Remote Sensing of Environment 331 (2025) 115016

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A novel Normalized Harvest Phenology Index (NHPI) for corn and soybean harvesting date detection using Landsat and Sentinel-2 imagery on Google Earth Engine

Yin Liu [a], Chunyuan Diao [a,*], Zijun Yang [b,c], Weiye Mei [d], Tianci Guo [a]

[a] Department of Geography and Geographic Information Science, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
[b] Department of Earth and Ocean Sciences, University of North Carolina Wilmington, Wilmington, NC 28403, USA
[c] Center for Marine Science, University of North Carolina Wilmington, Wilmington, NC 28409, USA
[d] Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Edited by Marie Weiss

Keywords:
Harvesting date
Phenology
Vegetation index
Agriculture
Google Earth Engine

ABSTRACT

The timing of harvesting is crucial for determining crop yield potential as it influences the final stages of the crop growth cycle and affects crop grain quality. Early harvesting can lead to yield losses from excessive moisture and insufficient dry matter, while delayed harvesting can degrade grain quality due to over-maturation and increased susceptibility to weather, pests, and diseases. Accurate monitoring of harvest timing is essential to assess yield gaps, support profitable and sustainable farming practices, and optimize agricultural supply chains. However, remote sensing-based harvesting date detection methods often suffer from biases due to the inconsistent relationship between end-of-season (EOS) metrics in vegetation index (VI) time series and actual harvesting dates. This inconsistency occurs because harvesting decisions are often influenced by human factors such as equipment availability, labor constraints, and fuel costs, rather than plant condition alone. In this study, we develop a novel Normalized Harvest Phenology Index (NHPI) that integrates the Normalized Difference Vegetation Index (NDVI) and the Near-Infrared (NIR) reflectance to accurately monitor whether fields of corn and soybean have been harvested. Leveraging the distinct separability of NIR reflectance for corn and soybean before harvesting (senescent plants) and after harvesting (crop residue), combined with the contrasting trends between NIR and NDVI during this transition, the NIR-to-NDVI ratio amplifies the harvesting signal in its time series, making it a robust indicator of harvesting events. As the first spectral index designed for scalable identification of crop harvesting stage, the developed NHPI is applied to map harvesting dates for corn and soybean fields across the U.S. Midwest from 2020 to 2023 using Landsat and Sentinel-2 imagery via Google Earth Engine (GEE). At the field level, the NHPI-based harvesting date estimation method achieves a mean absolute error (MAE) of 4 days and an R^2 of 0.85 when compared against field-recorded harvesting dates, significantly outperforming all advanced harvesting date estimation benchmarks (i.e., EOS phenometric-based method, shape model fitting method (SMF), and shape model fitting by the separate phenological stage method (SMF-S). The NHPI-based harvesting date mapping also shows strong alignment with the state-level cumulative distribution of harvesting dates of the USDA crop progress reports, achieving an average MAE of 3 days. Further analysis of NHPI values before and after harvesting events reveals its strong adaptability to diverse weather conditions at large scales, highlighting its efficiency and robustness.

1. Introduction

Food security will face growing challenges in the coming years due to increasing population, dwindling farmland, shifts in consumption habits, and climate change (Beddington, 2010). In response to these evolving conditions, adapting crop management practices will be crucial for addressing food security concerns (Liu et al., 2023). As one of the major crop management practices, the timing of harvesting is critical for ensuring that the grain reaches its highest possible quality and maximizing crop yield (Kusumastuti et al., 2016; Xu et al., 2019). Early

---

* Corresponding author.
E-mail addresses: yinl3@illinois.edu (Y. Liu), chunyuan@illinois.edu (C. Diao).

https://doi.org/10.1016/j.rse.2025.115016
Received 12 April 2025; Received in revised form 26 July 2025; Accepted 5 September 2025

harvesting can result in yield losses due to excessive grain moisture and insufficient dry matter accumulation, whereas delayed harvesting may degrade grain quality due to over-maturation, increased exposure to adverse weather, and higher susceptibility to pests and diseases (Kim et al., 2024). Furthermore, the timing of harvesting affects a range of supply chain activities, including labor distribution, machinery utilization, storage, processing, logistics, and grain market trading (Mardaneh et al., 2021). Harvesting date information is essential for understanding real-world farm practices and evaluating management efficiency at scale. It also informs models that monitor crop phenology, estimate crop yield, and support agricultural carbon accounting (Pugh et al., 2015; Zhang and Diao, 2023). As such, accurate detection of harvesting dates plays a key role in strengthening agricultural monitoring systems and enhancing food system resilience (Chandra et al., 2024).

Field surveying is the most traditional method for collecting crop harvesting ground truth data, recording the exact timing when farmers harvest their fields, and providing accurate harvesting date information. However, conducting these surveys across large geographic regions is labor-intensive, time-consuming, and expensive. Remote sensing addresses these limitations by providing extensive geographic coverage, rapid observations, and significantly lower costs. Since the 1990s, remote sensing time series analysis techniques have been increasingly utilized for monitoring crop harvesting (Reed et al., 1994). These methods can be broadly classified into three categories: (1) phenometric-based detection methods, (2) phenology matching methods, and (3) interferometric synthetic aperture radar (InSAR) coherence analysis methods.

Phenometric-based detection methods estimate harvesting dates using end-of-season (EOS) phenometrics derived from satellite time series of vegetation indices (VIs). These EOS phenometrics are typically extracted using specific thresholds (e.g., 10 % or 20 % of the curve amplitude) (Chen et al., 2004a; Delbart et al., 2006; Zeng et al., 2020) or curve inflection points (e.g., local minima in the rate of curvature change) (Diao, 2020; Diao and Li, 2022; Gao et al., 2017; Moulin et al., 1997; Schwartz et al., 2002; Wu et al., 2017; Zhang et al., 2003). Once identified, these EOS metrics are assumed to correspond to the field's harvesting date. However, the accuracy of detected harvesting dates is limited due to the inconsistent relationship between VIs-based EOS metrics and actual harvesting dates. This inconsistency arises because EOS metrics typically correspond to the physiological stage of crop senescence, whereas actual harvesting dates are influenced by human decisions based on factors such as labor availability, equipment scheduling, and fuel costs, and may therefore occur at different points along the VI trajectory. Consequently, EOS metrics lack a direct physical linkage to actual harvesting events, limiting their ability to capture variations in farmers' harvesting practices across space and time (Gao and Zhang, 2021). Applying the phenometric-based methods to retrieve the EOS metrics across different regions or crop species could thus lead to uncertainty and inaccuracy in harvesting date estimation.

Phenology matching methods estimate crop harvesting dates of target fields using pre-defined crop reference phenological time series and associated reference harvesting dates (Sacks and Kucharik, 2011; Diao et al., 2021; Dong et al., 2019; Liu et al., 2022a, 2022b; Shen et al., 2023; Cao et al., 2024). Representative phenology matching methods include the shape model fitting (SMF) method and the hybrid phenology matching method (Sakamoto et al., 2010; Diao et al., 2021). These methods usually rely on geometrical pattern matching techniques to map the reference phenological time series onto the target satellite time series by estimating mapping parameters (i.e., temporal shifts and scaling factors). These parameters quantify the adjustments required for the reference phenological trajectory to best align with the target field's time series. Once aligned, the reference harvesting date of the reference phenological time series is transformed using the derived mapping parameters to estimate the harvesting date for the target time series. The crop reference phenological time series are typically constructed by aggregating Vegetation Index (VI) time series of regional ground phenological observations to capture crops' typical seasonal growth patterns. The corresponding reference harvesting date of the time series is pre-defined or calibrated using ground harvesting date observing records. As a result, the estimated harvesting dates for the target time series are largely determined by the crop reference time series and reference harvesting dates (Zeng et al., 2020; Gao and Zhang, 2021; Diao et al., 2021). However, these crop phenological references are typically created with limited field observations, resulting in a lack of representativeness and accuracy (Sakamoto, 2018a; Diao et al., 2021). This limitation is particularly concerning for harvesting date detection, as the harvesting process is heavily influenced by human activity, resulting in significant variability across locations and years. Obtaining reliable crop reference harvesting dates therefore demands more diverse and comprehensive field observations than those required for other crop phenological stages, which is challenging at large scales.

The third type of methods, InSAR coherence analysis, leverages radar satellite data to infer harvesting dates through analyzing coherence between successive pairs of radar images to track changes in field conditions (Schlund and Erasmi, 2020; Shang et al., 2020; Amherdt et al., 2021; Pandit et al., 2022). InSAR coherence, calculated as a complex correlation coefficient between two adjacent radar images, measures the consistency of radar signal phases, which reflects surface stability. During harvesting, disturbances to the field surface could reduce surface stability and thus cause a drop in coherence value, typically followed by a sharp increase as the field stabilizes post-harvest. This characteristic pattern enables the identification of harvesting events in the coherence time series with defined criteria. This radar-based approach is effective particularly in cloudy regions, where optical sensors are limited, as radar can penetrate clouds and provide consistent observations regardless of weather or lighting conditions (Liu et al., 2022a, 2022b). Sentinel-1 is the most commonly used data source for InSAR-based agricultural monitoring, providing a revisit frequency of 6 to 12 days depending on geographic location. However, factors like rapid crop phenological changes, management practices, and surface moisture changes around the harvesting period can influence radar signal behavior, potentially resulting in low coherence values followed by high ones. This can produce signals similar to those detected by the harvesting identification criteria, leading to potential false positive detections (Löw et al., 2021, 2024). Additionally, inherent noise in InSAR data can disrupt phase consistency, complicating accurate coherence estimation and thus the identification of the harvesting date. Small fluctuations or noises in the coherence time series can affect harvesting event detection, leading to biased results (Xu et al., 2020; Nikaein et al., 2021; Wang et al., 2022).

The inconsistent relationship between EOS metrics and harvesting dates, the difficulty in defining phenological references with tailored reference harvesting dates, and the inherent noise in InSAR coherence data collectively highlight the limitations of current methods for accurate detection of crop harvesting dates. Field harvesting events involve significant surface changes, marked by the transition from crop-covered fields to crop residuals/bare soil, which likely exhibit unique patterns in certain spectral characteristics. Yet the spectral dynamics associated with these changes remain poorly understood, limiting the development of consistent and interpretable indicators across different regions and crop types. Jiang et al. (2024) recently proposed a machine learning-based approach that leverages all eight spectral bands and derived indices from PlanetScope imagery as input features to detect harvesting events. While this data-driven method shows potential, its reliance on large, high-quality training datasets limits its scalability and generalizability.

By analyzing a diversity of optical spectral dynamic changes during the harvesting season, we develop a novel Normalized Harvest Phenology Index (NHPI) to accurately retrieve crop harvesting dates. The NHPI is designed to achieve three key objectives: (1) enhance the spectral dynamic changes associated with field surface changes pertaining to harvesting, (2) accurately characterize the timing of the harvesting event from NHPI time series, and (3) facilitate robust, large-scale harvesting date mapping on the Google Earth Engine (GEE) platform. The proposed NHPI index is leveraged to generate harvesting date maps in the U.S. Midwest for corn and soybean under diverse climate conditions, landscape complexities, and cropping systems. The performance of the NHPI is evaluated upon comparison with three advanced benchmark methods: the EOS phenometric-based method (Diao, 2020), the shape model fitting (SMF) method (Sakamoto et al., 2010), and the shape model fitting by the separate phenological stage (SMF-S) method (Liu et al., 2022a, 2022b).

2. Study area and data

2.1. Study area

The study area spans 12 major agricultural states in the U.S. Midwest: Iowa (IA), Illinois (IL), Nebraska (NE), Minnesota (MN), Indiana (IN), Ohio (OH), South Dakota (SD), North Dakota (ND), Kansas (KS), Missouri (MO), Michigan (MI), and Wisconsin (WI) (Fig. 1). The U.S. Midwest is a vital region for global food production, dominated by the cultivation of corn and soybean as the primary crop species. The Midwest accounts for approximately 35 % of global corn production and about 31 % of global soybean production (Lobell et al., 2014). The diversity of environmental conditions in this region, including variations in climate, soil types, and farm management practices, creates an ideal setting for evaluating large-scale harvesting date estimation methods. The study period spans from 2020 to 2023.

2.2. Data

To analyze the optical spectral dynamical change before and after harvesting events, we conduct field surveys in Champaign County, IL, during the 2022 and 2023 harvesting seasons, to obtain ground truth information on field conditions (Fig. S1). Surveys are conducted every three days to record field statuses, including harvesting statuses and tillage conditions. The dataset includes records for 104 soybean and 107 corn parcels in 2022, and 105 soybean and 110 corn parcels in 2023. These field parcels span a range of soil types and microclimatic conditions within the county, capturing variability that may drive differences in harvesting decisions and providing a comprehensive dataset for interpreting spectral change patterns. On average, monitored corn fields are approximately 25.8 ± 18.0 ha in size, while soybean fields are about 20.8 ± 14.0 ha. Approximately half of the surveyed fields experience tillage operations during the observation period.

To evaluate the performance of our proposed NHPI-based harvesting date estimation method and three advanced benchmark methods at large scale, we utilize field-level corn and soybean harvesting data from Beck's Hybrids, a well-known agricultural seed company that publishes crop yield and management data. The dataset is publicly accessible at https://www.beckshybrids.com/Research/Yield-Data. The data are widely distributed across the U.S. Midwest, with the exception of ND (Fig. 1). This dataset includes 161 soybean fields in 2021, 253 in 2022, and 263 in 2023, along with 367 corn fields in 2021, 446 in 2022, and 479 in 2023.

Besides the field-level data, we further utilize the Crop Progress Report (CPR) data (USDA-NASS, 2024) to assess the harvesting date mapping capability of our proposed NHPI-based method at state level. This dataset provides weekly state-level cumulative statistics on major crops (e.g., corn and soybean) when they reach specific phenological stages of the study region, including key stages like maturity and harvesting. Additionally, to ensure our analysis focuses specifically on corn and soybean, we use the Cropland Data Layer (CDL), an annual crop type mapping dataset across the U.S., to identify the pixels of corn and soybean within the study area (Boryan et al., 2011). By aggregating NHPI-based harvesting date mapping results for corn or soybean to the state level annually, we compare NHPI-based results with CPR data and

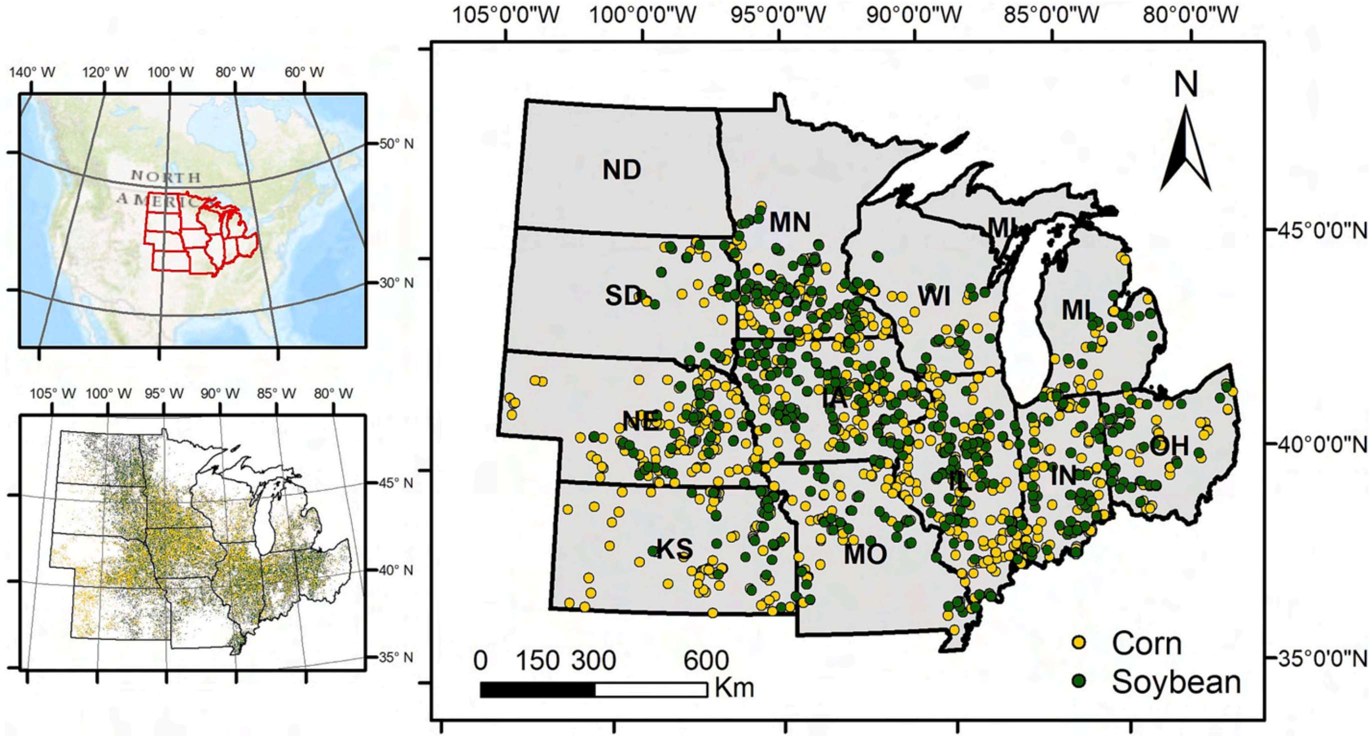


Fig. 1. Spatial distribution of corn and soybean harvesting date records from Beck's datasets (2021–2023) across 12 U.S. Midwestern states. Yellow points indicate recorded corn harvest dates, while green points represent recorded soybean harvest dates. The top-left inset map shows the geographic extent of the study region (in red) in North America. The bottom-left inset map highlights the spatial coverage of corn and soybean fields within the study area based on the 2022 Cropland Data Layer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

evaluate the method's accuracy and robustness across diverse agricultural landscapes in the U.S. Midwest.

The primary satellite imagery used for harvesting date estimation includes imagery from Landsat 7, 8, and 9, along with Sentinel-2 on GEE. We use Landsat and Sentinel-2 imagery from 2020 to 2023 to align with the study period. Landsat imagery offers a spatial resolution of 30 m, while Sentinel-2 provides higher spatial resolutions of 10 m for visible and near-infrared bands and 20 m for red-edge and two shortwave infrared bands. Sentinel-2 data are resampled to 30 m using the default nearest neighbor method on GEE to match the spatial resolution of Landsat imagery, ensuring consistency in spatial resolution and alignment for the following analysis. We utilize Landsat and Sentinel-2 imagery to retrieve the crop harvesting dates at the field level, with the farm field sizes in the U.S. Midwest averaged around 50 acres (Yan and Roy, 2016). The temporal resolutions of these datasets, with a 16-day revisit interval for each Landsat sensor and a 5-day interval for Sentinel-2, enable the construction of a relatively dense time series that increases the likelihood of capturing critical harvesting events, even with occasional cloud cover. Observation availability improves after 2022 with the integration of Landsat 9, which complements Landsat 8 to reduce the revisit interval to 8 days, further enhancing temporal coverage. To further enhance the consistency of observations, we harmonize the Sentinel-2 and Landsat datasets on GEE using the least square method, which reconciles their spectral and radiometric differences (Roy et al., 2016; Zhang et al., 2018; Fan et al., 2023). This harmonization creates a seamless dataset that enables more reliable and continuous tracking of crop growth and harvesting activities. As the official Harmonized Landsat-Sentinel (HLS) product is not currently available on the GEE platform, we instead perform our own harmonization within GEE to take full advantage of its data access and processing capabilities.

To calibrate the SMF and SMF-S benchmark methods (Sakamoto et al., 2010; Sakamoto, 2018b; Diao et al., 2021), we further utilize the MODIS MCD43A4 (version 6) nadir Bidirectional Reflectance Distribution Function (BRDF) adjusted reflectance product. This dataset provides consistent daily surface reflectance data at a 500-m resolution, which enables the building of high-quality satellite time series for characterizing crop growth patterns at a regional scale. We also evaluate the impact of weather conditions on spectral reflectance dynamics during the harvesting period using the DAYMET dataset (Thornton et al., 2020), which provides daily precipitation and temperature data at a 1-km resolution. Precipitation data is specifically analyzed to understand rainfall's impact on spectral reflectance and index values, as water can significantly alter surface characteristics.

3. Method

In this study, we develop a novel Normalized Harvest Phenology Index (NHPI) that combines the Normalized Difference Vegetation Index (NDVI) and the Near-Infrared (NIR) reflectance to enhance the spectral dynamic changes associated with field surface changes pertaining to harvesting. As shown in Fig. 2, the development of NHPI begins with a spectral dynamic analysis using Landsat and Sentinel-2 imagery. This analysis is guided by field survey data collected in Champaign County (2022–2023), which provides insights into field conditions before and after harvest. The NHPI is developed by first examining unique spectral dynamics across multiple spectral bands during the corn and soybean harvesting season. This analysis identifies NIR as a distinct indicator of harvest timing because of its sensitivity to the transition from senescent plants to crop residue during harvesting events and its low susceptibility to rainfall interference. Additionally, NIR aligns with NDVI's trend before harvest and exhibits an opposite trend after harvest, driven by changes in leaf structure and the physical properties of crop residue. As crops enter senescence, the breakdown of internal leaf structures reduces NIR scattering and lowers reflectance. After harvest, the presence of dry, fibrous residue increases NIR reflectance, while NDVI continues to decline during this transition. Building on these findings, NIR and NDVI are integrated by calculating their ratio to construct the NHPI. This integration amplifies harvest-related signals, resulting in a more robust and reliable index for identifying whether fields have been

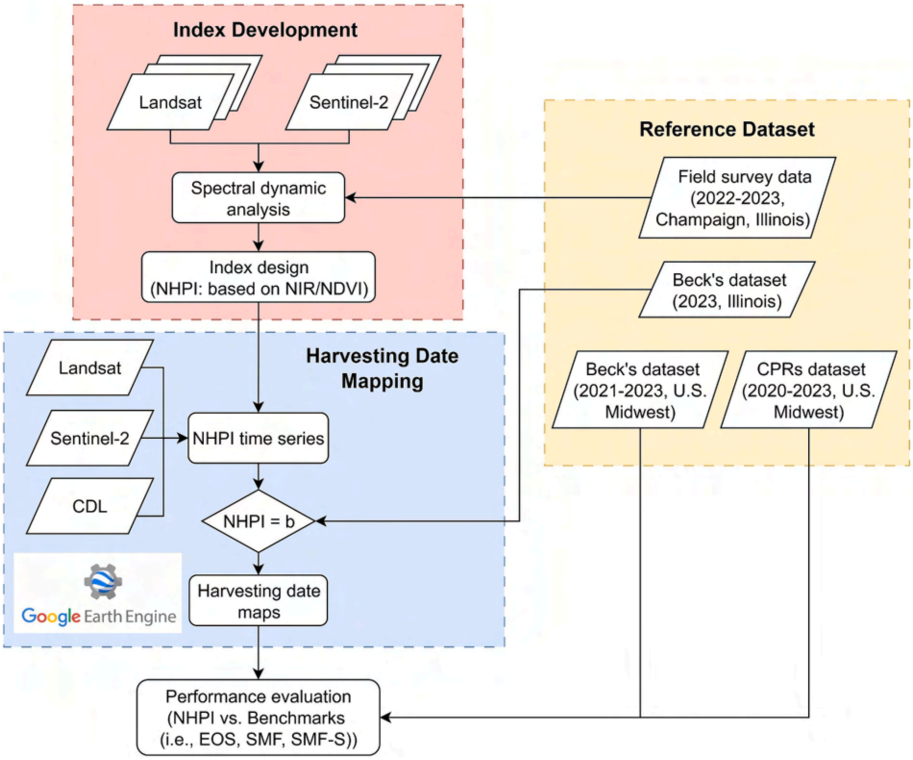


Fig. 2. Workflow of this study illustrating the development of the NHPI and its application in harvesting date mapping.

harvested. The innovation of NHPI lies in not only its unique spectral characteristic design but also its computational efficiency. Through the GEE platform, we construct NHPI time series with Landsat and Sentinel-2 imagery to map harvesting dates, defined as the first date when NHPI exceeds a predefined threshold, for corn and soybean fields from 2020 to 2023 across the U.S. Midwest. The threshold is calibrated using the Illinois 2023 subset of the Beck dataset. To evaluate the performance of the NHPI-based method at the field level, we compare NHPI-estimated harvesting dates with ground-truth data from the Beck dataset across the U.S. Midwest and benchmark the NHPI against three advanced harvesting date estimation methods. Furthermore, we assess the NHPI at the state level by comparing the cumulative distributions of NHPI-generated harvesting dates with those recorded in the CPRs, providing an additional layer of validation for its large-scale applicability. The codes are open source: https://github.com/rssiuiuc/NHPI/.

3.1. Development of normalized harvest phenology index (NHPI)

3.1.1. Temporal signature analysis of spectral reflectance

Based on field observations in Champaign, IL, we identify a typical sequence of three stages: crop senescence (stage 1), crop residue (stage 2), and post-harvest tillage (stage 3), as illustrated with representative corn and soybean fields in Fig. 3. The fields shown in Fig. 3 are selected as representative examples because their harvesting and tillage dates are well documented, and they have a high density of cloud-free observations during the late growing season, allowing for a clear depiction of the spectral transitions surrounding harvest. Similar patterns are also observed across many other fields. In the pre-harvest stage, fields are largely covered by senescent crops with drying foliage, creating a coarse texture for the field surface. During harvest, machinery removes the crops, leaving a smoother layer of residue spread across the surface. Depending on the tillage practice in the following weeks, some fields may be tilled, exposing bare soil, while other fields retain the residue layer. Based on harmonized Landsat and Sentinel-2 imagery time series (illustrated in Section 3.2), corn and soybean harvests are both characterized by a sharp increase in reflectance across all spectral bands immediately after harvest, with a subsequent decrease if post-harvest tillage occurs. This observed spike in reflectance is likely due to the exposure of crop residue after stalk removal, as its fibrous texture and high lignin and cellulose content scatter more light than senescent vegetation. In contrast, senescent vegetation reflects less light due to the structural breakdown of leaf tissues and increased shadowing within the canopy. When post-harvest tillage practice removes the residue and exposes bare soil, the reflectance drops to lower levels, comparable to that of senescent crops before harvest. In conclusion, the sharp increase in spectral reflectance observed during the harvest season is a direct and unique result of the harvesting of corn and soybean, which serves an effective indicator for detecting the field harvesting dates.

To identify the most effective spectral indicator for detecting harvesting events, we further investigate the differences in spectral reflectance patterns across crop stages and spectral bands of these field observations. As shown in Fig. 4, senescent crops and post-harvest tillage bare soil have similar reflectance. The contrast between crop residue and senescent plants is more evident in the NIR (wavelength around 800 nm) and shortwave infrared (SWIR) bands, including SWIR1 (wavelength 1100–1700 nm) and SWIR2 (wavelength 2000–2500 nm),

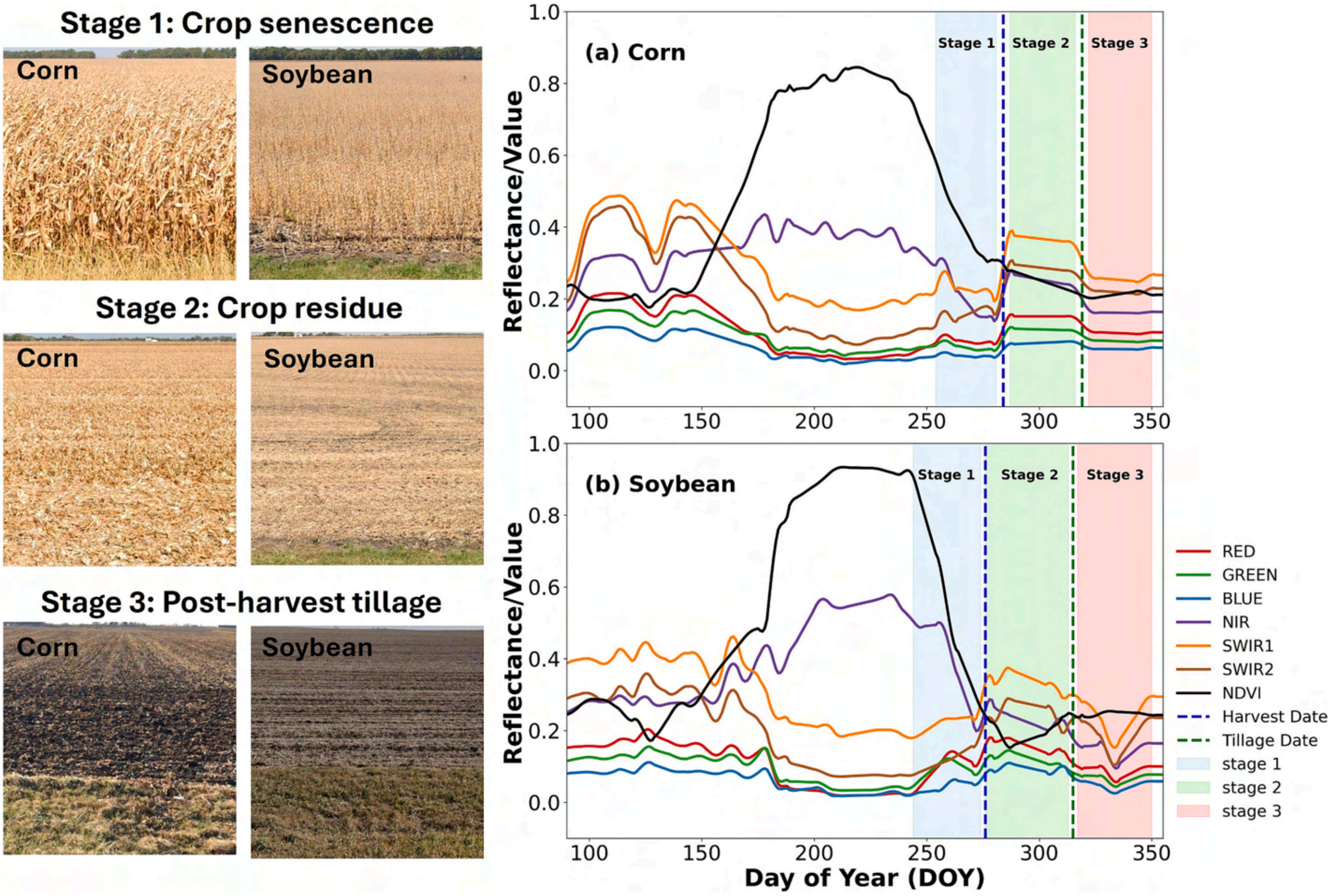


Fig. 3. Spectral reflectance and NDVI time series derived from harmonized Landsat and Sentinel-2 imagery for two representative fields, corn (a) and soybean (b), throughout the growing season. The figure also includes corresponding field photos depicting three key stages around harvesting: stage 1: crop senescence, stage 2: crop residue, and stage 3: post-harvest tillage.

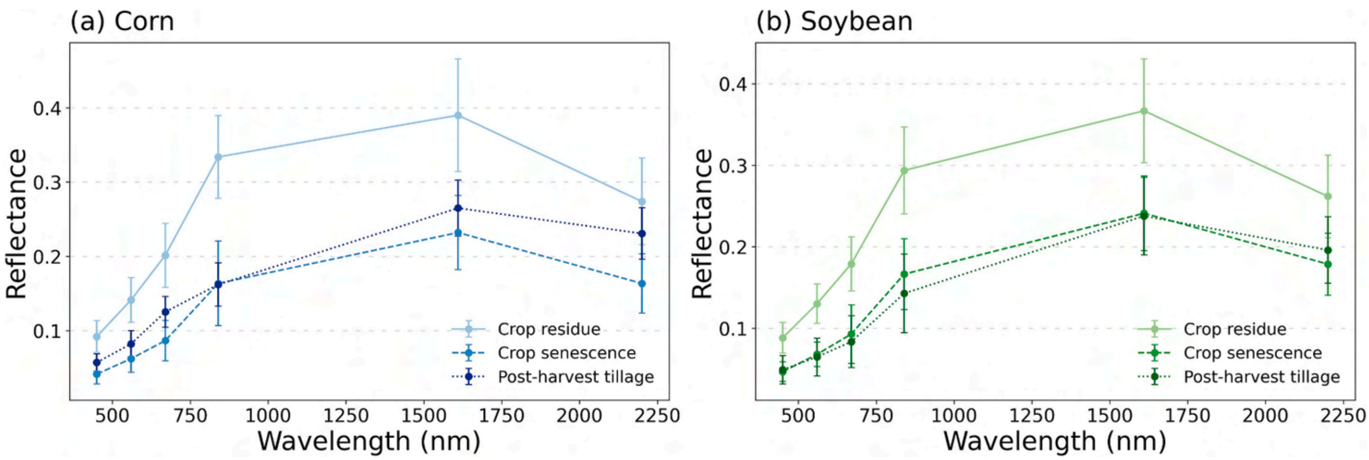


Fig. 4. Spectra for three key stages during the harvesting season for corn and soybean, derived from field observations and harmonized Landsat and Sentinel-2 time series data. These stages include senescent crops prior to harvest, crop residue immediately after harvest, and cropland following tillage. The spectral patterns are based on field survey data collected from 217 corn fields and 209 soybean fields in Champaign, IL. Error bars indicate one standard deviation across the sampled field parcels.

than in the visible spectrum. In the NIR band, healthy plants exhibit high reflectance due to their internal leaf structures, composed of cell walls and air spaces, which efficiently scatter NIR light. As plants enter senescence stage, these structures deteriorate, with cell walls losing their integrity and air spaces within the leaf tissue diminishing. This structural breakdown reduces the scattering of NIR light, leading to lower reflectance and increased light absorption. As senescent plants transition to crop residue during harvesting, their tissues change from partially broken-down structures to dry, fibrous, and fragmented material. This shift leads to higher NIR reflectance due to the loss of organized plant structures, the dominance of reflective fibers (e.g., lignin and cellulose) in crop residue, and the increased scattering from its rough and fragmented surfaces (Daughtry and Hunt, 2008; Baldocchi et al., 2020). In the SWIR band, reflectance is predominantly influenced by water content due to the strong absorption of SWIR light by water molecules. Senescent crops, which retain residual moisture, exhibit moderate reflectance in this spectral range, as the remaining water absorbs a significant portion of the incoming radiation. In contrast, crop residue, characterized by its dry, fibrous composition and fragmented surface, reflects more SWIR light due to the absence of moisture and enhanced scattering properties. Tilled field typically exhibits lower reflectance, as its optical properties are influenced by soil moisture, organic matter, and mineral composition, which collectively contribute to greater light absorption in this spectral region (Hively et al., 2021). This strong separation in NIR and SWIR reflectance between pre-harvest senescent plants and post-harvest residue highlights the potential of these bands as indicators for identifying harvesting events.

3.1.2. *Design of NHPI*

Building on the analysis, we develop the NHPI to estimate harvesting dates by detecting the transition from senescence plant to crop residue from satellite observations. The Harvest Phenology Index (HPI) is defined as the ratio of NIR reflectance to NDVI. To ensure numerical stability, we assume that all observations have NDVI values greater than zero when calculating HPI, as the index is designed for cropland harvesting analysis and cropland pixels typically exhibit positive NDVI values. The HPI is calculated as follows:

$$HPI(t) = \frac{NIR(t)}{NDVI(t)} \tag{1}$$

The normalized HPI (NHPI) is expressed as:

$$NHPI(t) = \frac{HPI(t) - HPI_{min}(t_{start}, t_{end})}{HPI_{max}(t_{start}, t_{end}) - HPI_{min}(t_{start}, t_{end})} \tag{2}$$

where $HPI_{min}(t_{start}, t_{end})$ and $HPI_{max}(t_{start}, t_{end})$ represent the minimum and maximum HPI values observed during the harvesting window [t_{start}, t_{end}], respectively. Within this window, the HPI time series is normalized to a 0–1 scale (Eq. (2)). The harvesting window is defined as the period starting from the timing of middle of senescence (MOS) metric, identified as the point where NDVI decreases to 50 % of its amplitude (the difference between maximum and minimum NDVI values during the senescence phase), to the two months afterward (Fig. 5). This definition is informed by harvesting dates of extensive fields in Beck's field-level data across the U.S. Midwest, which shows that the intervals between MOS and actual harvesting dates consistently fall within this two-month range. Defining this harvesting window eliminates the interference from signals unrelated to the harvesting period to improve the detection accuracy and efficiency.

We employ NIR and NDVI for devising NHPI based on three key considerations. First, the NIR band offers stronger contrast between crop residue and senescent plants compared to other spectral bands (Fig. 4), making it more effective in distinguishing field conditions before and after harvesting. Second, the pre- and post-harvest trends in NIR and NDVI values exhibit distinct and contrasting patterns, enhancing the precision and reliability of harvesting event detection when analyzed through their ratio. Before harvesting, NIR and NDVI follow similar decreasing patterns, keeping NHPI values low and stable during senescence. After harvesting, NIR reflectance increases sharply while NDVI decreases, resulting in a significant rise in NHPI that effectively signals the harvesting event (Fig. 5). The decrease in NDVI occurs as the removal of senescent plants further reduces chlorophyll content and photosynthetic activity, exposing non-vegetative surfaces (i.e., soil or crop residue). Third, NIR is less sensitive to variations in water content than SWIR, reducing the interference from rainfall and making it a more reliable indicator across different moisture conditions. In conclusion, NHPI effectively highlights the harvesting event, displaying relatively low values prior to harvest and significantly elevated values afterward as shown in Fig. 5. Harvesting can be detected when NHPI exceeds a specified threshold, which signals the occurrence of a harvesting event.

3.2. *Workflow of harvesting date mapping using NHPI*

The entire workflow of utilizing the NHPI to map corn and soybean harvesting dates across the U.S. Midwest (i.e., data harmonization, data quality control, index time series generation, and harvesting date detection) is conducted on GEE using its parallel matrix computation capabilities (Fig. 6). Due to the computational limits of GEE to regular

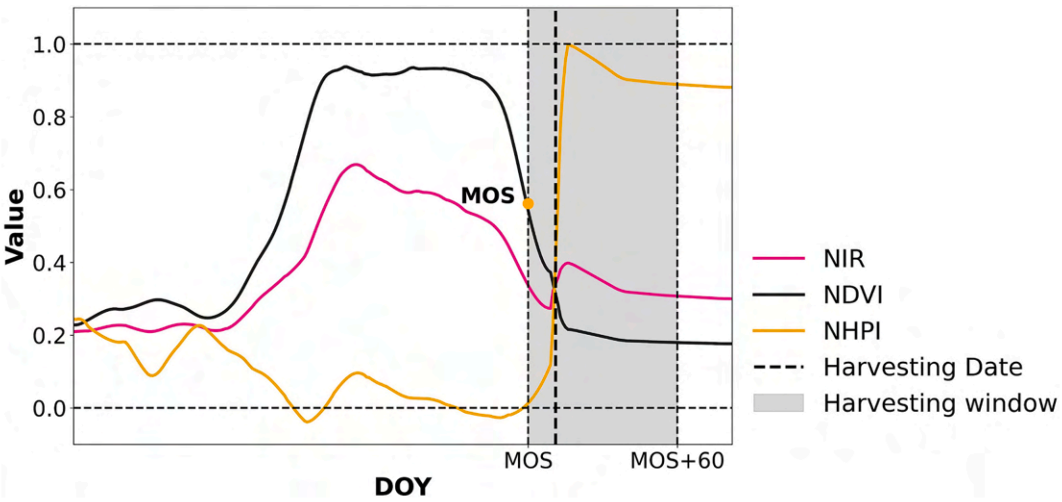


Fig. 5. Schematic diagram of the NHPI time series across the whole harvesting season and the corresponding NIR and NDVI source curves. MOS refers to middle of senescence metric.

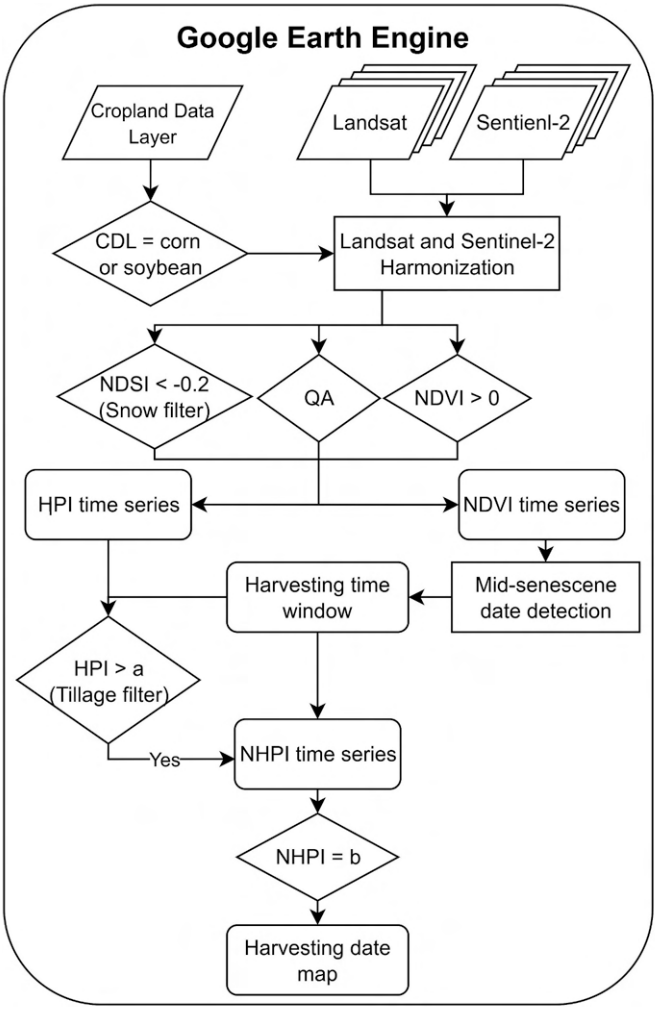


Fig. 6. Workflow for mapping corn and soybean harvesting dates across the U. S. Midwest using the Google Earth Engine platform. *a* is applied to exclude pixels without a clear harvesting signal, often disrupted by tillage practices. *b* is used to locate the estimated harvesting date on the NHPI time series.

users, we process data on a county-by-county basis and later merge the results to create the final Midwest-wide harvesting date product.

Specifically, the first step is data harmonization, which ensures consistency across sensors for robust time-series analysis. Landsat 7, Landsat 9, and Sentinel-2 imagery are harmonized to Landsat 8 standards using least-squares regression parameters derived from established studies (Table 1) (Roy et al., 2016; Zhang et al., 2018; Fan et al., 2023). We utilize CDL data to locate corn and soybean fields. The second step is data quality control, where rigorous filtering is applied to ensure reliable observations. Low-quality pixels, identified using the Quality Assurance (QA) layer for conditions such as cloud cover, shadows, or snow, are masked. Since the QA layer does not always effectively remove snow contamination, we further refine the data by retaining only observations where the Normalized Difference Snow Index (NDSI) is below -0.2, as lower NDSI values indicate a reduced likelihood of snow cover. While NDSI >0.4 is typically used to indicate snow presence, we adopt a conservative threshold of -0.2 to minimize the risk of including any snow-contaminated pixels (Riggs et al., 2016). In addition, to ensure numerical stability in the HPI and NHPI calculations, we retain only observations where NDVI is greater than zero. The third step focuses on the generation of the index time series. With high-quality satellite observations, we generate NDVI and HPI time series to capture phenological development and harvesting transitions. To address temporal gaps between valid observations, we apply linear interpolation to produce continuous NDVI and HPI curves. The NDVI time series is first analyzed to determine the harvesting window using the MOS as a key marker for each field (Fig. 5). Defining this window is crucial for isolating harvesting-related signals, eliminating interference from

Table 1
Coefficients used to harmonize Landsat 7, Landsat 9, and Sentinel-2 data to Landsat 8 across the six spectral bands.

Bands	Landsat-7	Landsat-9	Sentinel-2
Blue	$L8 = 0.0003 + 0.8474 \times L7$	$L8 = 0.0021 + 0.9518 \times L9$	$L8 = 0.0003 + 0.9570 \times S2$
Green	$L8 = 0.0088 + 0.8483 \times L7$	$L8 = 0.0024 + 0.9568 \times L9$	$L8 = 0.0015 + 1.0304 \times S2$
Red	$L8 = 0.0061 + 0.9047 \times L7$	$L8 = 0.0021 + 0.9690 \times L9$	$L8 = 0.0041 + 0.9533 \times S2$
NIR	$L8 = 0.0412 + 0.8462 \times L7$	$L8 = 0.0112 + 0.9545 \times L9$	$L8 = 0.0077 + 0.9644 \times S2$
SWIR1	$L8 = 0.0254 + 0.8937 \times L7$	$L8 = 0.0086 + 0.9560 \times L9$	$L8 = 0.0034 + 0.9522 \times S2$
SWIR2	$L8 = 0.0172 + 0.9071 \times L7$	$L8 = 0.0048 + 0.9596 \times L9$	$L8 = 0.0004 + 0.9711 \times S2$

unrelated periods, and enhancing the accuracy and efficiency of harvesting date detection for each field. Within this window, we then derive NHPI values by normalizing HPI to a 0–1 scale based on local minimum and maximum HPI values (Eq. (2)), ensuring comparability across different fields and environmental conditions. The final step is harvesting date detection from the retrieved HPI and NHPI index time series. Harvesting-related signals can become obscured when tillage follows harvest quickly and satellite observations fail to capture the short interval due to limited data availability. This limitation directly affects the accuracy of harvesting date estimation. To overcome this, we design a thresholding strategy: HPI must exceed a (set to 0.8, calibrated using 2023 Illinois ground-truth data from Beck's dataset) to confirm the presence of key observations of the harvesting events. The harvesting date is then identified as the first date when NHPI exceeds b (set to 0.6). This threshold is determined through calibration using 2023 Illinois ground-truth data from Beck's dataset, where it yields the lowest error for both corn and soybean (Fig. S2). This approach ensures accurate detection of harvesting events by using a tillage filter to minimize the impact of concurrent post-harvest transitions and satellite data gaps.

3.3. Benchmarks

All benchmark methods utilize the harmonized Landsat and Sentinel-2 NDVI time series, processed with the least square parameters (Table 1), as input for harvesting date estimation for each target farm field. This ensures consistency in the data source used by the NHPI-based method and the benchmark methods, enabling a fair comparison of their performance.

3.3.1. EOS phenometric-based method

As a widely used and established method for detecting harvesting dates, the EOS phenometric-based method estimates crop harvest phenological stage by extracting the phenological metrics characteristic of end of season. These EOS metric extraction strategies include threshold-based approaches (e.g., 10 % or 20 % of the NDVI curve amplitude) and inflection point-based approaches (e.g., identifying local minima in the rate of curvature change) (Diao, 2020). The selection of each crop type's corresponding EOS extraction strategy is guided by previous literatures, ensuring that the derived EOS metric closely aligns with the harvesting date's position on the NDVI time series for that crop, as validated by ground truth data.

Specifically, the crop's growth profile is first modeled by Beck's double logistic method to minimize noise while preserving unique growth patterns from the original raw data. Beck's double logistic method fits the NDVI time series using a generalized double logistic function with six parameters to model crop phenological development (Eq. (3)).

$$f(t) = V_{base} + (V_{max} - V_{base})^* \left(\frac{1}{1 + e^{(-m_1{}^*(t-m_2))}} + \frac{1}{1 + e^{(-n_1{}^*(t-n_2))}} - 1 \right) \tag{3}$$

where t is the day of year (DOY), and $f(t)$ is the corresponding fitted NDVI value. V_{base} is the NDVI value during the off-season, and V_{max} is the peak NDVI value for the year. m_2 and n_2 indicate the timing of inflection points during the curve's rising and declining phases, respectively. m_1 and n_1 describe the rates of change at these two inflection points (m_2 and n_2), respectively. The six parameters are estimated by the least square method which minimizes the root mean square errors (RMSEs) between $f(t)$ and the NDVI time series. This logistic-based fitting function provides a smooth and consistent representation of the crop growth cycle.

Then, the EOS metrics are extracted using GU-based strategy and curvature-based strategy from the fitted curve as estimated corn and soybean harvesting dates, respectively, in reference to literature (Diao, 2020). Among all EOS metric extraction strategies, these two strategies have shown superior performance, with the GU-based EOS metric showing the highest alignment with observed corn harvesting dates and the curvature-based EOS metric showing the highest alignment with observed soybean harvesting dates, as evidenced by Tables 4 and 5 of Diao, 2020. The GU-based strategy identifies the EOS metric as the intersection between the senescence line and base line of the fitted NDVI curve, termed "Recession" (Fig. 7). The base line is defined as the horizontal line at the minimum value of the fitted NDVI time series curve. The senescence line is defined as the tangent line to the fitted NDVI time series curve at the curve point with the minimum of the first gradient line. The curvature-based strategy identifies the EOS metric when the change rate of curvature of the fitted NDVI time series reaches its last local minimum in the harvesting window, termed "Dormancy" (Fig. 7).

3.3.2. Shape model fitting (SMF) method

As a classic phenology matching method, the SMF method is designed to estimate crop phenological transition dates, including harvesting dates, by matching the geometrical patterns of the target and reference phenological time series (Sakamoto et al., 2010). The SMF method operates under the assumption that the reference phenological time series can be geometrically matched to the target phenological time series through scaling and shifting. Once these two time series are aligned, the target harvesting date can be estimated by transferring the reference harvesting date of the reference phenological time series to the target one (Fig. 8).

The target phenological time series is the pixel-level NDVI time series of the target field derived from harmonized Landsat and Sentinel-2 dataset. The reference phenological time series represents a characteristic NDVI phenological curve for the specific crop (i.e., corn or soybean). For each crop species, it is typically pre-defined as the 90th percentile of all the pure MODIS pixels' NDVI curves on an annual and state-specific basis in reference to the previous study (Diao et al., 2021). The reference phenological time series defined per state and year can adapt to region and interannually varying crop growing conditions and have been found to outperform conventional phenological references. The 500-m spatial resolution of MODIS can help locate an adequate number of pure pixels of corn and soybean fields across U.S. Midwest while maintaining the computational efficiency of the calibration. The 90th percentile is selected for its ability to reliably characterize crop phenological profiles under optimal growth conditions. For each state and year, we calibrate its reference harvesting date on the reference phenological curve by examining all dates within a two-week window around the median harvesting date from the USDA CPRs. CPRs are generated based on aggregated field-level phenological observations and are used instead of Beck's field-level harvesting phenological records due to their consistent, long-term availability across states and years, which facilitates scalable region-wide calibration on an annual basis. Based on each candidate reference harvesting date, the harvesting dates of all pure corn (or soybean) pixels (i.e., target pixels) are estimated through the geometric matching of the SMF method (Eq. 4). The optimal reference harvesting date is selected as the one that minimizes the RMSE between the temporal cumulative distribution of SMF-estimated harvesting dates of target pixels and the corresponding ground harvesting date distribution from the CPRs. This calibration of reference harvesting date helps reduce potential uncertainties when aligning satellite time series with regional phenological observations from CPRs. With the calibrated reference harvesting date and the pre-defined reference phenological time series, the SMF method is then applied to estimate the harvesting dates for target farm fields using corresponding NDVI time series.

The geometrical matching process of the SMF method is defined in Eq. (4).

$$f(DOY) = Scale_{VI} \cdot r\left(Scale_{DOY} \cdot \left(DOY + Shift_{DOY}\right)\right) \tag{4}$$

where the function $r(DOY)$ represents the NDVI value of the reference phenological time series of the date DOY, and $f(DOY)$ refers to the NDVI

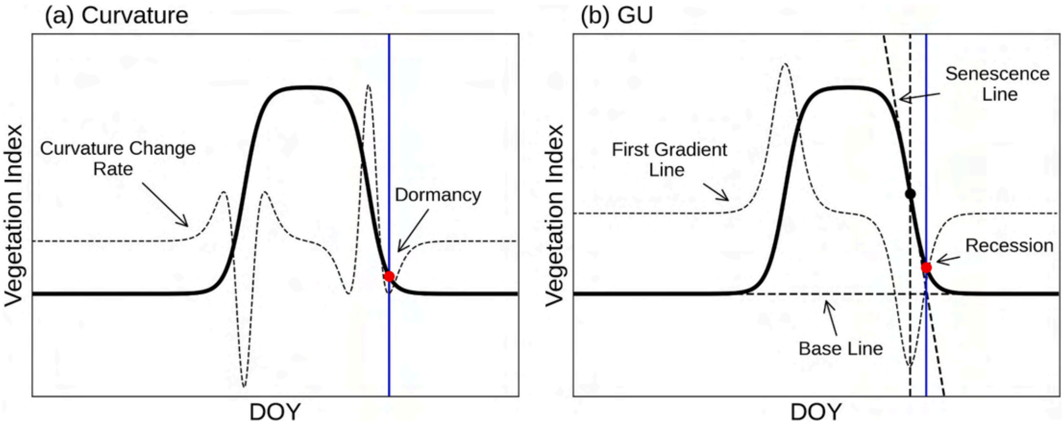


Fig. 7. Schematic of EOS metric extraction using curvature-based strategy (a) and GU-based strategy (b).

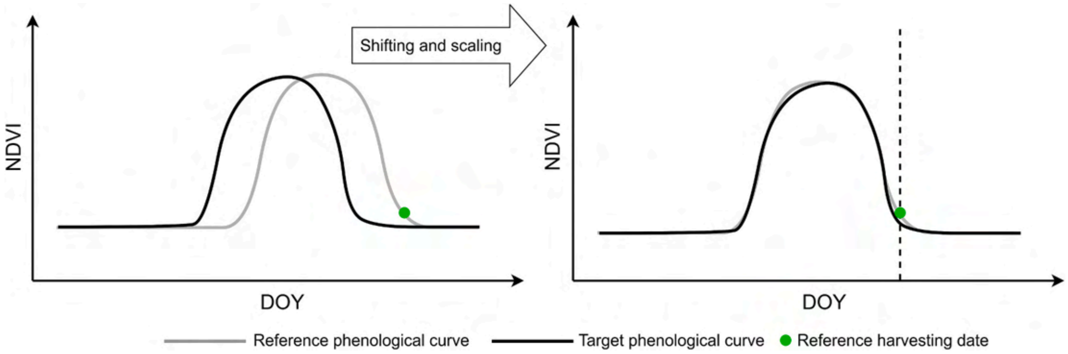


Fig. 8. Schematic of geometric matching between reference and target phenological time series curves using the SMF method.

value of the fitted phenological curve on the date *DOY*. The functions *r (DOY)* and *f(DOY)* denote the reference and aligned phenological trajectories in shape matching. $Scale_{VI}$ and $Scale_{DOY}$ denote the magnitude of scaling of the reference phenological curve on the vertical and horizontal axes (the dimensions of NDVI value and time), respectively. $Shift_{DOY}$ denotes the relative shift of phenological timing of the reference phenological curve.

The optimum scaling and shifting parameters (i.e., $Scale_{VI}$, $Scale_{DOY}$, and $Shift_{DOY}$) are determined by minimizing the RMSE between the fitted phenological curve *f(DOY)* and the target phenological curve *t (DOY)* (Sakamoto et al., 2010).

With the calibrated reference phenological harvesting date and the optimum scaling/shifting parameters specifically fitted for each target phenological curve, the harvesting date on the target phenological curve is estimated through the geometric transformation equation (Eq. (5)).

$$HD_t = \frac{1}{Scale_{DOY}} \cdot HD_r - Shift_{DOY} \tag{5}$$

where HD_t is the estimated harvesting date on the target phenological curve, and HD_r is the calibrated reference harvesting date on the reference phenological curve. $Scale_{DOY}$ and $Shift_{DOY}$ are the optimum horizontal scaling and shifting parameters, respectively.

3.3.3. Shape model fitting by separate phenological stages (SMF-S) method

As another benchmark of phenology matching method, the SMF-S method is developed to improve the accuracy of phenology characterization by addressing two limitations of SMF (Liu et al., 2022a, 2022b). First, the SMF method employs a global matching strategy which aligns the entire growing season of the target NDVI time series through linear shifting and scaling. However, this global matching strategy leads to synchronized length changes among all phenological stages. Such uniformity may not reflect actual variations in phenological stage lengths observed in practice. To overcome this, SMF-S applies stage-adaptive windows that allow independent scaling for each phenological stage, thereby enabling unsynchronized phenological changes tailored to individual phenology stages throughout the growing season. Second, the phenological transition date estimates by SMF may exhibit larger variance in the later stages (e.g., harvesting date) than those in the earlier stages (e.g., planting stage) since the reference phenological time series is stretched from the left endpoint (Eq. (4)). This stretching in SMF causes the scaling factor calibration process to prioritize the alignment of the initial stages, resulting in a dependence of phenological variance on the stage of growth. To address this issue, SMF-S performs shape model fitting using a modified fitting function within stage-adaptive windows. By fitting the curve from both sides, the modified fitting function addresses the dependence of phenological variance on the stage of growth and improving the robustness of late-stage phenological estimates (e.g., harvesting date).

Specifically, when applying SMF-S for harvesting date estimation, SMF-S introduces the term $(1 - Scale_{DOY}) \cdot HD_r$ to allow flexible stretching on both sides of harvesting-adaptive window. HD_r is the calibrated reference harvesting date on the reference phenological curve. The harvesting-adaptive window is centered at $DOY = (HD_r - Shift_{DOY})$ with the half-window width w. The width w is determined by both the local curve characteristics during harvest and the data noise levels. Additionally, SMF-S strengthens the nonlinear parameter optimization process by eliminating the NDVI scaling parameter, $Scale_{VI}$, as phenology estimates are primarily determined by the scaling and shifting in the time dimension (Eq. (6)). The modified function is expressed as:

$$f(DOY, HD_r) = r\left(Scale_{DOY} \cdot \left(DOY + Shift_{DOY}\right) + (1 - Scale_{DOY}) \cdot HD_r\right) \tag{6}$$

The optimum scaling/shifting parameters (i.e., $Scale_{DOY}$ *and* $Shift_{DOY}$) are determined by maximizing the correlation coefficient between the fitted phenological curve $f(DOY, HD_r)$ and the target phenological curve *t(DOY)* within harvesting-adaptive time window (Liu et al., 2022a, 2022b).

With the calibrated reference phenological harvesting date and the optimum scaling/shifting parameters specifically fitted for harvest phenological curve, the harvesting date on the target phenological curve is estimated through the geometric transformation equation (Eq. (7)).

$$HD_t = HD_r - Shift_{DOY}^{opt} \tag{7}$$

where HD_t is the estimated harvesting date on the target phenological curve, and HD_r is the calibrated reference harvesting date on the reference phenological curve. $Shift_{DOY}^{opt}$ is the optimum shifting parameter.

3.4. Evaluation

3.4.1. Index evaluation

To evaluate the suitability of different spectral indices for harvesting date detection, we compare the proposed NHPI with two band ratio indices that share similar structural formulation and with three commonly used residue-based indices, in terms of their ability to distinguish between pre-harvest and post-harvest conditions. Specifically, we include SWIR1/NDVI and SWIR2/NDVI to test whether SWIR bands can serve as alternatives to NIR for capturing harvesting signals. SWIR bands are selected due to their established sensitivity to crop residue, which has potential to detect the spectral transition from senescent vegetation to exposed residue. We also evaluate three residue sensitive indices: the Normalized Difference Tillage Index (NDTI) (Eq. (8)) (Zheng et al., 2012), the Normalized Difference Index 7 (NDI7) (Eq. (9)) (Jin et al., 2015), and the Shortwave Tillage Index (STI) (Eq. (10)) (Stern et al., 2023), which are frequently used for detecting surface residue and tillage intensity. All indices are applied to the Beck's dataset for comparative analysis. To account for the potential influence of rainfall on index performance, we divide the dataset into two groups based on whether precipitation occurs within seven days surrounding the recorded harvest date, using daily precipitation data from DAYMET. This evaluation framework enables a consistent and robust assessment of each index's effectiveness under both dry and wet harvesting conditions.

$$NDTI = \frac{SWIR1 - SWIR2}{SWIR1 + SWIR2} \tag{8}$$

$$NDI7 = \frac{NIR - SWIR2}{NIR + SWIR2} \tag{9}$$

$$STI = \frac{SWIR1}{SWIR2} \tag{10}$$

3.4.2. Field-level evaluation

At the field level, we validate NHPI-estimated harvesting dates using ground-truth data of the Beck's dataset across the U.S. Midwest from 2021 to 2023. We quantify its performance using Mean Absolute Error (MAE) (Eq. (11)) and the coefficient of determination (R^2) (Eq. (12)) between estimated and observed harvesting dates of the corn (or soybean) fields. Additionally, NHPI's performance is compared with that of the three benchmarks for estimating field-level harvesting dates using the Beck's dataset. To ensure a fair comparison, all benchmark methods are evaluated using the same dataset filtered by the tillage filter applied during the NHPI-based harvesting date estimation process (Fig. 6).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{12}$$

where y_i is the observed harvesting date of the sample i, $\bar{y}$ is the mean of all the observed harvesting dates, and $\hat{y}_i$ is the estimated harvesting date of the sample i. n denotes the number of samples. $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ is the sum of squared errors and $\sum_{i=1}^{n} (y_i - \bar{y})^2$ is total sum of squares.

3.4.3. State-level evaluation

At the regional level, we assess the performance of the NHPI-based method in estimating harvesting dates across the U.S. Midwest by comparing its aggregated distributions across states and years to the cumulative harvesting date distributions derived from CPRs for corn and soybean. Harvesting dates for all corn (or soybean) pixels are aggregated to generate state-level cumulative distributions formatted consistently with CPR data for the years 2020 to 2023. To assess the consistency between the NHPI-based estimates and CPR data, we sample points at 5 % intervals along each cumulative distribution curve (from 20 % to 80 %) and evaluate their differences using MAE and R^2 metrics.

Considering the high computational resources required for the advanced benchmark methods, we do not evaluate them at the state level. Instead, we generate crop harvesting date maps for Champaign County for the years 2022 and 2023 using the NHPI-based method and benchmarks. These maps are compared with ground-truth data from field surveys to assess the performance of each method across a continuous spatial scale.

4. Result

4.1. Index evaluation

Fig. 9 presents the differences in spectral index values between pre- and post-harvest periods for six candidate indices, evaluated separately under dry and wet harvesting conditions. Under dry conditions, all three NHPI-wise indices (i.e., NIR/NDVI, SWIR1/NDVI, and SWIR2/NDVI) demonstrate strong separability between pre- and post-harvest stages, with SWIR1/NDVI exhibiting the greatest mean difference. Under wet conditions, the separability of SWIR1/NDVI and SWIR2/NDVI decreases considerably, with the differences becoming indistinct or nearly imperceptible. In contrast, NIR/NDVI maintains relatively clear and

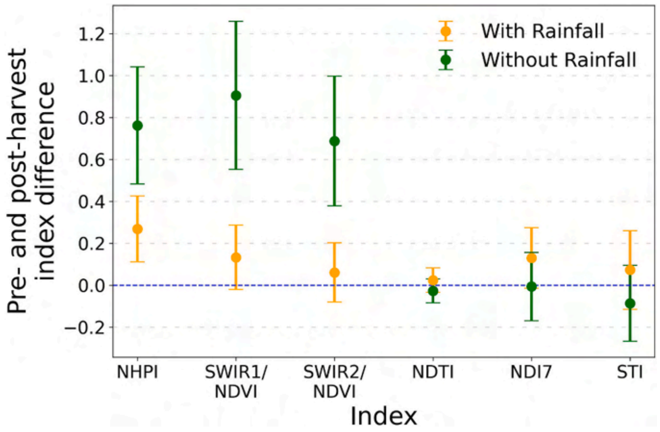


Fig. 9. Differences in indices of NIR/NDVI, SWIR1/NDVI, SWIR2/NDVI, NDTI (Normalized Difference Tillage Index), NDI7 (Normalized Difference Index 7), and STI (Shortwave Tillage Index) between pre- and post-harvest corn and soybean fields in the U.S. Midwest under varying rainfall scenarios. Rainfall presence is determined by whether precipitation occurred within seven days of the recorded harvest date. Error bars represent the standard deviation of the differences.

pronounced separability, while a moderate reduction is still observed. This is because NIR, SWIR1, and SWIR2 are all affected by surface moisture, while NIR is less sensitive to moisture-induced reflectance suppression than SWIR due to its weaker water absorption features. As a result, NIR serves as a more reliable and resilient numerator in ratio-based indices for detecting harvest transitions. The residue-based indices (i.e., NDTI, NDI7, and STI) show minimal differences between pre- and post-harvest periods, regardless of rainfall conditions. This is likely due to their primary sensitivity to surface residue and soil composition, rather than to the phenological transition from senescence to residue, which is critical for capturing the harvesting signal. Overall, NIR/NDVI demonstrates superior and stable performance under both dry and wet conditions, supporting the selection of NIR as the optimal spectral band in the NHPI formulation.

4.2. Field-level evaluation

Fig. 10 shows the scatterplots of the harvesting dates estimated by NHPI-based method against the ground-truth dates from the Beck's dataset for corn and soybean from 2021 to 2023. Overall, the NHPI-based method reliably estimates harvesting dates for both crops, with data points close to the 1:1 line (solid diagonal line) and consistently falling within the ±10-day boundaries (dashed lines). Over the three years, the harvesting date estimation performance of the NHPI-based method has demonstrated consistent year-over-year improvement. For corn, the NHPI-estimated harvesting dates show strong agreement with the ground truth, achieving the R^2 of 0.887 and MAE of 4.29 days in 2021. In 2022, the R^2 remains stable at 0.886, while the MAE drops to 3.675 days. By 2023, the R^2 rises to 0.929, and the MAE decreases further to 3.665 days, reflecting continued refinement in the method's performance. For soybean, the NHPI-based method also demonstrates reliable performance and consistent improvement over time, despite its slightly lower performance compared to that of corn. In 2021, the method yields the R^2 of 0.702 and MAE of 5.071 days. The method's accuracy improves in 2022, with R^2 increasing to 0.722 and MAE dropping to 3.388 days. In 2023, the method achieves the R^2 of 0.841 and the MAE of 3.695 days, showing its enhanced reliability for soybean harvesting date estimation over time.

To assess the improvement in harvesting date estimation performance over time, we analyze the impact of the number of high-quality satellite images within harvesting windows on NHPI-based harvesting date estimates. Fig. 11 shows the boxplot of the absolute errors (i.e., MAE) in the harvesting date estimates from NHPI-based method in relation to the number of high-quality harmonized Landsat and Sentinel-2 images available within the ±10-day window around the harvesting date. In general, a larger number of high-quality observations near the

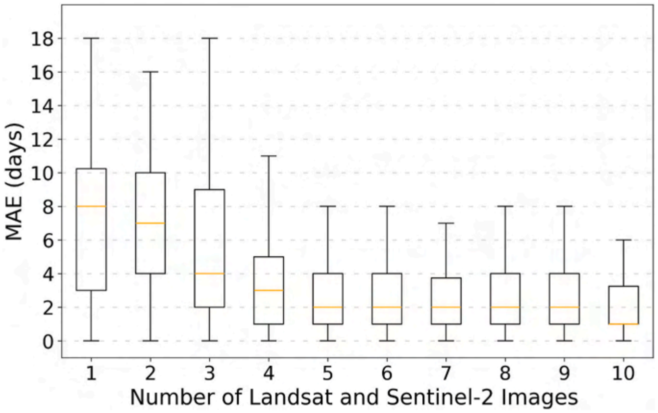


Fig. 11. The number of high-quality Landsat and Sentinel-2 images within the ±10-day window around the harvesting date vs the absolute error (i.e., MAE) of harvesting dates estimated by the NHPI-based method. The yellow lines represent the median of absolute error between actual and estimated harvesting dates. Boxes represent the interquartile range (25th–75th percentiles), and whiskers extend to the most extreme data points within 1.5 times the interquartile range, depending on the actual data distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

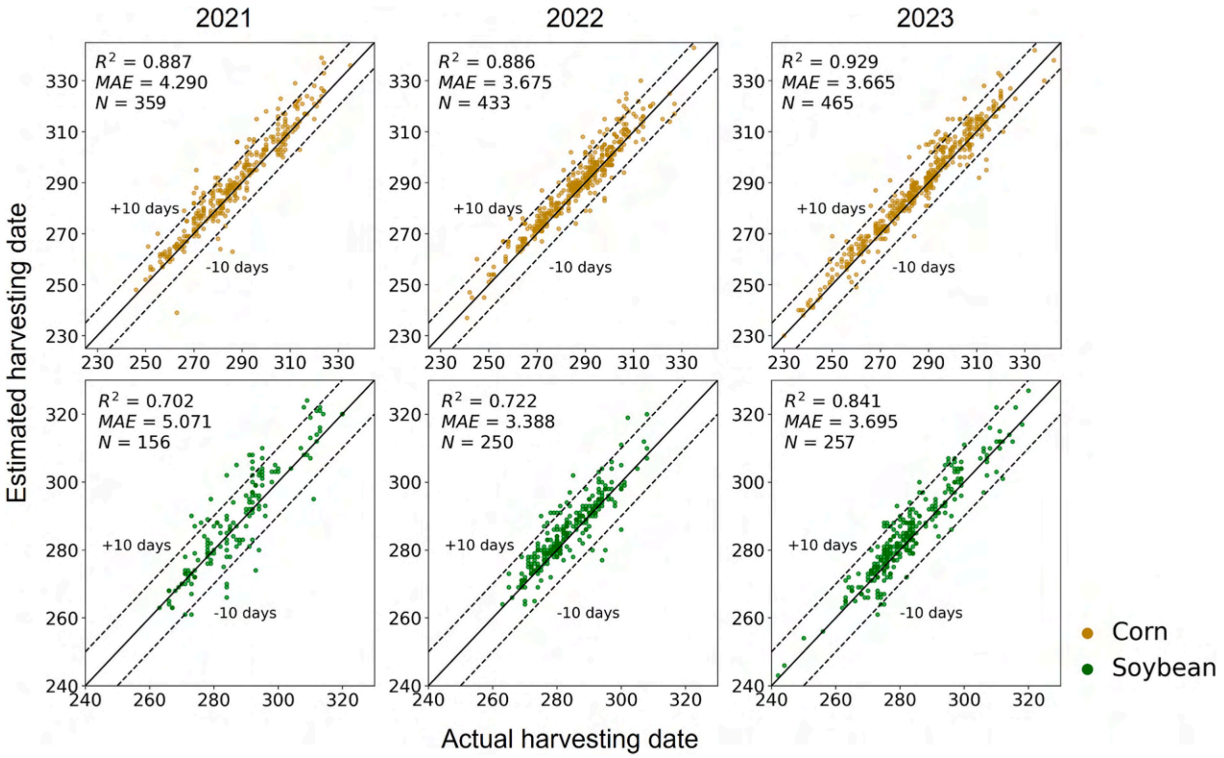


Fig. 10. Annual field-level harvesting date estimation results of NHPI-based method for corn and soybean.

harvesting date usually results in lower absolute errors and more stable performance, as reflected by reduced median errors and narrower error ranges. The median absolute error decreases from 8 days to a stable value of approximately 2 days. This relationship indicates the critical role of enhanced satellite data availability in improving harvesting date estimation accuracy. With more frequent observations, the two satellite images captured before and after harvesting are more likely to be temporally closer to the actual harvesting date. This narrower temporal window restricts the final harvesting date detection to a smaller range, improving its precision. Therefore, the increased availability of observations from different Landsat generations and Sentinel-2 in 2022 and 2023 contributes to the improved accuracy of NHPI-based harvesting date estimates. In contrast, fewer observations are available in 2021 because of the delayed launch of Landsat 9 (Masek et al., 2020). With fewer observations, the harvesting signal in NHPI time series appears weaker (i.e., a gradual slope instead of a sharper transition), resulting in relatively larger estimation errors.

Three state-of-the-art benchmark methods are selected and compared with the NHPI-based method for field-level harvesting date estimation. All methods are validated against Beck's field data. The NHPI-based method demonstrates superior performance in estimating field-level harvesting dates, achieving an average R^2 of 0.83 and a MAE of 3.9 days for corn and soybean fields (Fig. 12). The three benchmark methods show lower performance, with R^2 approximately 0.4 and MAE around 10 days for corn, and R^2 about 0.3 and MAE of 7 days for soybean. Among the benchmark methods, the two phenology-matching methods (i.e., SMF and SMF-S) exhibit comparable performance in harvesting date estimation, providing field-level estimates that generally align with in-situ harvesting data and forming scattered clusters near the 1:1 line. This clustering pattern results from the calibration of the reference harvesting date and phenological curve, which enhances overall matching accuracy. However, the calibrated reference harvesting date and phenological curve on the state and annual basis fail to fully capture the field-level variability in the relationship between harvesting and the VI curve, which results in relatively high estimation errors. The EOS phenometric-based method exhibits relatively larger biases with systematic underestimation for both corn and soybean, as well as distinct horizontal clustering patterns for soybean. The underestimation likely results from the inflection-based feature point extraction (i.e., GU-based and curvature-based strategies) failing to capture late harvest when NDVI has already declined to a plateau and remains unchanged. The horizontal clustering pattern observed in soybean is likely due to the curvature-based strategy, which relies solely on curvature changes and consistently identifies similar EOS points across different curves, especially when variations are minimal (Fig. 7). In contrast, the EOS point extracted by GU-based strategy is determined by both the senescence line and the baseline, making them more sensitive to variations in overall trajectory shape. Overall, the NHPI-based method delivers the most accurate and reliable results, reducing underestimation and enhancing consistency in field-level harvesting date estimation compared to the advanced benchmark methods.

The NHPI-based method outperforms the three benchmark methods in harvesting date estimation, achieving an MAE of approximately 4 days across 2022 and 2023 in Champaign, IL, when validated against field survey observations (Fig. 13). It effectively captures significant local variations in harvesting dates of the area with differences of up to 50 days between the earliest and latest harvesting dates, closely reflecting the real-world spatial variability of harvesting practices (Fig. 13). In contrast, the EOS phenometric-based method shows less stable performance, estimating later harvesting dates in 2022 (MAE: 12.7 days) and earlier dates in 2023 (MAE: 6.3 days). This inconsistent performance highlights interannual variability in the relationship between EOS metrics derived from VI time series and actual harvesting dates. The two phenology-matching benchmark methods, SMF and SMF-S, exhibit relatively stable and similar performance, with MAEs of around 8 days across both years. While the phenology-matching methods generally provide more consistent results, the EOS phenometric-based method may perform better in a given year depending on selected EOS extraction strategy and crop harvesting patterns. In addition, the NHPI-based method appears less affected by drainage compared to the benchmark methods, which show more irregular spatial patterns such as curved features aligned with drainage on the estimated harvesting date maps. This is likely due to two key factors: (1) the tillage filter excludes drainage-affected pixels that lack a distinct post-harvest crop residue signal, and (2) the NHPI trajectory is less sensitive to mixed-pixel effects, where a pixel includes both cropland and drainage. In such cases, the strong spike from post-harvest residue in the cropland portion typically dominates the NHPI signal,

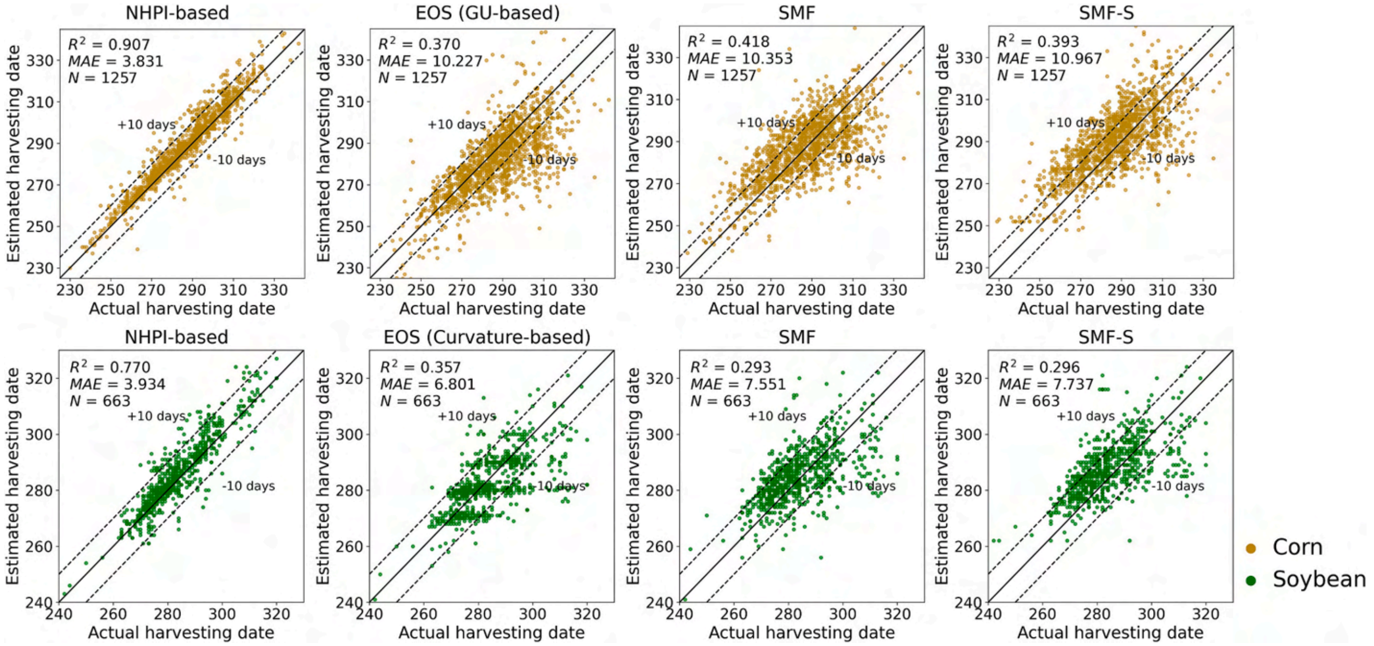


Fig. 12. Field-level harvesting date estimation results from 2021 to 2023 for corn and soybean using the NHPI-based method, EOS phenometric-based method (EOS), SMF, and SMF-S.

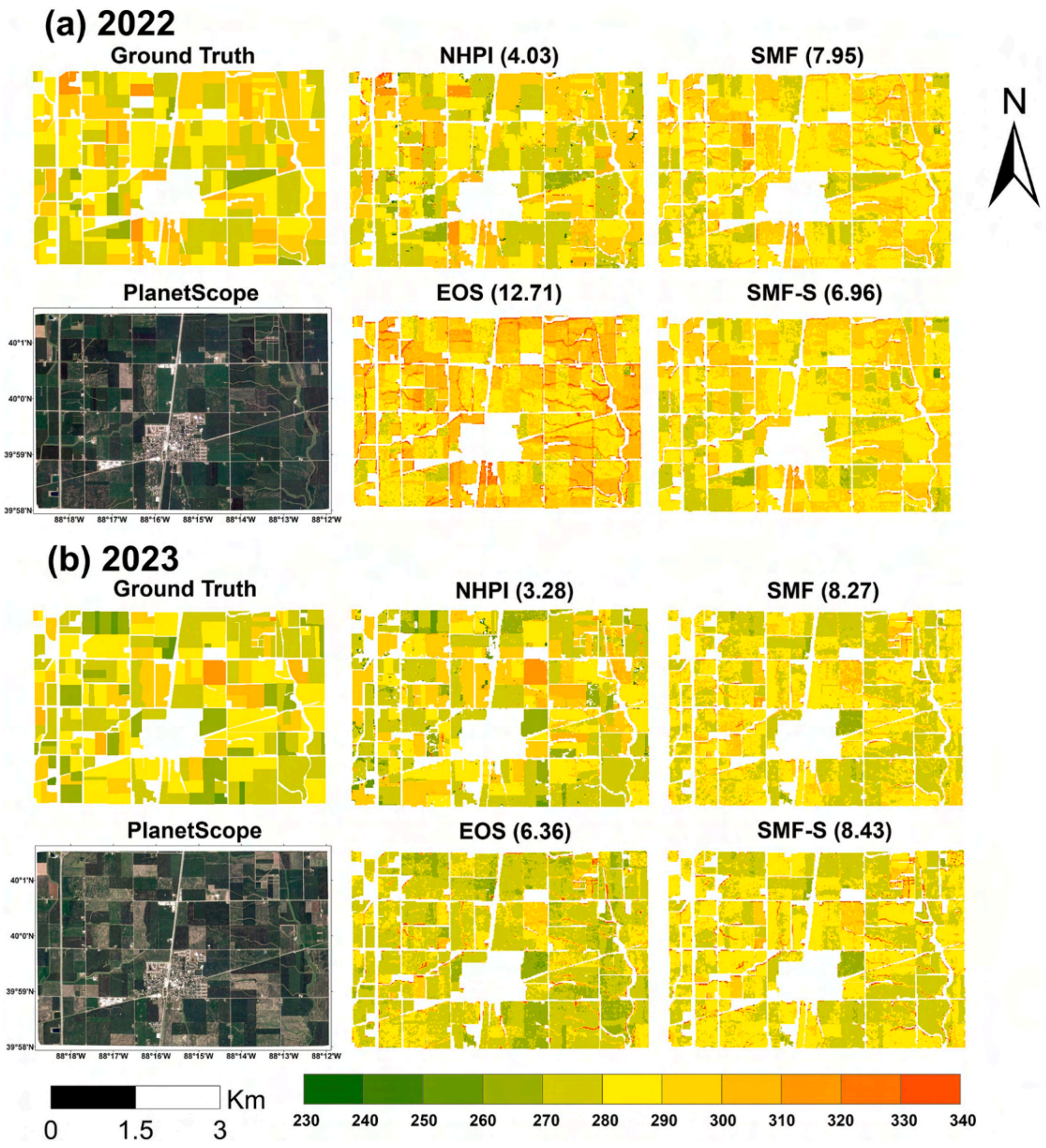


Fig. 13. Field-level harvesting date mapping results for Champaign, IL, in 2022 (a) and 2023 (b), with MAE values (shown in parentheses) for four methods: NHPI-based, SMF, SMF-S, and EOS phenometric-based method (EOS). Ground-truth harvesting dates and the corresponding growing season PlanetScope imagery (sourced from the Planet website) are provided for reference. The location of the evaluated fields within the U.S. is shown in Fig. S1.

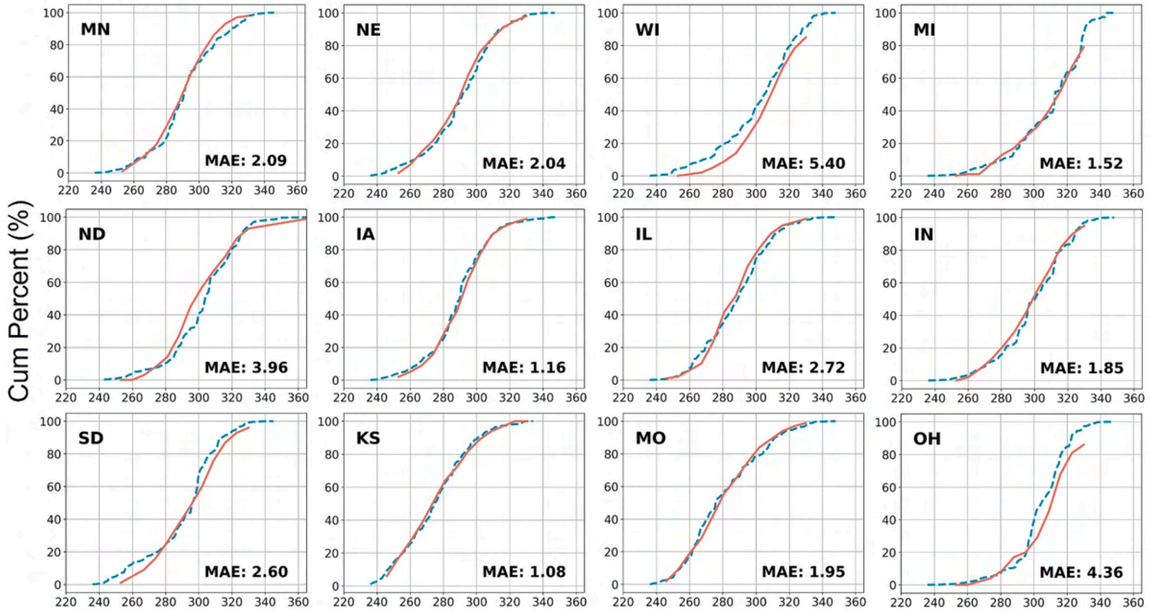
while the drainage portion remains spectrally stable during this transition and has minimal influence on the spike pattern. Overall, the NHPI-based method excels in capturing the real-world spatial variability of harvesting patterns, whereas the benchmark methods tend to produce overly uniform harvesting date estimates that do not adequately represent localized variations across crop fields.

4.3. State-level evaluation

With 2023 as an example, Fig. 14 presents a comparison of the temporal cumulative distributions of NHPI-estimated harvesting dates and state-level CPR data for corn and soybean across 12 states in the U.S. Midwest. In general, the temporal cumulative distributions of estimated harvesting dates align with those of state-level CPRs in 2023, with comparable harvesting progress for both crops across states. MAEs between the actual and estimated harvesting curves are all within 6 days for both crops across all states. 9 states achieve MAEs below 3 days for corn, while 7 states achieve these for soybean. Furthermore, differences in the cumulative distributions of estimated harvesting dates highlight the variability in harvesting decisions across U.S. Midwestern states, driven by diverse environmental, climatic, and farming conditions.

NHPI-estimated harvesting dates for corn and soybean fields in the U.S. Midwest from 2020 to 2023 are aggregated at the state level and validated against corresponding CPR data by sampling points from cumulative distribution percentiles (20 %–80 %) at 5 % intervals. Across all states and years combined, these aggregated results show a high level of consistency with the CPR state-level statistics, achieving the R^2 of 0.94 and the MAE of 2.59 days for corn, and the R^2 of 0.85 with the MAE of 2.98 days for soybean (Fig. 15). Notably, these results demonstrate improved accuracy compared to state-level harvesting date estimates from previous phenology characterization studies that utilized EOS phenometric-based methods (Yang et al., 2020; Shen et al., 2022a), which reported MAEs exceeding one week. This improvement stems from the stronger and more consistent alignment between actual harvesting dates and the feature points identified from the NHPI curve, compared to the relationship between the harvesting dates and the feature points derived from the VIs time series. Per-state evaluations against CPR data further highlight the robustness of the NHPI-based

(a) Corn



(b) Soybean

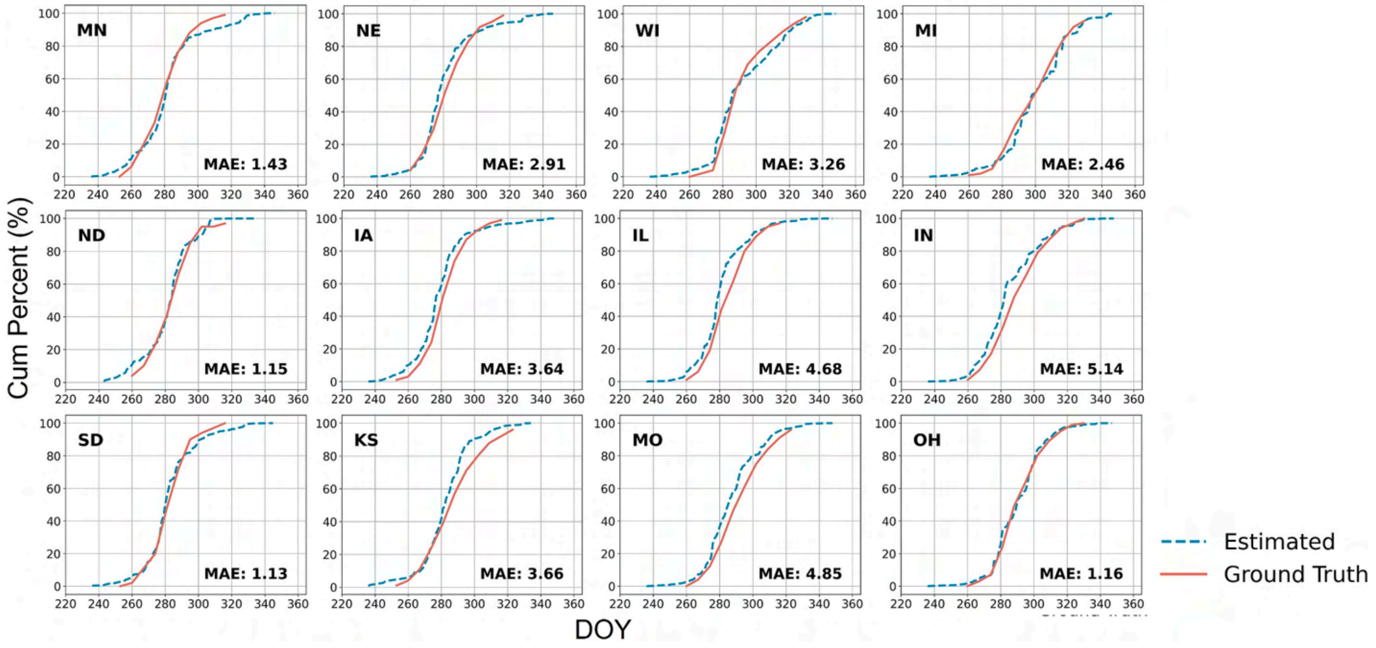


Fig. 14. State-level temporal cumulative distributions of 2023 corn (a) and soybean (b) harvesting dates from the NHPI-based method (dashed blue lines) versus those from ground-based CPRs (solid red lines) across 12 U.S. Midwestern states.

method (Tables S2 and S3). The method achieves MAEs below 5 days and R^2 values above 0.85 for both crops across most Midwestern states. These findings demonstrate the strong performance of the NHPI-based method in capturing state-level harvesting dynamics.

4.4. Spatiotemporal analysis of harvesting dates

The 30-m harvesting date maps for corn and soybean in the U.S. Midwest from 2020 to 2023 reveal distinct spatial and temporal patterns in harvesting practices (Figs. 16 and 17). For corn, harvesting dates tend to be earlier in the south of U.S. Midwest, with a clear spatial trend toward later harvests in the northern states (Fig. 16). In the northern U.S. Midwest, the state with the latest corn harvesting date varied annually from 2020 to 2023, highlighting year-to-year shifts in harvesting patterns likely driven by climate variability, crop management practices, and fluctuations in growing season length. Compared to corn harvesting dates, soybean harvesting dates display more stability across both space and time (Fig. 17). These high-resolution maps provide insights into fine-scale harvesting pattern differences, capturing substantial within-state variations. The zoomed-in consistent view of an area in Iowa across four years highlights year-to-year variation in harvesting dates. Additionally, northern Illinois consistently experiences later corn harvests compared to the southern part of the state. Such spatial details are not captured by state-level CPRs, underscoring the value of these detailed maps in understanding both local and regional harvesting dynamics.

To characterize spatial variability in corn and soybean harvesting dates across the U.S. Midwest, we further analyze the relationship

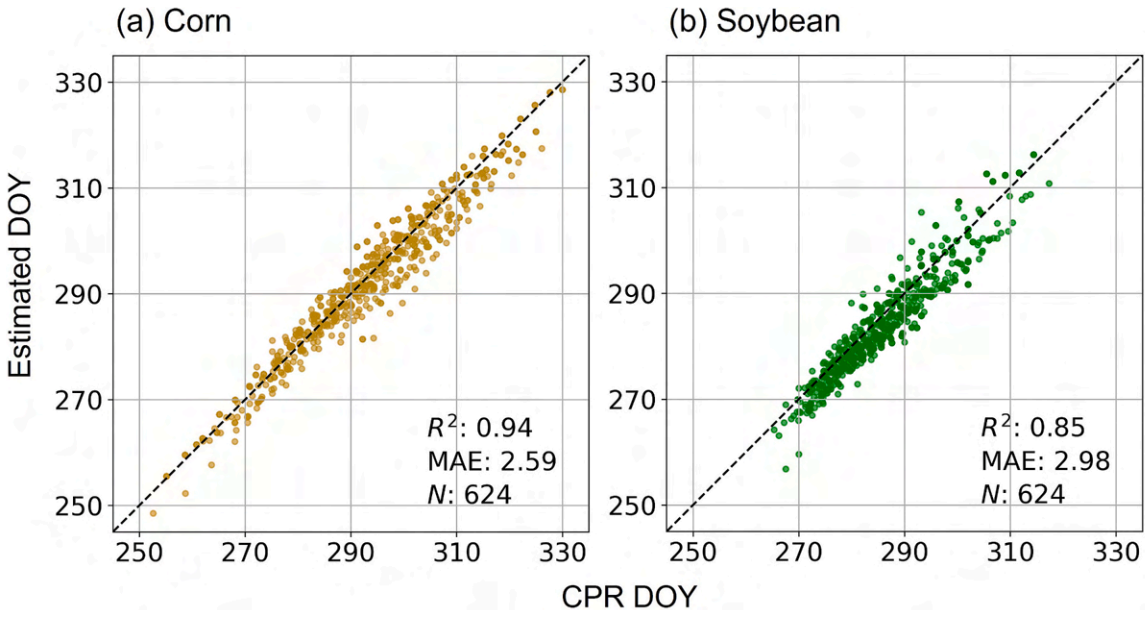


Fig. 15. Validation of NHPI-estimated harvesting dates against CPR's harvesting phenology data, using sampling points from cumulative distribution percentiles (20 %–80 %) at 5 % intervals, for both corn and soybean of the U.S. Midwest from 2020 to 2023. This results in a total of 624 sampling points (12 states × 4 years × 13 percentiles).

between harvesting dates and geographic location. Harvesting dates are aggregated for each county, and the county-level harvesting dates are then associated with the latitudes of the corresponding county centroids, given that latitude is a key determinant of temperature, which in turn influences agricultural management decisions. The fitting method with the lowest RMSE among linear, bilinear, and quadratic models is applied to model the spatial trend of harvesting dates across latitudes. As shown in Fig. 18, corn shows an increasing trend in harvesting dates in lower-latitude regions of the U.S. Midwest but exhibits a slight decreasing trend in higher-latitude ones. For soybean, a trend toward slightly earlier harvesting dates is observed as latitude increases. These patterns are also observed from state-level CPR results.

The earlier corn harvesting dates observed in lower-latitude regions are likely attributed to increased planting flexibility with relatively warm climate conditions. Warmer spring temperatures in these areas accelerate soil warming and promote earlier attainment of the temperatures needed for seed germination and crop growth. This flexibility allows farmers to start planting earlier relative to the northern part (Zheng et al., 2016). At the same time, corn seed sales data show that farmers in these regions often select cultivars with similar relative maturity dates, which may not fully take advantage of the extended growing season (Abendroth et al., 2021). As a result, corn in lower-latitude regions follows a temperature-driven planting pattern, where higher temperatures enable earlier planting and, consequently, earlier harvesting. In contrast, at higher latitudes, cooler temperatures delay planting decisions and slow crop development, prolonging the time required for corn to reach full physiological maturity. This temperature-driven delay, combined with a shorter growing season (frost-free period) constrained by climatic factors, prompts farmers to choose cultivars adapted to local conditions, leading to more synchronized harvest timing across northern regions. For soybean, which has a shorter growing season and is typically planted after corn, harvest timing remains more consistent across latitudes. Its growth cycle fits within a relatively narrow seasonal window, making it less sensitive to regional temperature variations.

The field-level harvesting date maps in the U.S. Midwest reveal that (1) soybean fields generally have later harvest dates than corn fields in regions below 40 degrees latitude, but align more closely with corn fields in regions above 40 degrees latitude for the mapping years, (2) corn harvesting dates vary by more than three months across the Midwest, while soybean harvesting dates vary by up to two months, and (3) the NHPI-based method demonstrates both its feasibility and effectiveness for field-level harvesting date estimation on the GEE platform.

5. Discussion

5.1. Advantages of NHPI

In this study, we introduce a novel NHPI to detect crop harvesting dates from optical satellite time series and demonstrate its effectiveness for corn and soybean fields across the U.S. Midwest, a region with diverse climates and cropping systems. NHPI achieves high accuracy and efficiency at both field and state levels, primarily due to three key advantages.

Firstly, the NHPI-based method innovatively estimates harvesting dates through directly identifying the special spectral characteristics associated with the harvesting event. By capturing the unique spectral reflectance changes through the combination of NIR and NDVI that occur during harvesting, the NHPI time series identifies the transition from senescent plants to crop residue, enabling accurate and reliable harvesting date estimation. In contrast, existing methods, such as phenometric-based detection and phenology matching methods, determine the harvesting date by identifying a related feature point on the VI time series curve (Sacks and Kucharik, 2011; Diao et al., 2021; Dong et al., 2019; Liu et al., 2022a; Shen et al., 2023; Cao et al., 2024; Chen et al., 2004b; Delbart et al., 2006; Zeng et al., 2020; Diao, 2020; Gao et al., 2017; Moulin et al., 1997; Schwartz et al., 2002; Wu et al., 2017). However, these methods assume a consistent relationship between harvesting and the VI curve, an assumption that does not always hold. Variations in field-level management practices can lead to differences in both the shape of VI curves and the timing of harvesting relative to these curves. The same harvesting date may correspond to different positions of VI curves—some near turning points and others approaching the curve minimum (Fig. 19). These discrepancies arise because crops are harvested at different phases of senescence across fields, with some harvested while vegetation remains partially green, resulting in relatively higher VI values, and others after complete senescence, when VI has much declined. Additional factors such as crop variety, weather

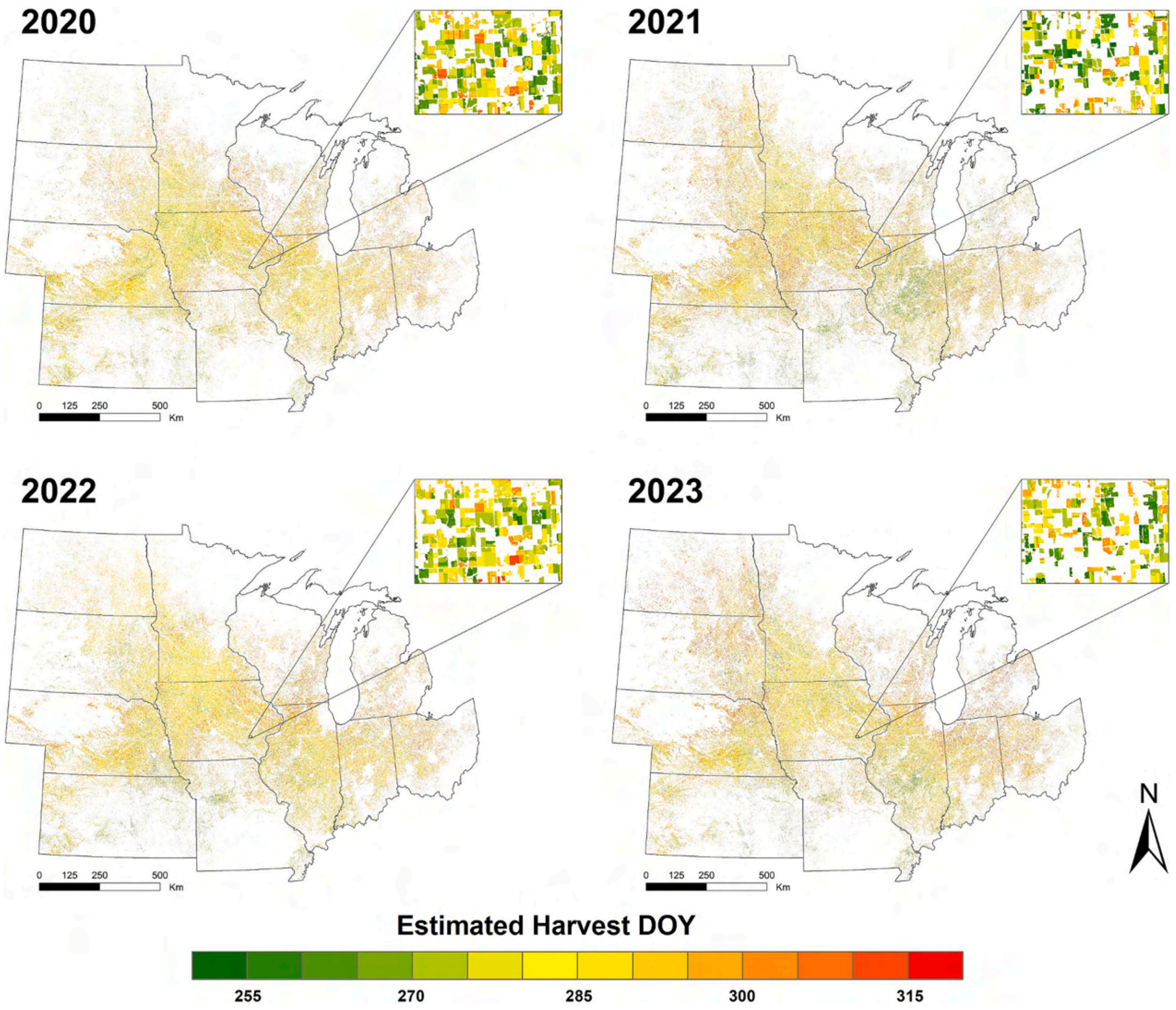


Fig. 16. NHPI-based mapping of 2020–2023 corn harvesting dates across the U.S. Midwest on GEE. The colour scale represents the estimated DOY of harvest, ranging from dark green (earlier than DOY 255) to dark red (later than DOY 315), with a gradient transitioning through yellow and orange in 5-day intervals. Insets display zoomed-in consistent views of a region in Iowa to show differences in estimated harvesting dates of corn fields at the local scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conditions, and management decisions further influence harvesting timing, causing variability in VI patterns and complicating the estimation process. Compared to these methods, the NHPI-based method demonstrates improved accuracy. At both the field and state levels, the NHPI-based method yields lower MAEs than those reported in previous phenology characterization studies, including state-level estimates using EOS phenometric-based (Yang et al., 2020; Shen et al., 2022a) and phenology-matching (Sakamoto, 2018b) methods, as well as field-level estimates based on EOS phenometrics (Shen et al., 2023), all of which report MAEs exceeding one week. This improvement is attributed to the stronger and more consistent alignment between actual harvesting dates and the transition points identified from the NHPI curve, as opposed to the less stable correspondence observed with VI-based metrics. By establishing a distinct linkage between the harvesting phenological stage and spectral characteristics, the NHPI-based method overcomes those limitations of conventional approaches and unlocks new possibilities for advancing crop phenology monitoring.

Secondly, the NHPI-based method effectively enhances the harvesting event signal on the time series, improving the harvesting date estimation accuracy. Among all optical spectral bands and their combinations available in Landsat and Sentinel-2, the NIR/NDVI ratio consistently exhibits the largest separability between pre- and post-harvest stages, regardless of rainfall conditions (Fig. 9). The high level of separability enables precise differentiation between pre- and post-harvest stages directly from the NIR/NDVI ratio time series. The effectiveness of NIR in this context can be attributed to its sensitivity to vegetation structure and its capacity to reduce soil background effects, which sets it apart from other spectral bands (Badgley et al., 2017). Specifically, in healthy plants, internal leaf structures, composed of cell walls and air spaces, efficiently scatter NIR light, resulting in high reflectance. As plants enter senescence stage, these structures deteriorate, with cell walls losing integrity and air spaces diminishing, leading to increased absorption and reduced NIR reflectance (Daughtry and Hunt, 2008; Baldocchi et al., 2020). After harvesting, when senescent vegetation is replaced by crop residue and soil background, NIR reflectance rises sharply due to the loss of organized plant structures, the dominance of reflective fibers (e.g., lignin and cellulose), and the enhanced scattering caused by the rough surfaces of crop residue. Meanwhile, NDVI value continues to decrease from senescent plants to post-harvesting conditions as the removal of senescent plants further reduces chlorophyll content and photosynthetic activity, exposing non-vegetative surfaces (i.e., soil or crop residue). Consequently, NIR generally follows a trend similar to NDVI before harvest but exhibits an opposite trend after harvest. The resulting distinct contrast makes the

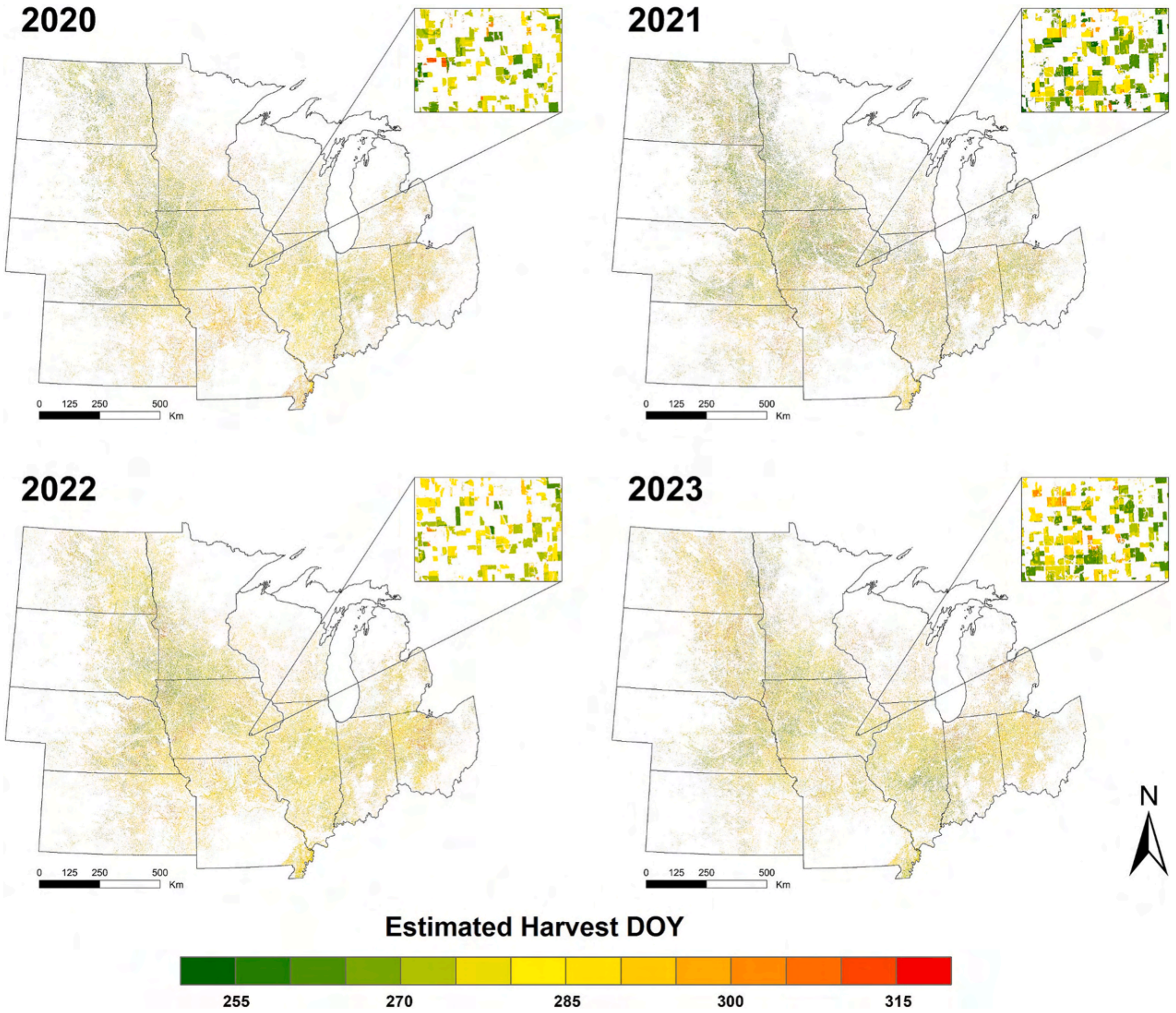


Fig. 17. NHPI-based mapping of 2020–2023 soybean harvesting dates across the U.S. Midwest on GEE. The colour scale represents the estimated DOY of harvest, ranging from dark green (earlier than DOY 255) to dark red (later than DOY 315), with a gradient transitioning through yellow and orange in 5-day intervals. Insets display zoomed-in consistent views of a region in Iowa to show differences in estimated harvesting dates of soybean fields at the local scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

NIR-to-NDVI ratio a clear indicator of the harvesting event. Moreover, as an index designed to detect harvesting dynamics by monitoring the transition from the senescent stage to the crop residue stage, NHPI offers significant advantages over traditional crop residue indices (e.g., NDTI and STI) in harvesting date detection. These crop residue indices are primarily intended for residue ratio analysis, operating under the assumption that harvesting has already occurred and no live crops remain in the field (Quemada et al., 2018). This reliance on post-harvest conditions limits their applicability for harvesting date estimation, as they are unable to reliably distinguish whether crops are still present or have already been harvested. In contrast, NHPI's design emphasizes the capturing of spectral changes during the harvesting event, offering a robust and transferable index for identifying harvesting dates across a wide range of environments.

Lastly, the straightforward design of the NHPI is well-suited for efficient implementation on geospatial data processing platforms like Google Earth Engine. It allows for direct application on Landsat and Sentinel-2 imagery time series, enabling efficient and robust harvesting date mapping without the need for complex preprocessing steps or specialized datasets, which are often required by advanced phenology matching or other harvest phenology detection methods. By leveraging GEE's vast data repositories and powerful computational capabilities, the NHPI-based method supports large-scale agricultural harvest monitoring, facilitating the efficient mapping of harvesting dates across extensive regions.

5.2. *Transferability of the proposed method*

The proposed NHPI-based method exhibits robust performance in estimating harvesting dates at both field and state levels across the U.S. Midwest over multiple years. Notably, this performance is achieved without the need for region-specific threshold calibration, underscoring the strong transferability of NHPI across diverse regions and years. A sensitivity analysis evaluating a range of NHPI thresholds (Fig. S2) indicates that values near 0.6 consistently produce low errors (MAE) and high agreement (R^2) for both corn and soybean across different states and years, further supporting its applicability without extensive recalibration. This transferability is attributed to two key design elements of NHPI, including: (1) the normalization of the HPI time series, and (2) the incorporation of the harvesting window.

Firstly, the normalization of the HPI time series effectively mitigates site-specific variability, reducing estimation errors in harvest date detection. A comprehensive analysis of all U.S. Midwest Beck's datasets reveals that HPI values exhibit a distinct pattern, with post-harvest

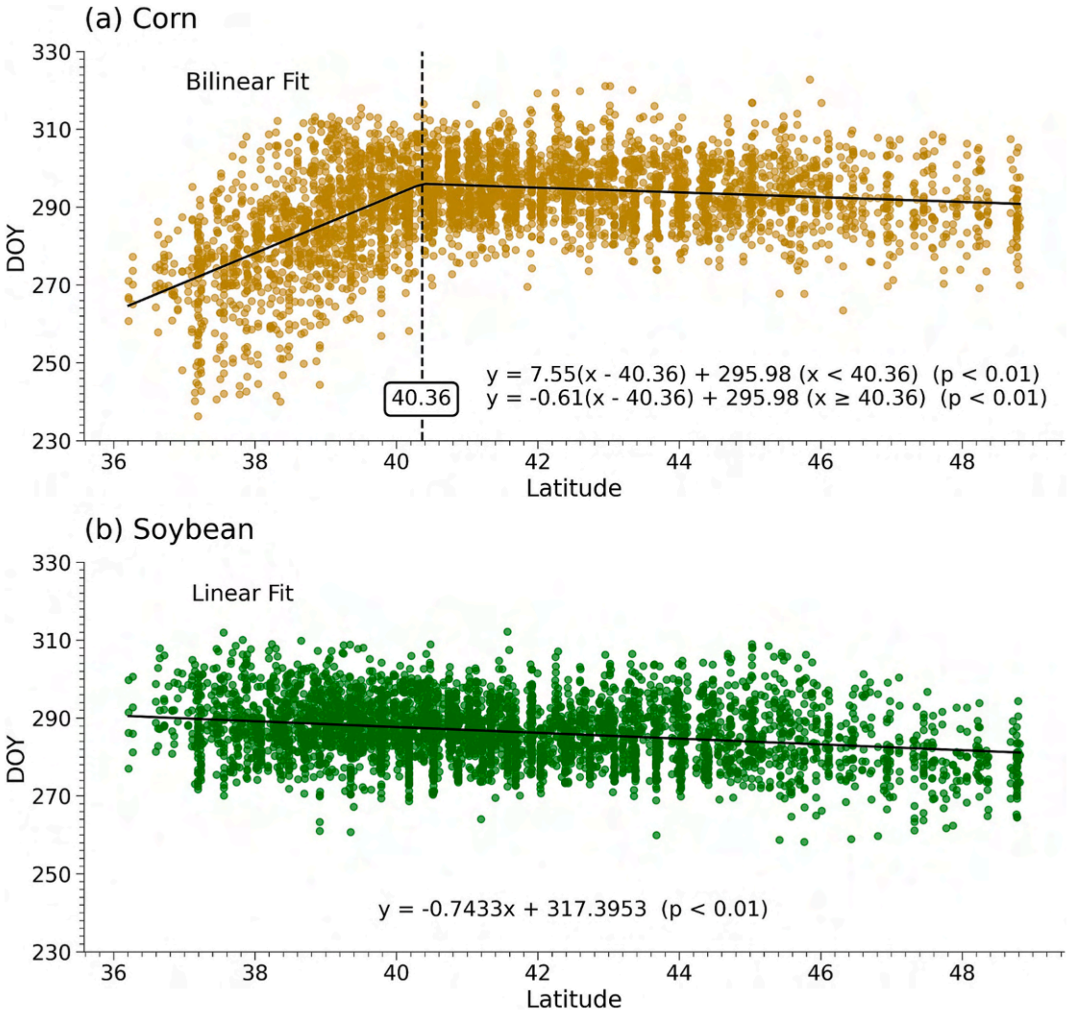


Fig. 18. Relationship between mean county-level harvesting dates and latitude in the U.S. Midwest for corn (a) and soybean (b) from 2020 to 2023, with a fitted trend line illustrating the latitude-dependent variation in harvesting timing. All fitted trends are statistically significant ($p < 0.01$).

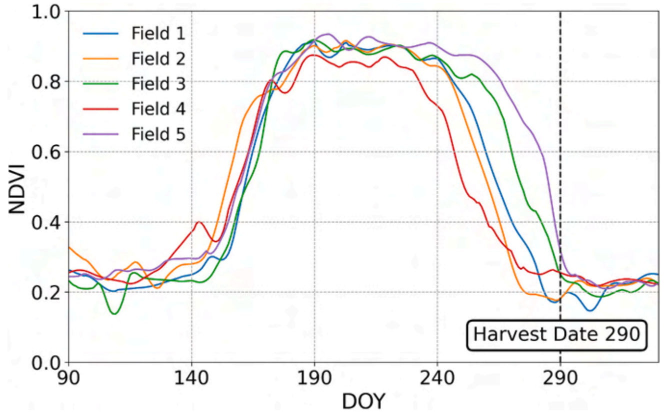


Fig. 19. NDVI profiles for sampled actual corn fields with the same harvesting dates on DOY 290 (October 17) in 2023, based on Beck's dataset of Illinois. The NDVI time series are derived from harmonized Landsat and Sentinel-2 observations.

values generally higher than pre-harvest values (Fig. 20(a)). This pattern highlights the spectral changes captured by the HPI time series before and after harvesting, and underscores the potential of using the HPI threshold for harvesting date detection. However, the overlapping between pre- and post-harvest value distributions can introduce errors when applying a uniform HPI threshold for harvesting date estimation. This overlap is likely caused by variations in crop cultivars, soil characteristics, and management practices, which influence the spectral reflectance patterns of fields during the transition from pre- to post-harvest stages. By standardizing HPI values to the 0–1 range, normalization significantly reduces the overlap between pre- and post-harvest value distributions (Fig. 20(b)). This process creates a clearer separation between pre- and post-harvest stages in the NHPI time series. As a result, NHPI maintains consistent and reliable performance across diverse agricultural regions. Additionally, normalization reduces the need for local calibration, increasing NHPI's adaptability and transferability across different landscapes and conditions. This improvement ensures more reliable and accurate harvest detection on a broader scale, making the method practical for region-wide harvest application.

Secondly, the incorporation of field-specific harvesting window in NHPI-based method ensures that the detection window aligns closely with the actual harvesting period, thereby filtering out confounding trends from earlier or unrelated portions of the season. High NIR/NDVI values typically occur after harvest, while similar values can also appear during the pre-planting season in fields where residual crop matter from the previous growing season remains due to the absence of tillage. The similarity in NHPI values between planting and harvesting periods can complicate harvesting date identification, potentially causing planting to be misclassified as harvesting, particularly when the detection window covers the entire growing season or year. Harvest timing may vary significantly across regions and years. In some cases, harvesting in one

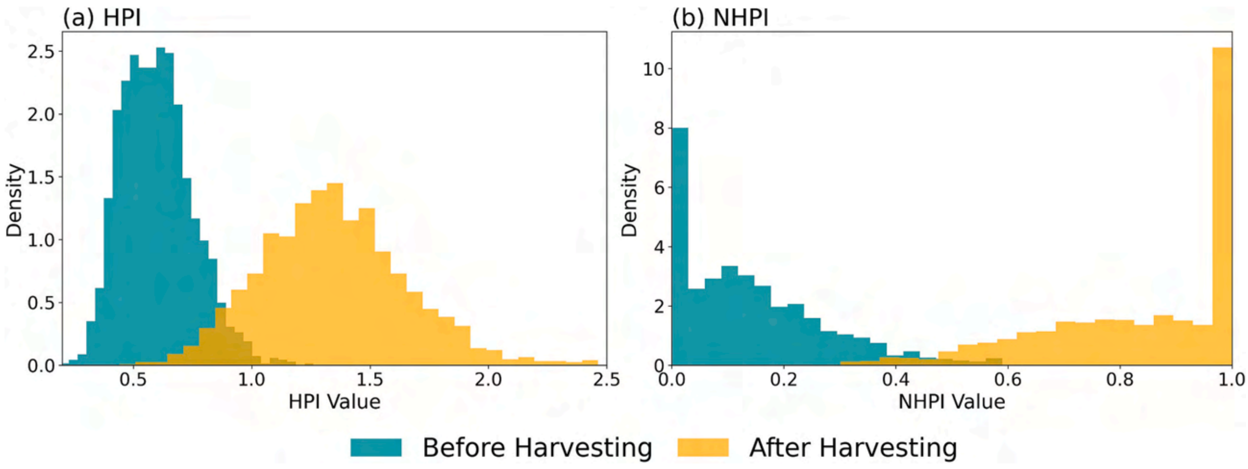


Fig. 20. HPI (a) and NHPI (b) values before and after the harvesting event, derived from harmonized Landsat and Sentinel-2 images for 2023 corn and soybean Beck fields across the U.S. Midwest.

area may coincide with planting in another area due to differences in crop types, management practices, and environmental conditions. By tailoring the harvesting window to each field, this method accounts for such seasonal and regional variations, ensuring more accurate detection. This adaptive approach enhances the overall generalizability and transferability of NHPI, enabling it to perform reliably across diverse regions and agricultural systems under varying conditions.

Overall, the NHPI-based harvest detection method exhibits strong transferability across broad regions and multiple years due to the normalization of the HPI time series and the incorporation of the harvesting window. In regions such as South America (e.g., Brazil) or subtropical Asia, where crops may remain in the field for 3 to 4 months after physiological maturity due to delayed dry-down, limited machinery, or market-driven harvest delays, the harvesting window length can be flexibly extended (e.g., from two to five months) to align with local crop calendars and management practices. Experimental results also indicate NHPI's adaptability to other data sources, such as PlanetScope and MODIS, enabling the creation of tailored harvest date maps for diverse applications.

5.3. Applicable conditions and limitations

The NHPI-based method shows strong performance in both field- and state-level harvesting date estimation. However, this method requires at least one high-quality satellite observation between the harvest event and typical post-harvest field practices (i.e., tillage). In certain scenarios, rapid post-harvest tillage combined with limited satellite data availability (due to cloud cover or infrequent revisit intervals) can obscure the harvesting signal in the HPI time series. This happens because the reflectance of senescent crops is similar to that of the field post-harvest tillage (Fig. 4), causing the HPI time series to appear flattened, with a direct transition from pre-harvest to post-harvest tillage. Consequently, when HPI is normalized to a 0–1 scale (NHPI), accurately distinguishing the harvesting event within the time series may not be easy, even though the harvesting date can still be identified using a threshold derived from the NHPI time series. It is worth noting that prompt no-till post-harvest planting of the next crop does not necessarily affect the detection of the previous crop's harvest date, as crop residues still remain undisturbed on the field surface. However, if the subsequent planting involves conventional tillage, the residue may be removed too quickly, likely preventing the harvesting signal from being captured by satellite observations. To ensure the accuracy of estimated harvesting dates, we utilize the absolute HPI value (rather than the normalized version) to filter out pixels that may lack a detectable harvesting signal due to immediate tillage and insufficient observations. This makes the availability of valid pixels with estimated harvest dates dependent on satellite image accessibility. In addition, the final estimation accuracy is highly related to the availability of the satellite data around harvest period (Fig. 11). To address the data gap-related challenge, advanced image fusion techniques, such as deep learning-based spatial and temporal fusion (Yang et al., 2021; Shen et al., 2022b; Liu et al., 2024; Lyu et al., 2025), could be employed to generate daily observations of both high spatial and temporal resolutions in the future. Additionally, utilizing high-temporal-resolution PlanetScope imagery for harvesting date estimation could also potentially mitigate the challenges posed by long revisit intervals and the close timing of post-harvest tillage, ultimately delivering more accurate and reliable mapping results.

The NHPI-based method estimates the harvesting date by detecting the transition from senescent vegetation to crop residue. Experimental results have demonstrated the robust harvesting date estimation performance of the NHPI-based method for crops such as corn and soybean. Preliminary testing on winter wheat fields also suggests its potential applicability to cereal species, though the absence of ground truth data prevents a full evaluation of its accuracy. However, extending NHPI to certain crops could pose potential challenges, particularly for those not fully removed during harvesting. For instance, in cotton fields, only the cotton bolls are harvested, leaving the stems intact. Such partial harvesting results in weaker spectral changes before and after harvest, which may limit NHPI's effectiveness in such scenarios. Future research could investigate the unique spectral signatures of partially harvested fields, enabling further adaptations of NHPI to improve its applicability across diverse crop types and agricultural systems.

6. Conclusion

In this study, we develop an innovative NHPI index that combines NIR and NDVI to robustly detect harvesting dates for corn and soybean fields across varied soil and weather conditions. Implemented on Google Earth Engine, NHPI supports scalable harvesting date mapping over extended areas using Landsat and Sentinel-2 imagery. The NHPI-based harvesting estimation method is evaluated against three benchmark methods (i.e., the EOS phenometric-based method, the shape model fitting method, and the shape model fitting by the separate phenological stage method). By uniquely identifying the distinct spectral reflectance change associated with the harvesting event, NHPI demonstrates enhanced capability in accurately estimating harvesting dates without requiring site-specific recalibration. In the U.S. Midwest, the NHPI-based method achieves a R^2 of 0.85 and MAE of less than 4 days for field-level harvesting dates from 2021 to 2023 and the state-level R^2 of 0.9 with MAE around 3 days from 2020 to 2023. NHPI demonstrates strong transferability for harvesting date estimation across various regions and years. The estimated field-level harvesting dates can further provide valuable insights for more accurate crop yield estimation and a deeper understanding of factors underlying yield gaps, while the spatiotemporal patterns of harvesting dates can contribute to developing proactive adaptation strategies to mitigate the adverse impacts of climate change.

CRediT authorship contribution statement

Yin Liu: Validation, Methodology, Data curation, Conceptualization, Writing – review & editing, Writing – original draft. **Chunyuan Diao:** Supervision, Resources, Methodology, Funding acquisition, Conceptualization, Writing – review & editing. **Zijun Yang:** Validation, Methodology, Writing – review & editing. **Weiye Mei:** Data curation, Writing – review & editing. **Tianci Guo:** Data curation, Writing – review & editing.

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 study is supported partly by the National Science Foundation (2048068 and 2518299) and partly by the National Aeronautics and Space Administration (80NSSC21K0946). It is also supported by the Taylor Geospatial Institute GISCoR (Geospatial Institute Seed Grant Program to stimulate Collaborative Research) (113356).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2025.115016.

Data availability

Data will be made available on request.

References

Abendroth, L.J., Miguez, F.E., Castellano, M.J., Carter, P.R., Messina, C.D., Dixon, P.M., Hatfield, J.L., 2021. Lengthening of maize maturity time is not a widespread climate change adaptation strategy in the US Midwest. Glob. Chang. Biol. 27, 2426–2440. https://doi.org/10.1111/gcb.15565.

Amherdt, S., Di Leo, N.C., Balbarani, S., Pereira, A., Cornero, C., Pacino, M.C., 2021. Exploiting Sentinel-1 data time-series for crop classification and harvest date detection. Int. J. Remote Sens. 42, 7313–7331. https://doi.org/10.1080/01431161.2021.1957176.

Badgley, G., Field, C.B., Berry, J.A., 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. Sci. Adv. 3, e1602244. https://doi.org/10.1126/sciadv.1602244.

Baldocchi, D.D., Ryu, Y., Dechant, B., Eichelmann, E., Hemes, K., Ma, S., Sanchez, C.R., Shortt, R., Szutu, D., Valach, A., Verfaillie, J., Badgley, G., Zeng, Y., Berry, J.A., 2020. Outgoing near-infrared radiation from vegetation scales with canopy photosynthesis across a Spectrum of function, structure, physiological capacity, and weather. J. Geophys. Res. Biogeosci. 125, e2019JG005534. https://doi.org/10.1029/2019JG005534.

Beddington, J., 2010. Food security: contributions from science to a new and greener revolution. Philos. Trans. R. Soc. B 365, 61–71. https://doi.org/10.1098/rstb.2009.0201.

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. https://doi.org/10.1080/10106049.2011.562309.

Cao, R., Li, L., Liu, L., Liang, H., Zhu, X., Shen, M., Zhou, J., Li, Y., Chen, J., 2024. A spatiotemporal shape model fitting method for within-season crop phenology detection. ISPRS J. Photogramm. Remote Sens. 217, 179–198. https://doi.org/10.1016/j.isprsjprs.2024.08.009.

Chandra, V.S., A.H, Albaaji, G.F., 2024. Precision farming for sustainability: an agricultural intelligence model. Comput. Electron. Agric. 226, 109386. https://doi.org/10.1016/j.compag.2024.109386.

Chen, J., Jönsson, Per, Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004a. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sens. Environ. 91, 332–344. https://doi.org/10.1016/j.rse.2004.03.014.

Chen, J., Jönsson, Per, Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004b. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sens. Environ. 91, 332–344. https://doi.org/10.1016/j.rse.2004.03.014.

Daughtry, C.S.T., Hunt, E.R., 2008. Mitigating the effects of soil and residue water contents on remotely sensed estimates of crop residue cover. Remote Sensing of Environment, Remote Sensing Data Assimilation Special Issue 112, 1647–1657. https://doi.org/10.1016/j.rse.2007.08.006.

Delbart, N., Le Toan, T., Kergoat, L., Fedotova, V., 2006. Remote sensing of spring phenology in boreal regions: a free of snow-effect method using NOAA-AVHRR and SPOT-VGT data (1982–2004). Remote Sens. Environ. 101, 52–62. https://doi.org/10.1016/j.rse.2005.11.012.

Diao, C., 2020. Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. Remote Sens. Environ. 248, 111960. https://doi.org/10.1016/j.rse.2020.111960.

Diao, C., Li, G., 2022. Near-surface and high-resolution satellite time series for detecting crop phenology. Remote Sens 14, 1957. https://doi.org/10.3390/rs14091957.

Diao, C., Yang, Zijun, Gao, F., Zhang, X., Yang, Zhengwei, 2021. Hybrid phenology matching model for robust crop phenological retrieval. ISPRS J. Photogramm. Remote Sens. 181, 308–326. https://doi.org/10.1016/j.isprsjprs.2021.09.011.

Dong, T., Shang, J., Qian, B., Liu, J., Chen, J.M., Jing, Q., McConkey, B., Huffman, T., Daneshfar, B., Champagne, C., Davidson, A., MacDonald, D., 2019. Field-scale crop seeding date estimation from MODIS data and growing degree days in Manitoba, Canada. Remote Sens 11, 1760. https://doi.org/10.3390/rs11151760.

Fan, C., Yang, J., Zhao, G., Dai, J., Zhu, M., Dong, J., Liu, R., Zhang, G., 2023. Mapping phenology of complicated wetland landscapes through harmonizing Landsat and Sentinel-2 imagery. Remote Sens 15, 2413. https://doi.org/10.3390/rs15092413.

Gao, F., Zhang, X., 2021. Mapping crop phenology in near real-time using satellite remote sensing: challenges and opportunities. J Remote Sens. https://doi.org/10.34133/2021/8379391, 2021, 2021/8379391.

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. https://doi.org/10.1016/j.rse.2016.11.004.

Hively, W.D., Lamb, B.T., Daughtry, C.S.T., Serbin, G., Dennison, P., Kokaly, R.F., Wu, Z., Masek, J.G., 2021. Evaluation of SWIR crop residue bands for the Landsat next Mission. Remote Sens 13, 3718. https://doi.org/10.3390/rs13183718.

Jiang, C., Guan, K., Huang, Y., Jong, M., 2024. A vehicle imaging approach to acquire ground truth data for upscaling to satellite data: a case study for estimating harvesting dates. Remote Sens. Environ. 300, 113894. https://doi.org/10.1016/j.rse.2023.113894.

Jin, X., Ma, J., Wen, Z., Song, K., 2015. Estimation of maize residue cover using Landsat-8 OLI image spectral information and textural features. Remote Sens 7, 14559–14575. https://doi.org/10.3390/rs71114559.

Kim, Y.-U., Webber, H., Adiku, S.G.K., de Nóia Júnior, R.S., Deswarte, J.-C., Asseng, S., Ewert, F., 2024. Mechanisms and modelling approaches for excessive rainfall stress on cereals: waterlogging, submergence, lodging, pests and diseases. Agric. For. Meteorol. 344, 109819. https://doi.org/10.1016/j.agrformet.2023.109819.

Kusumastuti, R.D., van Donk, D.P., Teunter, R., 2016. Crop-related harvesting and processing planning: a review. Int. J. Prod. Econ. 174, 76–92. https://doi.org/10.1016/j.ijpe.2016.01.010.

Liu, L., Cao, R., Chen, J., Shen, M., Wang, S., Zhou, J., He, B., 2022a. Detecting crop phenology from vegetation index time-series data by improved shape model fitting in each phenological stage. Remote Sens. Environ. 277, 113060. https://doi.org/10.1016/j.rse.2022.113060.

Liu, Y., Rao, P., Zhou, W., Singh, B., Srivastava, A.K., Poonia, S.P., Van Berkel, D., Jain, M., 2022b. Using Sentinel-1, Sentinel-2, and planet satellite data to map field-level tillage practices in smallholder systems. PLoS One 17, e0277425. https://doi.org/10.1371/journal.pone.0277425.

Liu, Y., Diao, C., Yang, Z., 2023. CropSow: an integrative remotely sensed crop modeling framework for field-level crop planting date estimation. ISPRS J. Photogramm. Remote Sens. 202, 334–355. https://doi.org/10.1016/j.isprsjprs.2023.06.012.

Liu, H., Zhang, H.K., Huang, B., Yan, L., Tran, K.K., Qiu, Y., Zhang, X., Roy, D.P., 2024. Reconstruction of seamless harmonized Landsat Sentinel-2 (HLS) time series via self-supervised learning. Remote Sens. Environ. 308, 114191. https://doi.org/10.1016/j.rse.2024.114191.

Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M., Hammer, G.L., 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. Science 344, 516–519. https://doi.org/10.1126/science.1251423.

Löw, J., Ullmann, T., Conrad, C., 2021. The impact of Phenological developments on interferometric and Polarimetric crop signatures derived from Sentinel-1: examples from the DEMMIN study site (Germany). Remote Sens 13, 2951. https://doi.org/10.3390/rs13152951.

Löw, J., Hill, S., Otte, I., Thiel, M., Ullmann, T., Conrad, C., 2024. How phenology shapes crop-specific Sentinel-1 PolSAR features and InSAR coherence across multiple years and orbits. Remote Sens 16, 2791. https://doi.org/10.3390/rs16152791.

Lyu, F., Yang, Z., Diao, C., Wang, S., 2025. Multistream STGAN: a spatiotemporal image fusion model with improved temporal transferability. IEEE J. Sel. Top. Appl. Earth

Observations Remote Sensing 18, 1562–1576. https://doi.org/10.1109/JSTARS.2024.3506879.

Mardaneh, E., Loxton, R., Meka, S., Gamble, L., 2021. A decision support system for grain harvesting, storage, and distribution logistics. Knowl.-Based Syst. 223, 107037. https://doi.org/10.1016/j.knosys.2021.107037.

Masek, J.G., Wulder, M.A., Markham, B., McCorkel, J., Crawford, C.J., Storey, J., Jenstrom, D.T., 2020. Landsat 9: empowering open science and applications through continuity. Remote Sens. Environ. 248, 111968. https://doi.org/10.1016/j.rse.2020.111968.

Moulin, S., Kergoat, L., Viovy, N., Dedieu, G., 1997. Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements. J. Clim. 10, 1154–1170. https://doi.org/10.1175/1520-0442(1997)010<1154:GSAOVP>2.0.CO;2.

Nikaein, T., Iannini, L., Molijn, R.A., Lopez-Dekker, P., 2021. On the value of Sentinel-1 InSAR coherence time-series for vegetation classification. Remote Sens 13, 3300. https://doi.org/10.3390/rs13163300.

Pandit, A., Sawant, S., Mohite, J., Pappula, S., 2022. Sentinel-1-derived coherence time-series for crop monitoring in Indian agriculture region. Geocarto Int. 37, 9497–9517. https://doi.org/10.1080/10106049.2021.2022008.

Pugh, T.A.M., Arneth, A., Olin, S., Ahlström, A., Bayer, A.D., Klein Goldewijk, K., Lindeskog, M., Schurgers, G., 2015. Simulated carbon emissions from land-use change are substantially enhanced by accounting for agricultural management. Environ. Res. Lett. 10, 124008. https://doi.org/10.1088/1748-9326/10/12/124008.

Quemada, M., Hively, W.D., Daughtry, C.S.T., Lamb, B.T., Shermeyer, J., 2018. Improved crop residue cover estimates obtained by coupling spectral indices for residue and moisture. Remote Sens. Environ. 206, 33–44. https://doi.org/10.1016/j.rse.2017.12.012.

Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O., 1994. Measuring phenological variability from satellite imagery. J. Veg. Sci. 5, 703–714. https://doi.org/10.2307/3235884.

Riggs, G.A., Hall, D.K., Román, M.O., 2016. MODIS Snow Products Collection 6 User Guide. NASA National Snow and Ice Data Center. Available at: https://modis-snow-ice.gsfc.nasa.gov/uploads/snow_user_guide_C6.1_final_revised_april.pdf.

Roy, D.P., Kovalskyy, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sens. Environ. 185, 57–70. https://doi.org/10.1016/j.rse.2015.12.024.

Sacks, W.J., Kucharik, C.J., 2011. Crop management and phenology trends in the U.S. Corn Belt: impacts on yields, evapotranspiration and energy balance. Agric. For. Meteorol. 151, 882–894. https://doi.org/10.1016/j.agrformet.2011.02.010.

Sakamoto, T., 2018a. Refined shape model fitting methods for detecting various types of phenological information on major U.S. crops. ISPRS J. Photogramm. Remote Sens. 138, 176–192. https://doi.org/10.1016/j.isprsjprs.2018.02.011.

Sakamoto, T., 2018b. Refined shape model fitting methods for detecting various types of phenological information on major U.S. crops. ISPRS J. Photogramm. Remote Sens. 138, 176–192. https://doi.org/10.1016/j.isprsjprs.2018.02.011.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A., Verma, S.B., Suyker, A.E., Arkebauer, T.J., 2010. A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data. Remote Sens. Environ. 114, 2146–2159. https://doi.org/10.1016/j.rse.2010.04.019.

Schlund, M., Erasmi, S., 2020. Sentinel-1 time series data for monitoring the phenology of winter wheat. Remote Sens. Environ. 246, 111814. https://doi.org/10.1016/j.rse.2020.111814.

Schwartz, M.D., Reed, B.C., White, M.A., 2002. Assessing satellite-derived start-of-season measures in the conterminous USA. Int. J. Climatol. 22, 1793–1805. https://doi.org/10.1002/joc.819.

Shang, J., Liu, J., Poncos, V., Geng, X., Qian, B., Chen, Q., Dong, T., Macdonald, D., Martin, T., Kovacs, J., Walters, D., 2020. Detection of crop seeding and harvest through analysis of time-series Sentinel-1 interferometric SAR data. Remote Sens 12, 1551. https://doi.org/10.3390/rs12101551.

Shen, Y., Zhang, X., Yang, Z., 2022a. Mapping corn and soybean phenometrics at field scales over the United States Corn Belt by fusing time series of Landsat 8 and Sentinel-2 data with VIIRS data. ISPRS J. Photogramm. Remote Sens. 186, 55–69. https://doi.org/10.1016/j.isprsjprs.2022.01.023.

Shen, Y., Zhang, X., Yang, Z., 2022b. Mapping corn and soybean phenometrics at field scales over the United States Corn Belt by fusing time series of Landsat 8 and Sentinel-2 data with VIIRS data. ISPRS J. Photogramm. Remote Sens. 186, 55–69. https://doi.org/10.1016/j.isprsjprs.2022.01.023.

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. https://doi.org/10.1016/j.rse.2023.113729.

Stern, A.J., Daughtry, C.S.T., Hunt, E.R., Gao, F., 2023. Comparison of five spectral indices and six imagery classification techniques for assessment of crop residue cover using four years of Landsat imagery. Remote Sens 15, 4596. https://doi.org/10.3390/rs15184596.

Thornton, M., Shrestha, R., Wei, Y., Thornton, P.E., Kao, S., Wilson, B.E., 2020. Daymet: daily surface weather data on a 1-km grid for North America. Version 4. https://doi.org/10.3334/ORNLDAAC/1840.

USDA-NASS, 2024. Crop Progress Report. USDA National Agricultural Statistic Service. URL. https://www.nass.usda.gov/Publications/National_Crop_Progress/ (accessed 6.30.25).

Wang, N., Zhang, X., Zhang, T., Pu, L., Zhan, X., Xu, X., Hu, Y., Shi, J., Wei, S., 2022. A sparse-model-driven network for efficient and high-accuracy InSAR phase filtering. Remote Sens 14, 2614. https://doi.org/10.3390/rs14112614.

Wu, C., Peng, D., Soudani, K., Siebicke, L., Gough, C.M., Arain, M.A., Bohrer, G., Lafleur, P.M., Peichl, M., Gonsamo, A., Xu, S., Fang, B., Ge, Q., 2017. Land surface phenology derived from normalized difference vegetation index (NDVI) at global FLUXNET sites. Agric. For. Meteorol. 233, 171–182. https://doi.org/10.1016/j.agrformet.2016.11.193.

Xu, J., Meng, J., Quackenbush, L.J., 2019. Use of remote sensing to predict the optimal harvest date of corn. Field Crop Res. 236, 1–13. https://doi.org/10.1016/j.fcr.2019.03.003.

Xu, G., Gao, Y., Li, J., Xing, M., 2020. InSAR phase Denoising: a review of current technologies and future directions. IEEE Geoscience and Remote Sensing Magazine 8, 64–82. https://doi.org/10.1109/MGRS.2019.2955120.

Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. Remote Sens. Environ. 172, 67–86. https://doi.org/10.1016/j.rse.2015.10.034.

Yang, Y., Ren, W., Tao, B., Ji, L., Liang, L., Ruane, A.C., Fisher, J.B., Liu, J., Sama, M., Li, Z., Tian, Q., 2020. Characterizing spatiotemporal patterns of crop phenology across North America during 2000–2016 using satellite imagery and agricultural survey data. ISPRS J. Photogramm. Remote Sens. 170, 156–173. https://doi.org/10.1016/j.isprsjprs.2020.10.005.

Yang, Z., Diao, C., Li, B., 2021. A robust hybrid deep learning model for spatiotemporal image fusion. Remote Sens 13, 5005. https://doi.org/10.3390/rs13245005.

Zeng, L., Wardlow, B.D., Xiang, D., Hu, S., Li, D., 2020. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sens. Environ. 237, 111511. https://doi.org/10.1016/j.rse.2019.111511.

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. https://doi.org/10.1016/j.isprsjprs.2023.09.025.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9.

Zhang, H.K., Roy, D.P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., Roger, J.-C., 2018. Characterization of sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. Remote Sens. Environ. 215, 482–494. https://doi.org/10.1016/j.rse.2018.04.031.

Zheng, B., Campbell, J.B., de Beurs, K.M., 2012. Remote sensing of crop residue cover using multi-temporal Landsat imagery. Remote Sens. Environ. Remote Sensing of Urban Environments 117, 177–183. https://doi.org/10.1016/j.rse.2011.09.016.

Zheng, B., Chenu, K., Chapman, S.C., 2016. Velocity of temperature and flowering time in wheat – assisting breeders to keep pace with climate change. Glob. Chang. Biol. 22, 921–933. https://doi.org/10.1111/gcb.13118.