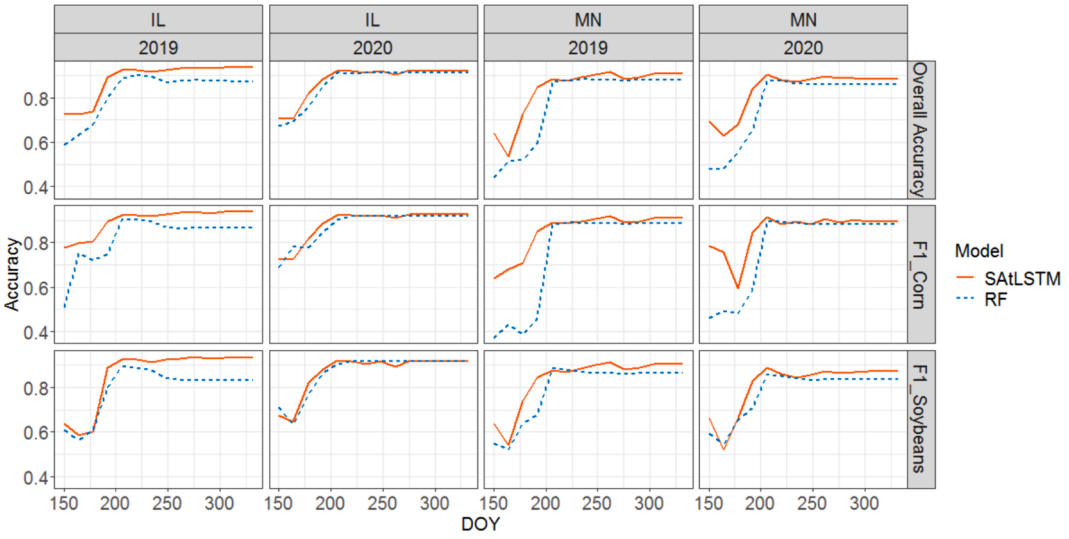

Fig. 15. Comparison of overall accuracy and F1 scores for corn and soybeans generated by the SAtLSTM and RF models.

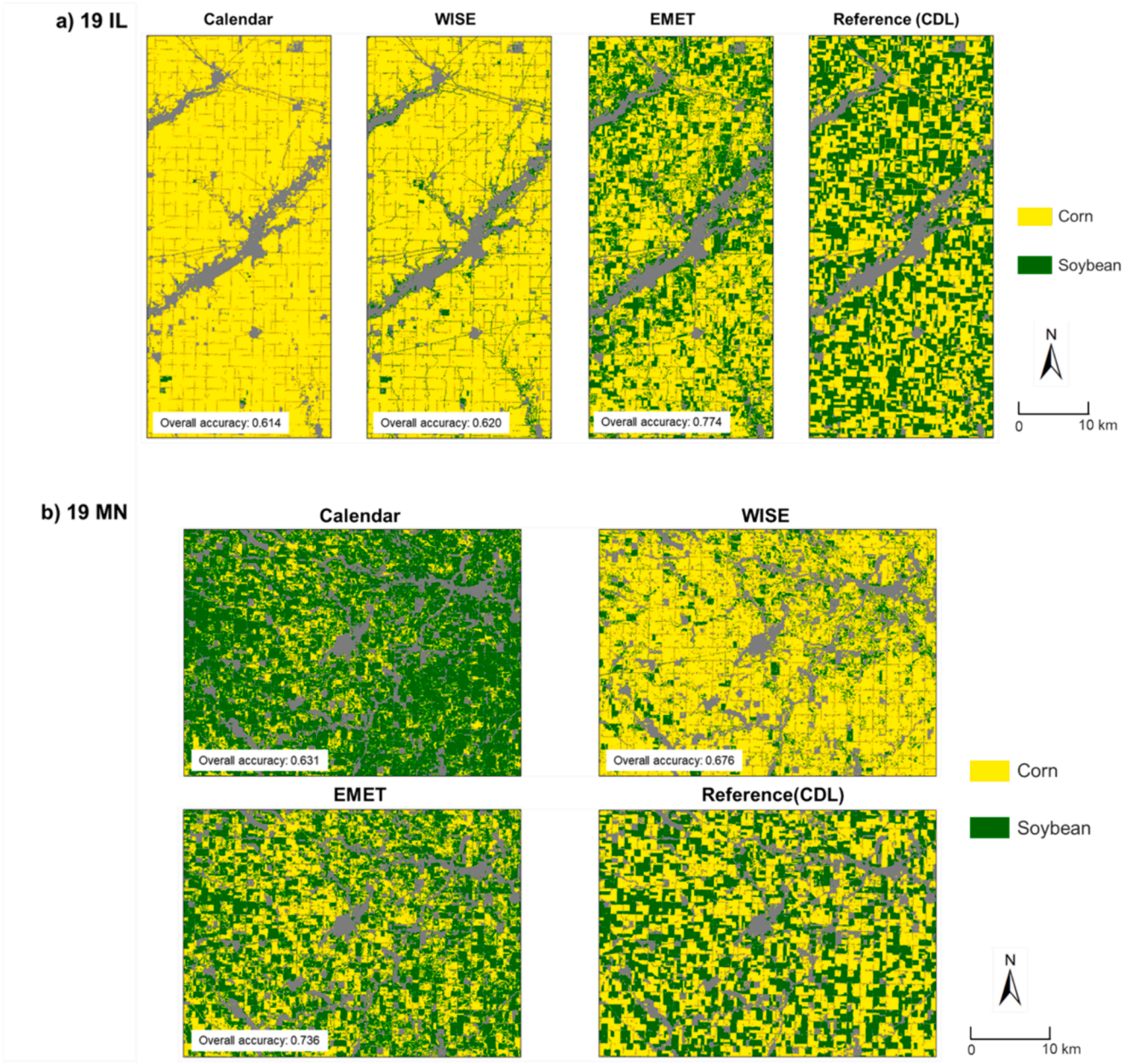

Fig. 16. Early-season crop type maps for (a) the Piatt County in the IL site and (b) the Watonwan County in the MN site. Maps are generated on DOY 171 in 2019.

timely and scalable crop mapping across space and time using training samples from previous years. Through the integration of spatiotemporal image fusion, thermal-based phenology normalization, and deep learning, the EMET framework demonstrates superior performance in NRT mapping of crop types at the field level in the U.S. Corn Belt. The hybrid deep learning fusion model plays an important role in monitoring the farm fields in NRT by providing time-series remote sensing images with high spatial and temporal resolutions throughout the growing season. The temporally dense satellite dataset lays a solid foundation for the subsequent NRT crop phenology normalization and crop type mapping. A thermal-based NRT phenology normalization approach is innovatively devised through integrating NRT crop emergence characterization (WISE) and AGDD normalization, which allows the timely normalization of the crop phenological patterns across space and time. An advanced deep learning model, SAtLSTM, is employed in this study in an effort to effectively extract and learn the complex temporal evolving features from the normalized satellite time series for accurate crop classification.

The advantage of the EMET framework particularly lies in its capability of NRT crop mapping throughout the growing season. While satellite missions provide valuable time-series datasets for agricultural applications, it remains challenging how we can leverage the limited information available during early- to mid-season and make reliable crop type map as early as possible. One prominent challenge is that data availability is highly uncertain during such a relatively short time period. For rainfed agricultural systems like the U.S. Corn Belt, the availability of frequent satellite observations can hardly be guaranteed due to cloud contamination. To that end, the EMET framework incorporates the hybrid deep learning fusion model, which captures the rapid phenological temporal changes and ensures the availability of dense time series at any time during the growing season. The temporal dynamics captured by the time series facilitates phenology characterization (e.g., emergence) and accurate crop type mapping. Quantitative accuracy assessment further confirms that the enhanced spatiotemporal resolution of the fusion dataset contributes to earlier and more accurate identification of crop types (Fig. S4). The satellite datasets used in this

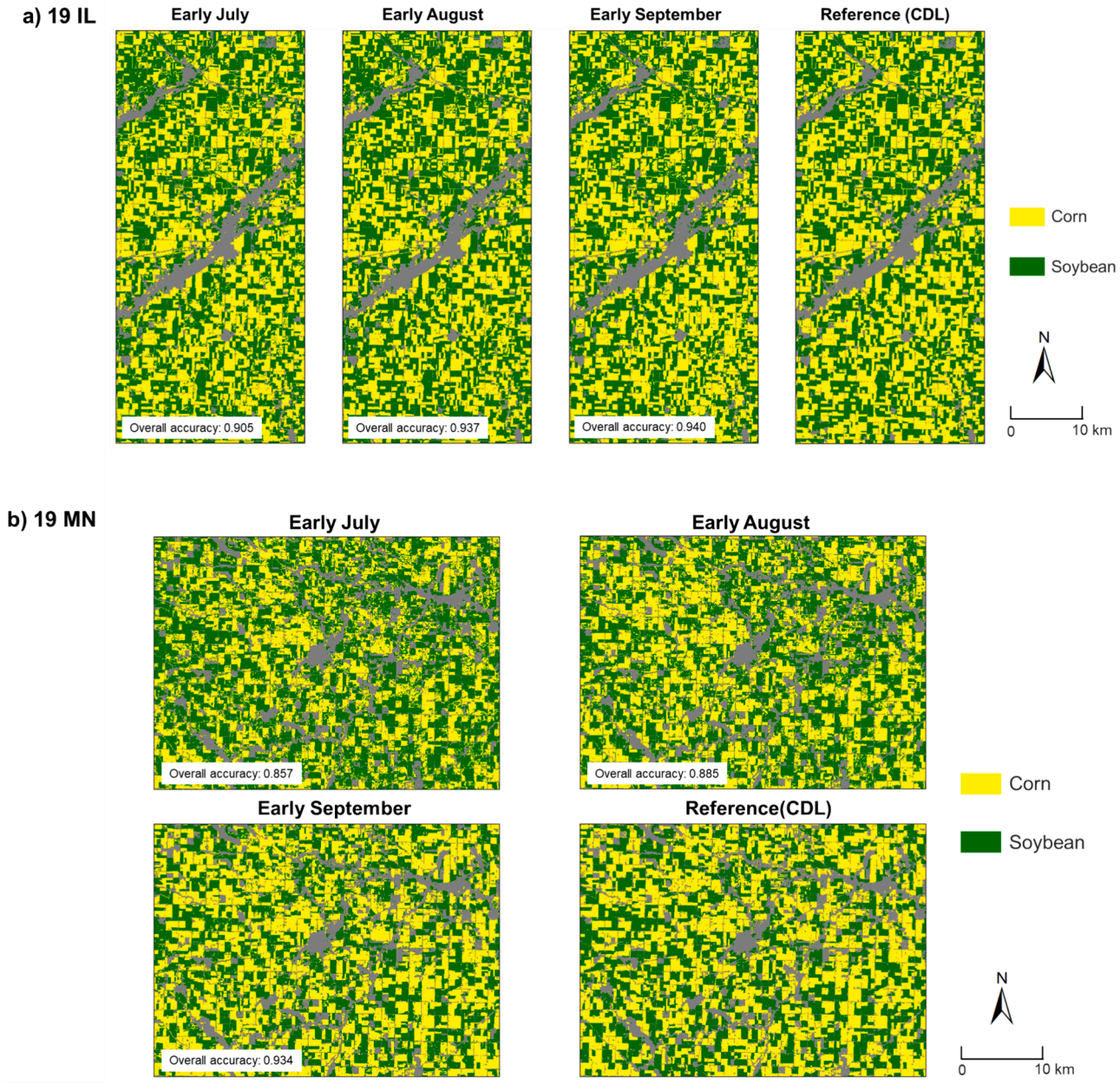

Fig. 17. Evolving crop type maps for (a) the Piatt County in the IL site and (b) the Watonwan County in the MN site during the early- to mid-season in 2019, generated by EMET.

study (i.e., HLS and MODIS) are publicly available only one to two days after the images are acquired by the satellite sensors, securing the low latency of the proposed EMET framework and facilitating NRT crop mapping.

Another challenge lies in NRT crop mapping is that early crop phenology is subject to the influence of a variety of environmental and anthropogenic factors, which largely hinders the performance of conventional calendar-based approach in the effective identification of crop types. While previous studies have demonstrated the feasibility of within-season crop mapping by coupling satellite images and ground truth labels from past seasons, the calendar-based approach is subject to the spatiotemporal variation in crop phenological progress (Xu et al., 2020; Zhong et al., 2019). The integration of fusion datasets enables EMET to uniquely leverage rich temporal phenological information to derive insights from dense time series at the field level. By innovatively integrating WISE and AGDD, crop phenological progress across space and time can be normalized in NRT with phenological patterns of the same crop type more comparable. Meanwhile, the phenological variations of different crop types can be preserved. Leveraging MACD function and the momentum criteria, WISE can utilize partial-season data to detect early signals of crop emergence. Different from the CPR dataset which records crop emergence dates at the state level, the WISE algorithm is able to estimate pixel-level crop emergence for each farm field within a week after crop emergence (Gao et al., 2020a; Gao et al., 2021). The WISE-derived emergence dates serve as the biofix dates for the AGDD estimation, which enables further normalization of the crop phenological patterns throughout the growing season. The EMET-normalized satellite time series contain comprehensive yet complex temporal features corresponding to different crop types. With the gate mechanism in LSTM coupled with the self-attention design, the SAtLSTM can model the temporal dependencies and capture the sequential features that are sensitive to the different crop species, better leveraging the limited information available during early- to mid-season.

The efficacy of the EMET phenology normalization approach is assessed through the design of three phenological scenarios. The calendar-based scenario is devised based on the commonly adopted approach which organizes satellite observations based on the acquisition dates (Cai et al., 2018; Xu et al., 2020). Through the NRT characterization of crop emergence stage, the WISE algorithm facilitates normalization of early-season phenology (i.e., emergence). The WISE scenario thus holds great potential in accommodating the spatiotemporal variability of crop phenological progress caused by the differences in anthropogenic factors and environmental conditions around the planting and emergence stages. Yet the WISE scenario may be limited in normalizing the phenological progress in the mid- to late-season. To that end, EMET normalization is designed to incorporate the heat accumulation, an important driver of crop growth progress, to normalize crop phenology in NRT. Current literature has discovered that thermal time has been found conducive to phenology characterization across locations and years (Liao et al., 2023; Nguyen et al., 2020; Qian et al., 2019; Zeng et al., 2016). In this study, EMET innovatively incorporates WISE and AGDD in an NRT manner for phenology normalization, which is found to be advantageous compared to the benchmark scenarios. Our results suggest that the EMET scenario yields consistently better crop type mapping accuracy across testing sites and years, with superior performance especially in the early season. The ability of EMET to yield accurate classification results earlier than the benchmarks makes it promising in providing timely estimates of crop type distributions for relevant stakeholders. The partial derivatives can further help us understand how the model learns critical information from the satellite observations in the three scenarios. In terms of feature importance, the SWIR1 band reflectance is identified as the most important variable for all the three scenarios, consistent with the findings in the existing literature that SWIR1 plays an important role in classifying corn and soybeans in the U.S. Corn Belt (Cai et al., 2018; Xu et al., 2021). The most critical period for identifying corn and soybeans is between approximately DOY 170 and 210 (mid-June to late-July). According to CPRs, this period corresponds to around the silking stage of corn and the blooming stage of soybeans. Since the SWIR1 band is sensitive to water content in crop plants, its high importance indicates that the difference in water content between corn and soybeans during their late vegetative to early reproductive stages may provide valuable information for accurate crop type mapping.

The advantage of EMET lies not only in higher mapping accuracy, but also in enhanced model scalability. Recent literature has found that the incorporation of crop phenological information facilitates better spatiotemporal transferability of crop classification models (Kerner et al., 2022; Yang et al., 2023). In the context of NRT crop type mapping, EMET also demonstrates its enhanced scalability compared to the calendar-based and WISE scenarios, particularly for the MN site with 85% overall accuracy achieved over a month earlier. The advantage is more obvious in 2019 during the early season when the excessive rainfall largely altered crop phenological progress. The degraded performance of the benchmark models in the MN site suggests that the spatial difference between the study sites may largely affect crop type mapping accuracy, and that the EMET framework can better accommodate not only interannually changing but also spatially varying phenological patterns. To confirm this finding, additional spatial transferability tests are carried out in Michigan (MI) and Kentucky (KY) in 2019 (Fig. S5). These two sites are located distantly from the IL and MN sites, with substantial differences in environmental and climatic conditions (Fig. S5a-c). The consistent advantage of EMET from early- to late-season in the MI and KY sites further confirms EMET's enhanced spatial scalability (Fig. S5d). Compared to the WISE normalization approach that only considers the crop emergence dates, the use of AGDD in EMET helps reduce the variability of the length for different phenological stages after crop emergence, facilitating better model scalability (Zeng et al., 2016). The analysis of intraspecific and interspecific differences further illustrates the advantages of EMET over the WISE scenario. While maintaining low intraspecific differences, EMET yields substantially larger interspecific differences across different years and study sites. The higher interspecific differences stem from EMET's ability to better account for the crop-specific relationship between crop growth progress and heat accumulation, allowing the model to better differentiate crop species.

While the EMET framework possesses its unique advantages and shows promising performance in NRT crop mapping, there still exist limitations. The thermal time metric, AGDD, is employed in EMET for NRT normalization of crop phenology. While the employment of AGDD improves model performance, it might also bring uncertainty in the phenology normalization process. For example, spatiotemporal variation in climatic conditions (e.g., water stress) may affect the responses of crop species to heat accumulation. In different regions, management practices (e.g., cultivar selection, planting timing, and fertilization) usually adapt to local environmental conditions in order to optimize crop yield, which may alter the relationship between crop growth and heat accumulation. Therefore, future work should be focused on more comprehensive NRT modeling of crop phenology, through the integration of a combination of environmental (e.g., water availability and soil properties) and anthropogenic factors (e.g., planting dates and cultivar) derived from satellite remote sensing and auxiliary data sources. In addition, the AGDD accumulation relies on the WISE-detected emergence dates, yet the validation of WISE results is mainly based on state-level CPR due to the limited availability of field-scale crop emergence records. In the future, leveraging in-situ phenology observations for validation at finer spatial scales will help us better understand model performance across space and time.

While the EMET framework can well capture the interspecific difference for NRT crop mapping, its ability in reducing the intraspecific difference over an extended area (e.g., intercontinental transfer) needs to be further evaluated. While the spatial transferability has been tested across the US Corn Belt, intercontinental transfer is faced with more drastic differences in phenological progress, which may require more comprehensive design to account various climatic and human factors for phenology normalization. As one of the most important crop production areas in the world, the U.S. Corn Belt is predominantly planted with corn and soybeans, which are selected as the two crop types of interest. In the future, we may explore the applicability of the proposed framework in other crop production regions in which more crop types and/or more complex landscapes may be observed. This study incorporates CDL for model building and validation, given the relatively high accuracy of CDL for both corn and soybeans classes. Yet the CDL quality may be compromised in study areas where more crop species are planted. Future studies may benefit from the integration of field observation data for model validation. Spatiotemporal image fusion can benefit time-series remote sensing applications by providing dense satellite time series at desired spatial and temporal resolutions. In recent years, satellite image products have become increasingly available (e.g., Landsat-9, PlanetScope, and Harmonized Landsat Sentinel-2) with improved spatial and temporal resolution characteristics. Synthesizing multi-source satellite data through more advanced data fusion models may represent another direction for future improvements. With the advent of new technologies and datasets, such as the Segment Anything Model (Kirillov et al., 2023) or the recently published Crop Sequence Boundaries from USDA (Abernethy et al., 2023; Hunt et al., 2023), aggregating pixel-level mapping results into field boundaries is promising for further enhancing model performance.

With the EMET framework, crop types can be identified in NRT as early as late June-early July. Such early identification of crop types provides critical information for relevant stakeholders with significant implications for food security, agricultural risk and damage assessment, and supply chain management. Satellite-based crop maps can serve as a complementary means to the traditional survey-based methods for estimating crop acreage and production. Since the crop acreage can be estimated with low latency, NRT crop mapping holds a large potential in

timely informing decision-makers to help them better understand the volatility in food market, identify early signs of food insecurity, and make proactive management plans accordingly. With the spatiotemporal fusion dataset at a 30-meter resolution, crop type information can not only be used for regional-scale acreage estimation, but also facilitate applications in precision agriculture, such as the estimation of crop growth progress, health conditions, and yields for specific crop fields, which can further contribute to more sustainable agricultural production.

6. Conclusion

This study proposes the EMET framework for NRT crop mapping through a systematic design that encompasses spatiotemporal image fusion, crop thermal-based phenology normalization, and deep learning. The framework is evaluated in the U.S. Corn Belt, trained with satellite observations from the IL site in 2017, and tested in both IL and MN sites in 2019 and 2020. The hybrid deep learning fusion model generates high spatiotemporal resolution imagery that timely captures rapid temporal phenological changes during the growing season. The thermal-based NRT phenology normalization process registers crop phenological progress across space and time through integrating the WISE-derived crop emergence dates and heat accumulation estimated by AGDD. The normalized time-series data enable the SAtLSTM model to better identify different crop types despite the spatiotemporal variations in crop phenological patterns. The EMET framework can be implemented in NRT, and the mapping results can be generated within a few days after satellite image acquisition. Compared to calendar and WISE scenarios, EMET demonstrates enhanced scalability with improved field-level crop mapping results across study sites and testing years throughout the growing season. EMET achieves an overall accuracy of 85%, approximately four weeks earlier than the benchmark scenarios; crop types can be accurately identified by EMET as early as late July with an overall accuracy over 90%. NRT crop type mapping can provide valuable information for reducing the volatility in the food market and enhancing food security. The 30-m crop type maps can further facilitate a variety of agricultural applications at the field level, including but not limited to crop growth progress characterization, health condition estimation, and yield prediction, holding great potential to optimize crop managements and facilitate precision agriculture.

CRediT authorship contribution statement

Zijun Yang: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chunyuan Diao:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Feng Gao:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Bo Li:** Writing – review & editing, Methodology.

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.

Acknowledgements

This work was supported partly by the National Science Foundation (number 2048068), partly by the National Aeronautics and Space Administration (number 80NSSC21K0946), and partly by the United States Department of Agriculture (number 2021-67021-33446). The USDA is an equal opportunity provider and employer. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2024.07.007.

References

Abernethy, J., Beeson, P., Boryan, C., Hunt, K., Sartore, L., 2023. Preseason crop type prediction using crop sequence boundaries. Comput. Electron. Agric. 208, 107768.
Akyuz, F.A., Kandel, H., Morlock, D., 2017. Developing a growing degree day model for North Dakota and Northern Minnesota soybean. Agric. For. Meteorol. 239, 134–140.
Angel, J.R., Widhalm, M., Todey, D., Massey, R., Biehl, L., 2017. The U2U corn growing degree day tool: Tracking corn growth across the US Corn Belt. Clim. Risk Manag. 15, 73–81.
Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., Brivio, P.A., 2016. Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. European J. Remote Sens. 49, 361–380.
Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., Ferraz, P.D.R., Lebourgeois, V., Lelong, C., Simões, M., Verón, R.S., 2018. Remote sensing and cropping practices: A review. Remote Sens. (Basel) 10, 99.
Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523.
Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. Geocarto Int. 26, 341–358.
Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. 210, 35–47.
Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161.
Diao, C., 2019. Innovative pheno-network model in estimating crop phenological stages with satellite time series. ISPRS J. Photogramm. Remote Sens. 153, 96–109.
Diao, C., 2020. Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. Remote Sens. Environ. 248, 111960.
Diao, C., Yang, Z., Gao, F., Zhang, X., 2021. Hybrid phenology matching model for robust crop phenological retrieval. ISPRS J. Photogramm. Remote Sens. 181, 308–326.
Feng, S., Zhao, J., Liu, T., Zhang, H., Zhang, Z., Guo, X., 2019. Crop type identification and mapping using machine learning algorithms and sentinel-2 time series data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 3295–3306.
Gao, F., Zhang, X., 2021. Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. *J. Remote Sens., 2021*.
Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44, 2207–2218.
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.
Gao, F., Anderson, M., Daughtry, C., Karnieli, A., Hively, D., Kustas, W., 2020a. A within-season approach for detecting early growth stages in corn and soybean using high temporal and spatial resolution imagery. Remote Sens. Environ. 242, 111752.
Gao, F., Anderson, M.C., Hively, W.D., 2020b. Detecting cover crop end-of-season using VENμS and sentinel-2 Satellite imagery. Remote Sens. (Basel) 12, 3524.
Gao, F., Anderson, M.C., Johnson, D.M., Seffrin, R., Wardlow, B., Suyker, A., Diao, C., Browning, D.M., 2021. Towards Routine Mapping of Crop Emergence within the Season Using the Harmonized Landsat and Sentinel-2 Dataset. Remote Sens. (Basel) 13, 5074.
Gao, F., Jennewein, J., Hively, W.D., Soroka, A., Thieme, A., Bradley, D., Keppler, J., Mirsky, S., Akumaga, U., 2023. Near real-time detection of winter cover crop termination using harmonized Landsat and Sentinel-2 (HLS) to support ecosystem assessment. Sci. Remote Sens. 7, 100073.
Gu, L., Post, W.M., Baldocchi, D.D., Black, T.A., Suyker, A.E., Verma, S.B., Vesala, T., Wofsy, S.C., 2009. Characterizing the seasonal dynamics of plant community photosynthesis across a range of vegetation types. *Phenol. Ecosyst. Proces.: Appl. Global Change Res.*, 35-58.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Hunt, K.A., Abernethy, J., Beeson, P., Bowman, M., Wallander, S., Williams, R., 2023. Crop Sequence Boundaries (CSB): Delineated Fields Using Remotely Sensed Crop Rotations. USDA-NASS, Washington, DC, USA. https://www.nass.usda.gov/Research_and_Science/Crop-Sequence-Boundaries/index.php.
Johnson, D.M., Mueller, R., 2021. Pre-and within-season crop type classification trained with archival land cover information. Remote Sens. Environ. 264, 112576.
Kerner, H.R., Sahajpal, R., Pai, D.B., Skakun, S., Puricelli, E., Hosseini, M., Meyer, S., Becker-Reshef, I., 2022. Phenological normalization can improve in-season classification of maize and soybean: A case study in the central US Corn Belt. Sci. Remote Sens. 6, 100059.
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., 2023. Segment anything. In, *Proceedings of the IEEE/CVF International Conference on Computer Vision,* pp. 4015-4026.
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.
Kluger, D.M., Wang, S., Lobell, D.B., 2021. Two shifts for crop mapping: Leveraging aggregate crop statistics to improve satellite-based maps in new regions. Remote Sens. Environ. 262, 112488.
Konduri, V.S., Kumar, J., Hargrove, W.W., Hoffman, F.M., Ganguly, A.R., 2020. Mapping crops within the growing season across the United States. Remote Sens. Environ. 251, 112048.
Liao, C., Wang, J., Shan, B., Shang, J., Dong, T., He, Y., 2023. Near real-time detection and forecasting of within-field phenology of winter wheat and corn using Sentinel-2 time-series data. ISPRS J. Photogramm. Remote Sens. 196, 105–119.
Lin, C., Zhong, L., Song, X.-P., Dong, J., Lobell, D.B., Jin, Z., 2022. Early-and in-season crop type mapping without current-year ground truth: Generating labels from historical information via a topology-based approach. Remote Sens. Environ. 274, 112994.
Liu, L., Zhang, X., Yu, Y., Gao, F., Yang, Z., 2018. Real-time monitoring of crop phenology in the Midwestern United States using VIIRS observations. Remote Sens. (Basel) 10, 1540.
Nguyen, L.H., Joshi, D.R., Clay, D.E., Henebry, G.M., 2020. Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier. Remote Sens. Environ. 238, 111017.
Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. (Basel) 11, 523.
Pörtner, H.-O., Roberts, D.C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R.A., Betts, R., Kerr, R.B., Biesbroek, R., 2022. Climate change 2022: Impacts, adaptation and vulnerability. IPCC Sixth Assessment Report.
Prosekov, A.Y., Ivanova, S.A., 2018. Food security: The challenge of the present. Geoforum 91, 73–77.
Qian, Y., Yang, Z., Di, L., Rahman, M.S., Tan, Z., Xue, L., Gao, F., Yu, E.G., Zhang, X., 2019. Crop growth condition assessment at county scale based on heat-aligned growth stages. Remote Sens. (Basel) 11, 2439.
Rußwurm, M., Körner, M., 2020. Self-attention for raw optical satellite time series classification. ISPRS J. Photogramm. Remote Sens. 169, 421–435.
Schaaf, C., Wang, Z., 2015. MCD43A4 MODIS/Terra+ Aqua BRDF/Albedo Nadir BRDF Adjusted RefDaily L3 Global 500 m V006. *NASA EOSDIS Land Processes DAAC*.
Serra, J., Arcos, J.L., 2014. An empirical evaluation of similarity measures for time series classification. Knowl.-Based Syst. 67, 305–314.
Shen, Y., Zhang, X., Yang, Z., Ye, Y., Wang, J., Gao, S., Liu, Y., Wang, W., Tran, K.H., Ju, J., 2023. Developing an operational algorithm for near-real-time monitoring of crop progress at field scales by fusing harmonized Landsat and Sentinel-2 time series with geostationary satellite observations. Remote Sens. Environ. 296, 113729.
Song, H., Liu, Q., Wang, G., Hang, R., Huang, B., 2018. Spatiotemporal Satellite Image Fusion Using Deep Convolutional Neural Networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 821–829.
Soudani, K., Le Maire, G., Dufrêne, E., François, C., Delpierre, N., Ulrich, E., Cecchini, S., 2008. Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. Remote Sens. Environ. 112, 2643–2655.
Tan, B., Morisette, J.T., Wolfe, R.E., Gao, F., Ederer, G.A., Nightingale, J., Pedelty, J.A., 2010. An enhanced TIMESAT algorithm for estimating vegetation phenology metrics from MODIS data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 4, 361–371.
Thornton, M., Shrestha, R., Wei, Y., Thornton, P., Kao, S., Wilson, B., 2022. Daymet: daily surface weather data on a 1-km grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. In.
Thornton, P.E., Shrestha, R., Thornton, M., Kao, S.-C., Wei, Y., Wilson, B.E., 2021. Gridded daily weather data for North America with comprehensive uncertainty quantification. Sci. Data 8, 190.
USDA-NASS, 2022. https://www.nass.usda.gov/Publications/National_Crop_Progress/.
Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Proces. Syst. 30.
Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. Remote Sens. Environ. 222, 303–317.
Wang, Y., Feng, L., Zhang, Z., Tian, F., 2023. An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery. ISPRS J. Photogramm. Remote Sens. 199, 102–117.
Wang, Z., Zhang, H., He, W., Zhang, L., 2022. Cross-phenological-region crop mapping framework using Sentinel-2 time series Imagery: A new perspective for winter crops in China. ISPRS J. Photogramm. Remote Sens. 193, 200–215.
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.
Xu, Y., Ma, Y., Zhang, Z., 2024. Self-supervised pre-training for large-scale crop mapping using Sentinel-2 time series. ISPRS J. Photogramm. Remote Sens. 207, 312–325.
Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H., Lin, T., 2020. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. Remote Sens. Environ. 247, 111946.
Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K., Ying, Y., Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. Remote Sens. Environ. 264, 112599.
Yang, Z., Diao, C., Li, B., 2021. A Robust Hybrid Deep Learning Model for Spatiotemporal Image Fusion. Remote Sens. (Basel) 13, 5005.
Yang, Z., Diao, C., Gao, F., 2023. Towards Scalable Within-Season Crop Mapping with Phenology Normalization and Deep Learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.
Yaramasu, R., Bandaru, V., Pnvr, K., 2020. Pre-season crop type mapping using deep neural networks. Comput. Electron. Agric. 176, 105664.
You, N., Dong, J., Li, J., Huang, J., Jin, Z., 2023. Rapid early-season maize mapping without crop labels. Remote Sens. Environ. 290, 113496.
Zeng, L., Wardlow, B.D., Wang, R., Shan, J., Tadesse, T., Hayes, M.J., Li, D., 2016. A hybrid approach for detecting corn and soybean phenology with time-series MODIS data. Remote Sens. Environ. 181, 237–250.
Zhang, C., Diao, C., 2023. A Phenology-guided Bayesian-CNN (PB-CNN) framework for soybean yield estimation and uncertainty analysis. ISPRS J. Photogramm. Remote Sens. 205, 50–73.
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.
Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. Remote Sens. Environ. 140, 1–13.
Zhong, L., Yu, L., Li, X., Hu, L., Gong, P., 2016. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. Sci. Rep. 6, 1–14.
Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443.
Zhu, X.L., Cai, F.Y., Tian, J.Q., Williams, T.K.A., 2018. Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. Remote Sens. (Basel) 10, 527.
Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36.
Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G., 1999. Unmixing-based multisensor multiresolution image fusion. IEEE Trans. Geosci. Remote Sens. 37, 1212–1226.