A novel Greenness and Water Content Composite Index (GWCCI) for soybean mapping from single remotely sensed multispectral images

Hui Chen [a, b], Huapeng Li [a *], Zhao Liu [a, b], Ce Zhang [c, d], Shuqing Zhang [a *], Peter M. Atkinson [c, e, f]

[a] *Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China*

[b] *University of Chinese Academy of Sciences, Beijing 100049, China*

[c] *Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster LA1 4YQ, UK*

[d] *UK Centre for Ecology & Hydrology, Library Avenue, Bailrigg, Lancaster LA1 4AP, UK*

[e] *Geography and Environmental Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

[f] *Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A Datun Road, Beijing 100101, China*

Abstract

As a critical source of food and one of the most economically significant crops in the world, soybean plays an important role in achieving food security. Large area accurate mapping of soybean has long been a vital, but challenging issue in remote sensing, relying heavily on large-volume and representative training samples, whose collection is time-consuming and

* Corresponding author.
E-mail addresses: lihuapeng@iga.ac.cn (H.P. Li); zhangshuqing@iga.ac.cn (S.Q. Zhang).

inefficient, especially for large areas (e.g., national scale). Thus, methods are needed that can map soybean automatically and accurately from single-date remotely sensed imagery. In this research, a novel Greenness and Water Content Composite Index (GWCCI) was proposed to map soybean from just a single Sentinel-2 multispectral image in an end-to-end manner without employing training samples. By capitalizing on the product of the NDVI (related to greenness) and the short-wave infrared (SWIR) band (related to canopy water content), the GWCCI theoretically provides the required information with which to discriminate between soybean and other land cover types. The effectiveness of the proposed GWCCI was investigated in four typical soybean planting regions with contrasting agricultural landscapes distributed in the four major soybean-producing countries in the world (i.e., China, the United States, Brazil and Argentina). In the experiments, the GWCCI method produced a consistently higher accuracy compared to three conventional benchmark classifiers (maximum likelihood classifier (MLC), support vector machine (SVM), random forest (RF)). The GWCCI achieved an average overall accuracy up to 95.66% and a Kappa coefficient up to 0.91 across the four study regions during the period 2017-2021. It was also found that the proposed method was applicable for soybean mapping using any cloud-free scene of imagery dated from, or even near, the time window, demonstrating the robustness of the GWCCI to image acquisition date. The proposed GWCCI method is straightforward, reliable and robust, and represents an important step forward for mapping soybean, one of the most significant crops grown globally.

Keywords: Soybean mapping; Crop classification methods; Automatic mapping methods;

Sentinel-2 imagery; Short-wave infrared (SWIR); Normalized Difference Vegetation Index (NDVI)

1. Introduction

Being rich in plant protein with high nutritional value, soybean is an important food and economic crop in the world, accounting for about 5% (Li et al., 2021a) of the world's cropland. As such, it has a vital role and impact on ensuring food security (Inglada et al., 2015; Matton et al., 2015). Soybean is employed primarily in processed food products, refined soybean oil, and animal feed. At the same time, soybean rhizobia can convert nitrogen in the air into nitrogen element that can be absorbed by vegetation, which is beneficial both to the soil and to the plant (Masuda et al., 2009). Detailed mapping of soybean spatial distribution and its real-time dynamic monitoring (Gómez et al., 2016; Defourny et al., 2019; Li et al., 2019; Li et al., 2021b, 2021c) provide strong decision-support for both government and private sectors on a variety of critical issues involving the plantation and production of soybean and its business management.

Traditional field survey methods require not only a great deal of human and material resources, but a large amount of time (Siyal et al., 2015). Moreover, governments usually report soybean statistics several months after the soybean crop is harvested. Such time delays in the availability of soybean information can hinder sound decision-making for soybean marketing and soybean planting (Nasrallah et al., 2018). Furthermore, the quality of the survey data cannot be guaranteed because of human subjectivity and errors (Liu et al., 2018). Compared to traditional field survey, remote sensing has unique advantages, including large-coverage (Song et al., 2017; Konduri et al., 2020; Ajadi et al., 2021; Li et al., 2021a), timely observation (Thenkabail et al., 2012; Dado et al., 2020) and low cost. Since the late 1990s, various remote sensing image classification methods that can be broadly categorised into two classes (classifier-based and threshold-based), have been applied increasingly for crop mapping and monitoring (Zhang et al., 2022).

For classifier-based methods, adequate samples are usually required to train a classifier and build the classification model (Picoli et al., 2018; Rußwurm and Körner, 2020). The Maximum Likelihood Classifier (MLC) is one of the most commonly used classifiers for crop classification (Zhong et al., 2016a; Ashourloo et al., 2019). However, suffering from the Hughes phenomenon (Chen et al., 1996), it is difficult for the MLC to achieve promising crop mapping results. Along with the rapid development of computer science, machine learning (ML) has entered the field of remote sensing-based crop classifications (Liu et al., 2018; Li et al., 2021c; Turkoglu et al., 2021). ML algorithms, such as the Support Vector Machine (SVM) and Random Forest (RF) (Teluguntla et al., 2018; Griffiths et al., 2019; Calderón-Loor et al., 2021), provide a wide array of opportunities for crop classification based on remotely sensed data. These data-driven algorithms can increase classification accuracy and efficiency by operating on multidimensional data independent of data distribution (de Souza et al., 2015; Li et al., 2021a; Xu et al., 2021). Whereas, they belong to shallow-structured models and, as such, they cannot extract and utilize deep features of remotely sensed imagery (Xu et al., 2021). Furthermore, manual feature engineering (e.g., to produce texture features) is often needed, which is laborious and challenging. Deep learning (DL), a new form of ML, has been shown in previous research to be capable of mining automatically deep information from time-series remotely sensed data (Castro et al., 2018; Marcos et al., 2018; Rußwurm and Körner, 2020). DL techniques can greatly enhance the capability to handle long sequential dependencies, and thus, typically outperform conventional ML algorithms in identifying crops (Zhong et al., 2019; Garnot et al., 2020). However, for these ML models (including DL models), they usually require a large number of training samples and, moreover, they can be hard to generalize the model to different regions and applications (Xu et al., 2020; Ajadi et al., 2021). These major issues limit the practical utility of ML methods for crop classification, especially over large spatial areas (Turkoglu et al., 2021).

Threshold-based approaches (Boschetti et al., 2017) identify crops by quantifying the magnitude of, or variation in, vegetation indices (VIs) or the phenological metrics derived from VIs during the crop growth period (Xu et al., 2021). These methods are usually implemented on time-series images. For example, crops including rice (Kontgis et al., 2015; Qiu et al., 2015; Lu et al., 2017), winter crops (Zhang et al., 2021) and canola (Sulik and Long, 2016) are the prevalent crops classified by threshold-based approaches due to their differences in spectral or phenological features observed in the time-series profile. However, it can be challenging for these methods to separate crops at similar phenological stages (Domínguez et al., 2015; Tian et al., 2019a, 2019b), such as soybean and corn (Zhong et al., 2016a). Besides, in order to exploit the classification potential of multisource and/or multitemporal data sources as much as possible, a relatively large number of thresholds need to be determined subjectively in a conventional threshold-based approach, thus significantly limiting its efficiency and accuracy

(Kontgis et al., 2015).

Belonging to an advanced threshold-based method, an index-based method with just one threshold being determined objectively, has received increasing attention in recent years (Ashourloo et al., 2019; Jia et al., 2019). Index-based approaches classify a specific crop type(s) from existing remote sensing data (or products) by enhancing the spectral differences between the targeted crop type (s) and others (Ashourloo et al., 2020; Qu et al., 2021). The advantages of index-based approaches include their mathematical simplicity and ease of computation (Ashourloo et al., 2019), thus, making them more practical. For instance, canola was accurately mapped from Sentinel-2 imagery at its flowering period by a novel index built on three spectral bands (red, green and near infrared bands) (Ashourloo et al., 2019); An index based on the spectral profile of near-infrared (NIR) and red bands during the cultivation, peak greenness and harvest stages was constructed to detect potato with time-series of Sentinel-2 images (Ashourloo et al., 2020); More recently, a Winter Wheat Index (WWI) was developed to identify winter wheat using the normalized difference vegetation index (NDVI) data acquired at the four critical growth stages of the crop (Qu et al., 2021). These previous researches well demonstrate that index-based approaches are not only more accurate, but more cost-efficient, in comparison with classifier-based approaches. Yet, to the best of our knowledge, an index that can well reflect the unique yet complex spectral characteristics of soybean and thus effectively distinguish this valuable crop from other land use/cover types from remotely sensed images, has so far not been designed.

In previous studies, NDVI time-series data from Sentinel (Radočaj et al., 2020; Wang et al., 2020), Landsat (Zhong et al., 2014; Cai et al., 2018) and Moderate Resolution Imaging Spectroradiometer (MODIS) (de Souza et al., 2015; Picoli et al., 2018), were used for soybean identification. However, high classification accuracies were not attained because although sufficient spectral differences existed between soybean and non-crop land covers, serious spectral confusion arose between soybean and some crop types (Zhang et al., 2020). Meanwhile, the SWIR band was found to be sensitive to canopy water content and, thus, capable of discriminating soybean from certain other crops that had similar phenological stages, but varied in water content with soybean (e.g., corn) (Zhong et al., 2016a; Elsherif et al., 2018; Zhang and Zhou, 2019; Zhang et al., 2020; Song et al., 2021b). Furthermore, the variations in the greenness and canopy water content--two important biophysical properties of soybean, are usually correlated during its growth process (Colombo et al., 2008). For example, while the greenness of soybean increases along with crop development, its canopy water content increases too. Such a biophysical process provides us a new perspective to solve the complex issue of soybean mapping based on single and appropriate imagery, by jointly and simultaneously mining the information of NDVI and SWIR band. In this paper, time-series data of the NDVI and the short-wave infrared (SWIR), were jointly investigated to quantify and highlight dynamic changes in the spectral response of soybean over time. On this basis, a novel index named Greenness and Water Content Composite Index (GWCCI) was established to identify and map soybean using only a single imagery during its peak growing season. The effectiveness and robustness of GWCCI were comprehensively tested in four major soybean-producing countries, with diverse and various environmental conditions, across the world.

2. Materials

2.1 Study areas

Four study sites under varying environmental conditions, located in four major soybean-producing countries (China, the United States (US), Brazil and Argentina) accounting for ~90% of the world's total soybean production (Wilcox et al., 2004; Schwalbert et al., 2020) (Fig.1), were chosen as our study areas to test the effectiveness and generalization of our method.

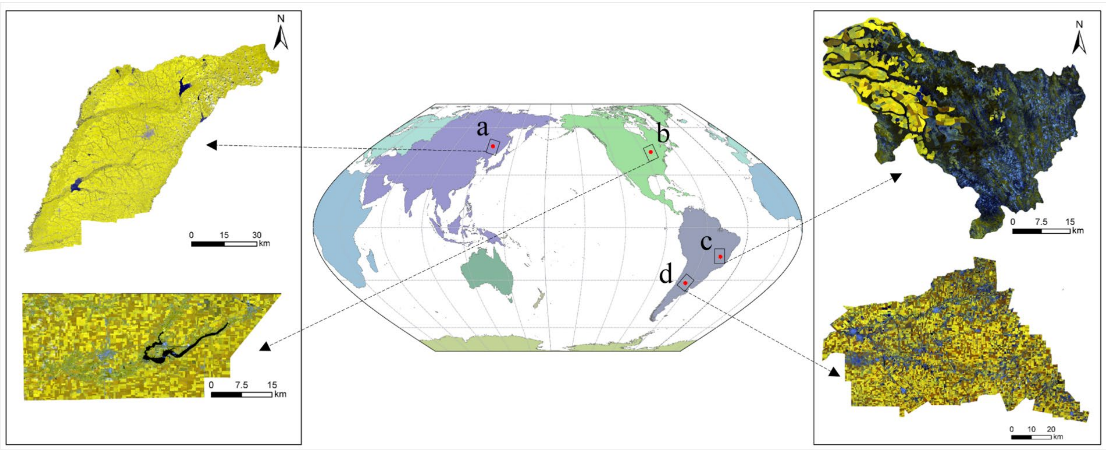

Fig. 1. The geographical locations of the selected four counties (Hailun, China (a); De Witt, US (b); Bonfinopolis de Minas, Brazil (c); Rio Segundo, Argentina (d); marked in black rectangle with red dots inside), distributed in three continents including Asia, North America and South America in the world. Fig. 1(a-d) (the left and right panels), the corresponding false colour Sentinel-2 images (R: B7, G: B6, B: B3) of the four counties, respectively.

The first selected site, Hailun (Fig. 1a) county, is located in Heilongjiang province, China. With rich, black soils, the province is a major soybean-producing region in China, accounting for over 40% of the country's total soybean production (Li et al., 2021a). The county has a temperate continental monsoon climate, with four distinct seasons (severely cold in winter; hot in summer), with rainfall and high temperatures concentrating in the same season. Soybean is seeded in spring (from late April to early May) and harvested in autumn (from late September to early October). The spatial pattern of soybean in this county is distributed densely. Due to the management of farmers (not farms), most of the soybean fields are small in area and have a long and narrow rectangular shape. In the US, the De Witt county (Fig. 1b), lying in the central area of the Corn Belt was chosen. The US Corn Belt consists of 12 states across the north-central region of the US, and makes up over 75% of the US soybean production (Johnson, 2014) and 28% of the global soybean production (Riccetto et al., 2020), respectively. The belt is relatively flat with fertile soil, and the climate of the region is temperate continental climate (Haigh et al., 2015). As one of the most important crops of the belt, soybean is generally planted from early May to June, and harvested from late September to late October (Wang et al., 2019). In De Witt, farms are managed by farmers which control the agricultural production processes including crop sowing, fertilization and harvesting. The agricultural landscape composition of this county is relatively simple, with soybean and corn accounting for over 90% of the county's crop production.

In addition to the two counties in China and the US, two additional counties in Brazil and Argentina, respectively, were also selected to further test the proposed method. The Bonfinopolis de Minas (denoted as Bonfinopolis hereafter) county (Fig. 1c) is sited in the state of Minas Gerais, Brazil, one of the most productive agricultural zones for decades (Wilcox et al., 2004). Bonfinopolis has a typical tropical savanna climate, with high temperatures throughout the year and most of annual precipitation concentrated in summer (Sayago et al., 2017). Agriculture in Bonfinopolis is intensified by a double cropping system, and soybean is generally planted in the first season in spring and summer (approximately from October to April) (Zhong et al., 2016a). The Rio Segundo (Fig. 1d) county, Argentina, is located in the heart of the Pampas where over 90% of the country's soybean is produced (Al-Mamoori et al., 2021). The climate of the county is classified as dry subhumid with an annual rainfall amount of nearly 800 mm (mainly concentrated in summer) (Sayago et al., 2017). Two summer crop types, corn and soybean, planted in October and harvested as late as May, dominate the county (Antonio et al., 2021).

Obviously, the selected four study areas (with detailed description in Table 1) differing in climate, phenology, cropping system, planting structure, and crop management, are suitable for testing the effectiveness of the proposed approach.

Table 1 Detailed descriptions of the study areas. MSC_SA: annual maps of soybean cover over South America by Song et al. (2021a).

Study sites (counties)	Location	Main crops	Soybean planting ratio (%)	Average field size of soybean (km^2)	Cropping system	Resources of Reference	Soybean phenology
Hailun	Heilongjiang province, China, Asia	soy, corn, rice, potato, onion, golden berry	34.50	0.032	Single cropping system	field survey	Soybean is seeded in spring (from late April to early May) and harvested in autumn (late September to early October).
De Witt	Illinois, USA, North America	soybean, corn	40.15	0.636	Single cropping system	CDL+ google earth	Soybean is generally planted from early May to June, and harvested from late September to late October.
Bonfinopolis	Minas Gerais, Brazil,	soybean, corn, cotton	17.50	0.429	Double cropping	MSC_SA +	Soybean is generally planted in the first season

	South America				system	google earth	in spring and summer (approximately from October to April).
Rio Segundo	Córdoba, Argentina, South America	soybean, corn	47.45	0.924	Double cropping system	MSC_SA + google earth	Summer crops predominate the county, where soybean is planted in October and harvested as late as May.

2.2 Data in this study

2.2.1 Ground reference data

Ground reference data were collected in each of the study areas. For Hailun (Fig. 1a), field survey was conducted along the road network in early August, 2021. A total of 670 field patches were identified and digitised using the ArcGIS 10.7 software. For the De Witt in the US (Fig. 1b), ground reference data was acquired according to the Cropland Data Layer (CDL) achieved annually by the US Department of Agriculture (USDA) (Boryan et al., 2011). The CDL has been used as the reference in a wide range of crop mapping applications (Li et al., 2019; Zhang et al., 2019b; Li et al., 2020; Li et al., 2021b) in the light of its very high quality. For example, the average accuracy of corn and soybean is over 90% for the year of 2021. As for the two counties in South America, the reference of soybean was obtained from the product of annual maps of soybean cover over South America (denoted as MSC_SA) by Song et al. (2021a) with an average overall accurate of 95% from 2017 to 2019. In the latter three counties in the North America and South America, 40 crop patches (including soybean and non-soybean) were randomly selected and digitised in each county with the support of the ArcGIS 10.7 software.

To ensure that training and testing samples were taken from different patches, all the collected reference patches of each study area were split randomly into two subsets, respectively: approximately 70% for training and the remaining 30% for testing. A stratified random sampling scheme was adopted to produce samples within the training and testing patches. For each county, pixels falling into the training patches were used as training samples to train the supervised classifiers, whereas a total of 2000 pixels (1000 soybean pixels and 1000 non-soybean pixels) were randomly selected from the testing patches and used for classification accuracy testing (Liu et al., 2020).

2.2.2 Sentinel-2 data

The Sentinel-2 (S2) images with cloud cover of less than 10% were employed for soybean mapping and classification across the study areas. The S2 images dated during the period 2017-2021 were collected from the Sentinel Scientific Data Hub of the European Space Agency (ESA) (2017-2018) and the Google Earth Engine (GEE) (2019-2021) (da Silva Junior et al., 2020) respectively, for all study areas except for the county of China, where only images of the year 2021 were used due to the lack of ground reference data during the year of 2017-2020 (Table 2). Note, the proposed approach was established based on the analysis of spectral variations over Hailun, and thus their S2 time-series images of a full year ("full year" in Table 2) in 2021 were employed for the derivation of index; whereas in each of the other study sites, only S2 images within a time period (roughly corresponding to the peak growing season) were acquired. The preprocessing stage of the S2 images from ESA (2017- 2018), involving the default atmospheric and topographic corrections and the resampling of bands, was undertaken by using the Sen2Cor toolbox (version.9.0, ESA, 2021) and the SNAP software. There are 13 bands in S2: five visible and near-infrared bands (Bands 2-4, and 8, 10m; Band 8A, 20m); three red edge bands (Bands 5-7; 20m); two SWIR bands (Bands 11-12; 20m) (Liu et al., 2021), and three other bands (Bands 1, 9 and 10; 60m). The first 10 bands (Bands 2-8, 8A and Bands 11-12) were used in this research as they were designed for vegetation monitoring (Berger et al., 2012), and the original bands with a spatial resolution of 20m were resampled to 10m to maintain consistency of spatial resolution. Moreover, since the spatial extent of each county covers more than one scene of remotely sensed image, the image dataset of each date in each study area was acquired through image mosaicking (with usable images in a specific time period (e.g., 7.15-8.30)) and masking with the corresponding county's boundary.

Table 2 Date of Sentinel-2 image acquisitions for soybean mapping in each study site. In Hailun, S2 images with a "full year" of 2021 were included for deriving GWCCI, but only a single dated S2 image in the parenthesis was used for soybean mapping. In other study sites, an image dataset of each year might be mosaicked with usable images acquired within a time period (e.g., 7.15-8.30 for De Witt in 2021).

Counties	Location	2021	2020	2019	2018	2017
Hailun	China	Full year (7.19)	\	\	\	\
De Witt	USA	7.15-8.30	7.15-8.30	7.15-8.30	8.4	8.29
Bonfinopolis	Brazil	1.15-3.13	\	1.15-3.13	1.20	2.24
Rio Segundo	Argentina	2.05-2.20	2.05-2.20	2.05-2.20	2.14	2.19

3. Methods

In this paper, we developed a novel Greenness and Water Content Composite Index (denoted as GWCCI) that combines NDVI and the shortwave infrared (SWIR) band to accurately map soybean in its peak growing season. The GWCCI was established by first investigating the critical phenological stage for soybean identification and then designing a proper means of combining both NDVI and SWIR to maximally reflect the information difference between soybean and other land cover types. The workflow of this research consists of the following three components (Fig. 2): (1) data preparation, including data acquisition and standard image preprocessing (i.e., atmospheric and geometric correction, and image mosaicking); (2) derivation of the GWCCI, which includes phenological analysis of soybean and other land cover types, and the analysis of the spectral dynamics of both NDVI and SWIR, and the definition of the index; and (3) application of the GWCCI for soybean mapping and validation of the soybean mapping results.

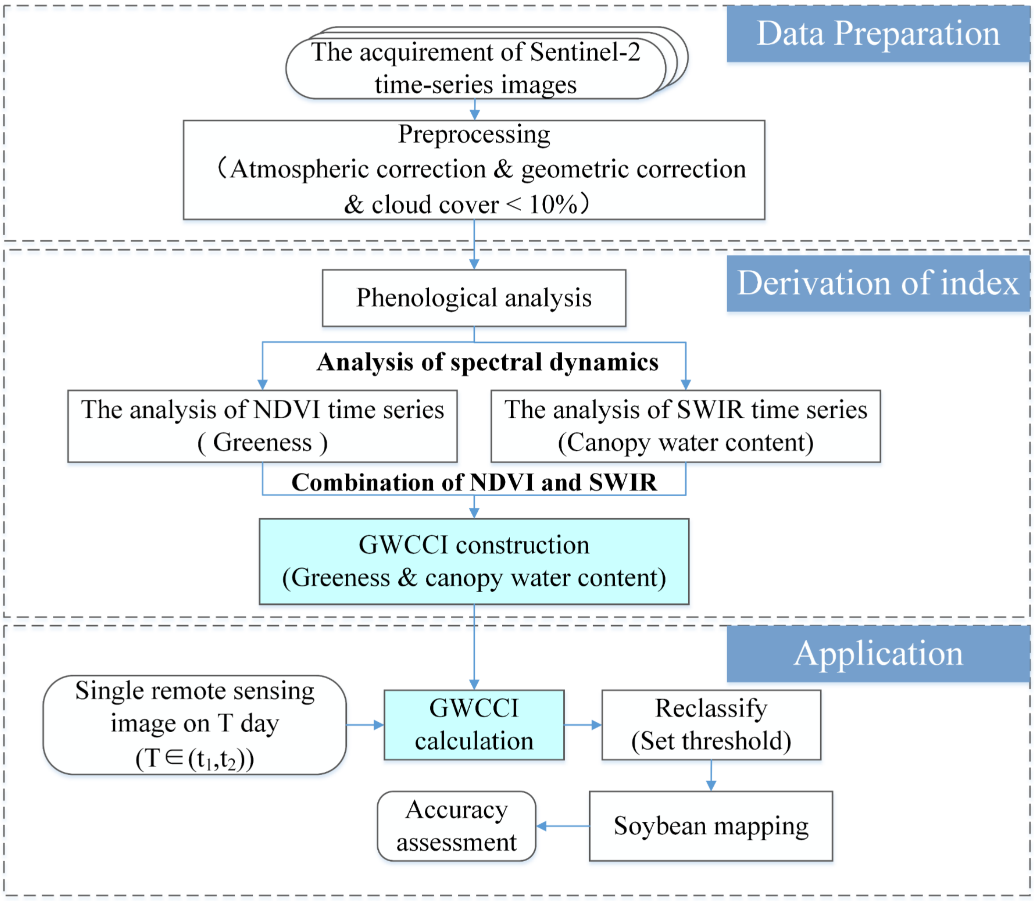

Fig. 2. Workflow of this research for soybean mapping.

3.1 Phenological analysis

Here, the GWCCI was derived based on the full year S2 images and the corresponding ground reference data in Hailun. Based on field investigation and the related literature (Paul et al., 2021), the land cover types in the two counties include three major crops (soybean, corn, rice), woodland, built-up area, water, and some other minor crops (e.g., potato, adzuki bean, gold berry, which are uniformly called "other crops").

The phenological calendars (Zhang et al., 2020) of three main crops (i.e., soybean, corn and rice) and other crops, are shown in Fig. 3. Specifically, the rice is sowed before mid-April, followed by its phenological stages of emergence, planting, greening up, tillering, booting, heading, milky, maturity and senescence. The sowing stage of dryland crops normally occurs in late April (soybean and other crops) or around mid-April (corn), followed by the phenological stages of emergence, three leaves, blooming, pod bearing, seed filling and senescence (Hu et al., 2018). All these detailed phenological stages can be broadly divided into three distinct growth stages; the early growing season (early May to mid-July), the peak (mid) growing season (mid-July to late August) and the late growing season (late August to mid-October) (Ashiq et al., 2021). The peak growing season of the crops is critical since it is during this time that biomass reaches a peak, as do other indicators like greenness and canopy water content. From Fig. 3 it can be seen that there is some overlap between the phenological properties of these crops, since they are all grown in the same term with a hydrothermal synchronization.

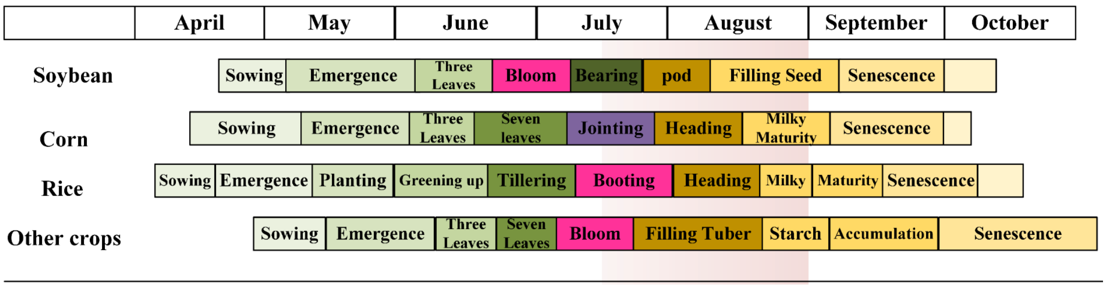

Fig. 3. Phenological calendars for major crops and other crops.

3.2 Analysis of spectral dynamics

3.2.1 NDVI time-series profile

NDVI is a widely used vegetation index that represents the greenness of vegetation. The greater the NDVI, the greater the greenness (Chen et al., 2021). For Sentinel-2 imagery, NDVI can be calculated using the red (ρ_{RED}) and near-infrared (ρ_{NIR}) spectral bands (Eq. (1)):

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{RED}}} \qquad (1)$$

By using Eq. (1) and Sentinel-2 time-series data, the NDVI time-series profile was acquired to explore the greenness of the land cover types. The NDVI time-series profiles derived from GEE during the whole year of 2021 are shown in Fig. 4 for each individual land cover type. Specifically, the whole growing season of soybean includes the early growing season (EGS) (day of year (DOY)$\in$[120,195]), the peak growing season (PGS) (DOY$\in$[195,240]) and the late growing season (LGS) (DOY$\in$[240,285]), as marked by the black line segments above the time series profiles (Fig. 4).

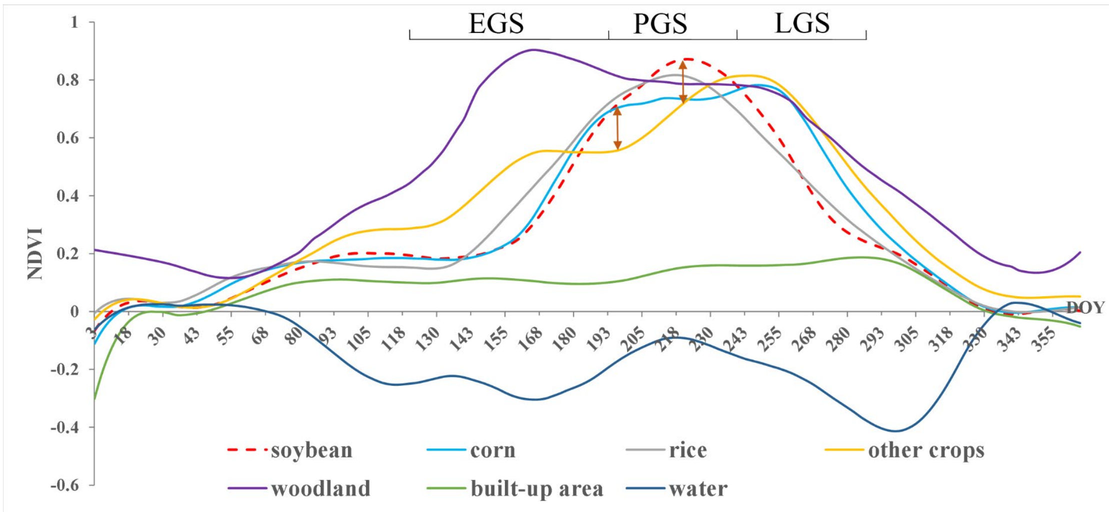

Fig. 4. The NDVI time-series profiles of the main land cover types.

As shown in Fig. 4, there exist distinct differences in NDVI dynamics between the crops (i.e., soybean, corn, rice and other crops) and non-crop land covers (i.e., built-up area, woodland and water). The NDVI values of non-crops tend to be relatively stable, whereas those of crops (soybean, corn, rice and other crops) become gradually larger with an increase in the day of the year (DOY) during the EGS of soybean, peaking at its DOY≈218, and then declining in its LGS. In addition, the difference between the main crops (soybean, corn and rice) and other crops in NDVI increased gradually during the PGS of soybean (as illustrated by the orange double arrows in Fig. 4). Though the NDVI of soybean is confused with that of corn, woodland and rice at the PGS of soybean, the latter two types of land cover can be identified readily by using the SWIR band, due to their extremely high water content. Therefore, separating soybean and corn, both of which have higher biomass and similar phenological features, will be the key to the success of soybean mapping.

3.2.2 Separability of spectral bands for soybean and corn

To help to identify more features with which to separate soybean and corn, 100 sample pixels were selected randomly for the crops of soybean and corn in each scene of the Sentinel-2 time-series imagery based on ground reference data. The time-series spectral reflectance profiles of soybean and corn, ranging from January to late November, 2021, were then plotted (Fig. 5) with the support of the GEE platform (Xiong et al., 2017).

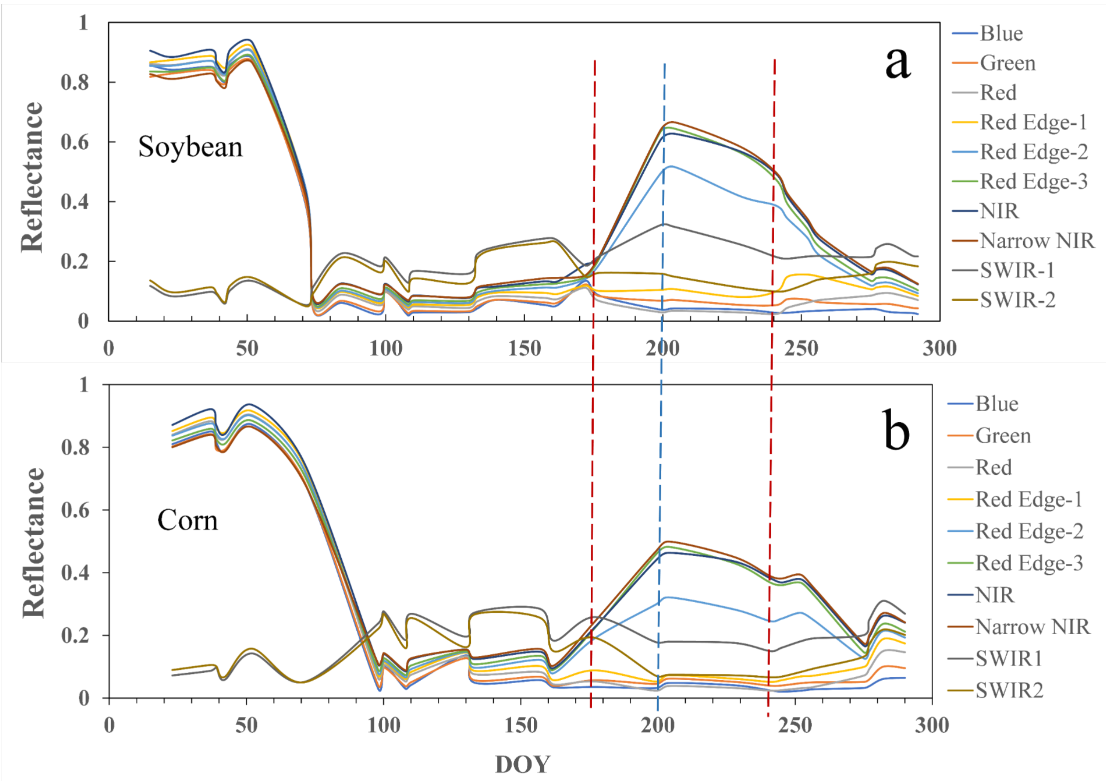

Fig. 5. The annual variation of 10 spectral bands from the S2 images for soybean (a) and corn (b).

As shown by Fig. 5, the major difference in the spectral profile of soybean (Fig. 5a) and corn (Fig. 5b) occurred in the time period from late June (DOY≈175) to late August (DOY≈240), as marked by the two red dotted lines. The reflectance of both soybean and corn started to increase from DOY≈175, peaked at DOY≈200, and then began to decline slowly until DOY≈240. Within the time period of DOY∈(175,200), both the reflectance of soybean and corn rose continuously in a linear manner but with different gradients; the reflectance of soybean increased more sharply than corn, achieving the largest spectral difference at DOY≈200 (the dotted blue line in Fig. 5), at which a peak value of 0.60 and 0.50 was obtained for soybean and corn respectively. Therefore, the middle of July (around day 200) was identified as the optimum date, based on which an index can be designed for distinguishing soybean and corn.

To quantify the separability of soybean and corn, a metric D_i^t was proposed as follows:

$$\mathrm{D}_i^t = \{\mathrm{S}_{\mathrm{min}}^{t,i} - \mathrm{C}_{\mathrm{max}}^{t,i}\} \tag{2}$$

$$\mathrm{S}_{\mathrm{min}}^{t,i} = \mathrm{Min}\{\mathrm{S}_j\mathrm{B}_i^t \,|\, j = (1,2,3,\dots,n); i = (1,2,3,\dots,10)\} \tag{3}$$

$$\mathrm{C}_{\mathrm{max}}^{t,i} = \mathrm{Max}\{\mathrm{C}_j\mathrm{B}_i^t \,|\, j = (1,2,3,\dots,n); i = (1,2,3,\dots,10)\} \tag{4}$$

where i is band i in the Sentinel-2 image, j is the sample ID of soybean or corn; and t is the date (day of the year, DOY, $t \in \mathrm{int}[0,365]$). In band i, D_i^t is the difference between the band's minimum spectral value of soybean ($\mathrm{Min}_{\mathrm{soybean}}$) and the band's maximum spectral value of corn ($\mathrm{Max}_{\mathrm{corn}}$) on day t; $\mathrm{S}_{\mathrm{min}}^{t,i}$ and $\mathrm{C}_{\mathrm{max}}^{t,i}$ is the minimum and the maximum reflectance of soybean and corn in band i on the day t, respectively, which were obtained at 5th and 95th percentiles of the samples for each of the type aiming to eliminate the effect of noise (e.g., residual cloud and poor-quality pixels) (Zhang et al., 2022). A positive value of D_i^t refers to the situation without spectral intersections between soybean and corn.

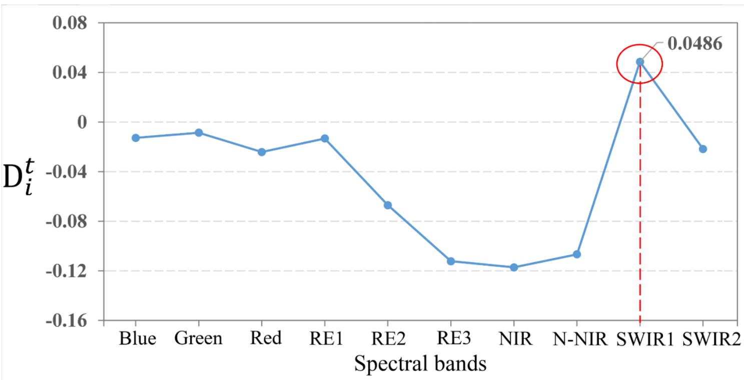

Fig. 6. The D_i^t metric plotted for each of the 10 spectral bands. (RE: Red edge; N-NIR: Narrow NIR).

Based on Eqs. (2)-(4), D_i^t was calculated for each of the 10 spectral bands (Fig. 6). As illustrated by Fig. 6, the only positive D_i^t value (0.0486) appeared for the SWIR1 band (denoted as SWIR hereafter), indicating that SWIR was the most sensitive spectral band for separating soybean and corn. To test the effectiveness of the SWIR band for separating soybean and corn, the SWIR time-series profile was generated for each of the land covers (Fig. 7) using the Sentinel-2 images derived from the GEE platform based on the reference data.

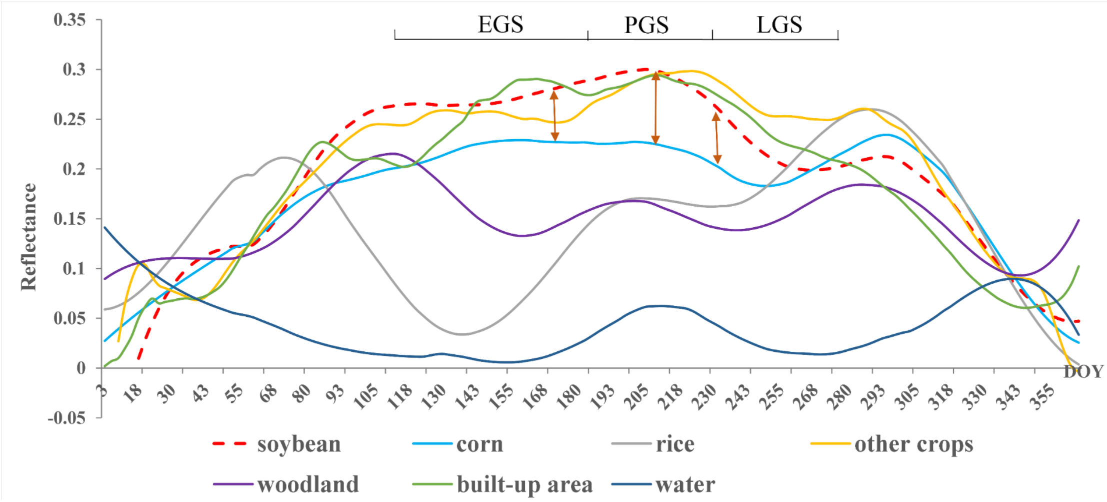

Fig. 7. The SWIR time-series profiles of the main land cover types.

The shortwave infrared reflectance is inversely proportional to canopy water content of vegetation (Yilmaz et al., 2008; Jacquemoud et al., 2009; Zhang et al., 2019a). From the SWIR time-series profiles of the main land cover types (Fig. 7), it can be seen that there was a distinct spectral difference between soybean vs. corn, rice, woodland and water. It was noteworthy that a clear spectral difference between soybean and corn appeared in the PGS of soybean, peaking at DOY≈218 (similar to the timing of peak NDVI (Fig. 4)). In addition, small differences between the two crops could also be observed from the figure during the EGS and LGS of soybean.

3.3 Definition of the Greenness and Water Content Composite Index (GWCCI)

From the above analysis, it was found that during its PGS, soybean maintained an especially high-level greenness and canopy water content, and based on which, soybean could be separated from the rest of the land cover types. Therefore, a Greenness and Water Content Composite Index (GWCCI) was developed by combining the NDVI and SWIR band as follows:

$$\text{GWCCI} = \text{NDVI}_t * \rho_{\text{SWIR}_t} \tag{5}$$

where, ρ_{SWIR} is the reflectance of the SWIR band, t is the date (within PGS) of S2 image. Fig. 8 depicts the GWCCI time-series profile for each land cover type across a full year of 2021.

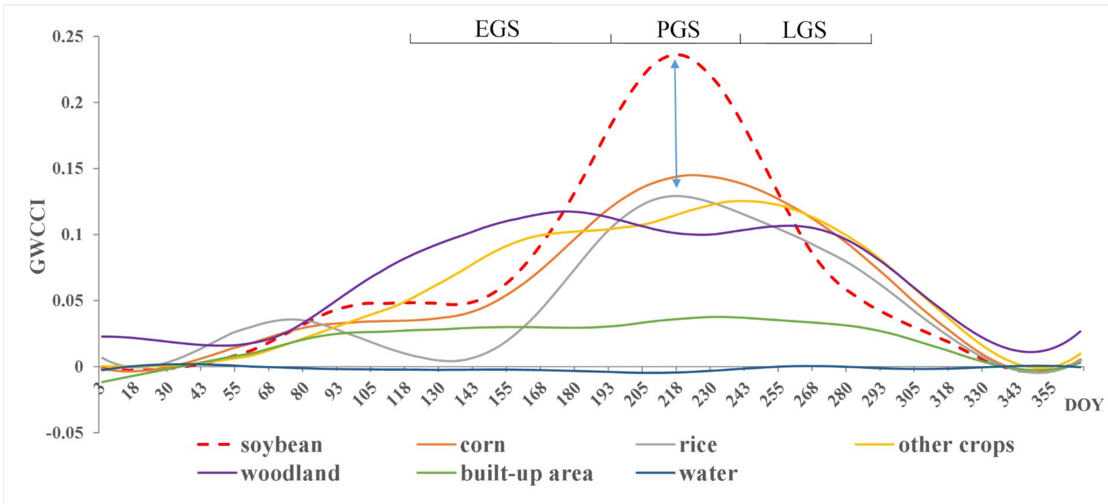

Fig. 8. The GWCCI time-series profiles of the main land cover types.

As shown in Fig. 8, the signal of the soybean is greatly enhanced by the GWCCI during the peak growing season (PGS), whereas those of the other land cover types are normally reduced, thus greatly highlighting the signal difference between them. Therefore, based on the proposed GWCCI, soybean can be identified and mapped with just one scene of remotely sensed imagery within PGS. Soybean mapping using the GWCCI approach was implemented by first calculating GWCCI map from S2 image, and then map soybean though an automatic determination of optimal threshold with a grid search method (Zhang et al., 2018b).

The time period of PGS with a start and end date (t_1, t_2), which is defined as Time Window (TW), is a prerequisite for image acquisition and soybean mapping. To determine TW, we adopted the index of Green Chromatic Coordinate (GCC) (Sonnentag et al., 2012; Zhang et al., 2018a; Shen et al., 2022), which is calculated with the following steps:

$$\mathrm{GCC} = \frac{\mathrm{G}}{\mathrm{R} + \mathrm{G} + \mathrm{B}} \tag{6}$$

where R, G and B denote the value of red, green and blue bands, respectively.

GCC increases gradually as crop plants germinate and grow, and tends to be stable during the peak growing season with a little change rate (ρ_t), which can be calculated as follows:

$$\rho_t = \frac{\mathrm{GCC}_t - \mathrm{GCC}_{t-1}}{\mathrm{GCC}_{t-1}} \tag{7}$$

where t is the date; GCC_t and GCC_{t-1} is the GCC value on date t and $(t-1)$, respectively; ρ_t is the change rate of GCC on date t. TW is defined as follows:

$$\{\mathrm{TW} \in (t_1,\ t_2) | \rho_{t_1} = -\varepsilon < \rho_t < \rho_{t_2} = \varepsilon, |\rho_t| < \varepsilon\} \tag{8}$$

where t_1, t_2 are the start and end date of TW respectively, and ε is a user-defined threshold.

3.4 Accuracy assessment

The performance of the proposed method was evaluated with four commonly used accuracy assessment indices, including the overall accuracy (OA%), producer's accuracy (PA%), user's accuracy (UA%), and Kappa coefficient (k). A benchmark comparison study was also undertaken with three traditional classification approaches (MLC, SVM, and RF) (Xu et al., 2021). MLC, one of the most commonly used classifiers (Zhong et al., 2016a; Ashourloo et al., 2019), is based on the assumption of a Gaussian distribution. SVM, a non-parametric classifier that makes no assumptions about the distributions of the underlying data, classifies imagery by establishing a hyper-plane using kernel functions (Azadbakht et al., 2019; Zhang et al., 2021). RF is essentially an ensemble classifier which has been employed widely for soybean mapping due to its robustness and convenience (Zhong et al., 2016b; Liu et al., 2019; Li et al., 2020). The control parameters of the three benchmark classifiers were determined following the suggestions of Zhang et al. (2020).

4. Results

4.1 Time window and optimal thresholds of GWCCI for soybean mapping

The change rate (ρ)of GCC and the corresponding TW were determined for each country by using Eqs. (6) -(8) (Fig. 9). As shown by Fig. 9 (a) and (b), located in the North Hemisphere, the two counties in China and the US have a similar TW, with an interval of 45 days from day 195 to day 240 (as marked by the black bold line segments) determined by the change rate of GCC with a threshold (ε) of 0.001 (Eqs. (7) -(8)). In contrast, TWs for the selected counties in the South Hemisphere were relatively narrower, with an interval of 30 days from day 40 to day 70 (ε=0.01) for Brazil's Bonfinopolis (Fig. 9 (c)), and 20 days from day 40 to day 60 (ε=0.005) for Argentina's Rio Segundo (Fig. 9 (d)), respectively.

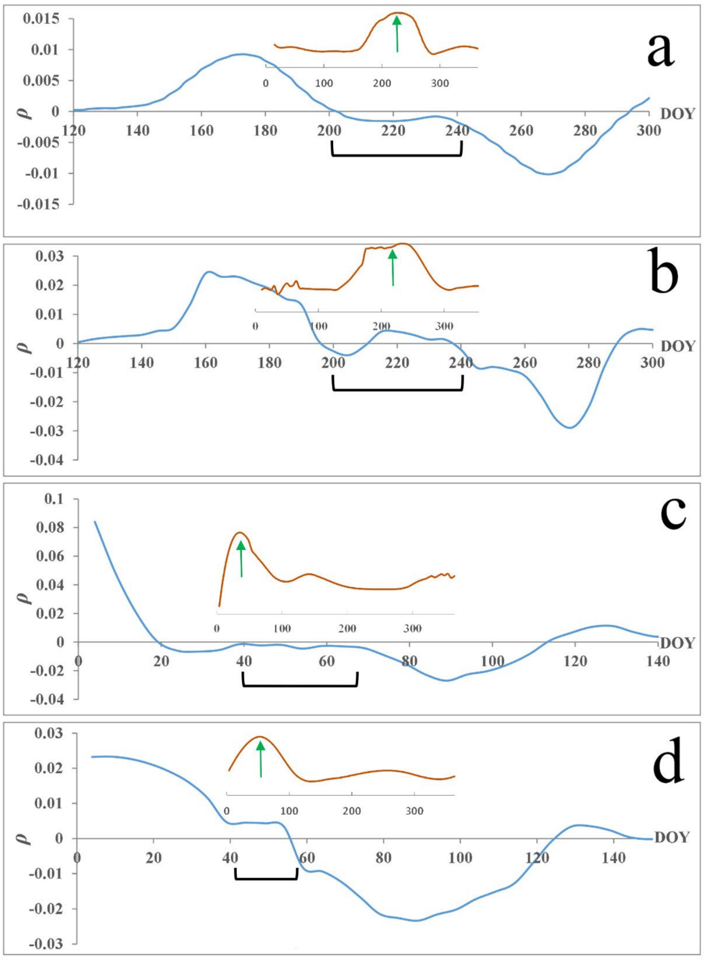

Fig. 9 Variations in change rate (ρ) of the GCC time-series (blue lines) during the whole growing season for Hailun, China (a); De Witt, US (b); Bonfinopolis, Brazil (c); and Rio Segundo, Argentina (d). The black bold line segment in each subfigure illustrates the length of PGS. Red curve shows the profile of the GCC time series, with the peak value marked by a green arrow.

Based on the calculated TWs, the optimal threshold of GWCCI of each county in each of the years (listed in Table 2) was calculated and summarized in Table 3. It can be seen from the table that the optimal threshold values of GWCCI are relatively stable across the study areas from 2017 to 2021, mainly fluctuating around 0.17.

Table 3 The optimal thresholds of the GWCCI determined using a grid search algorithm.

Counties	2021	2020	2019	2018	2017
Hailun	0.1700	\	\	\	\
De Witt	0.1700	0.1739	0.1766	0.1705	0.1700
Bonfinopolis	0.1759	\	0.1695	0.1703	0.1709
Rio Segundo	0.1695	0.1735	0.1701	0.1703	0.1700

4.2 Soybean mapping results

The soybean mapping results by the GWCCI approach across the four counties in 2021 were achieved and presented in Fig. 10 (left panel) (Hailun (Fig. 10A), De Witt (Fig. 10B), Bonfinopolis (Fig. 10C) and Rio Segundo (Fig. 10D)). The proposed method was also compared with the benchmarks. For a better visual comparison effect, some typical areas marked by red rectangles in the GWCCI's soybean maps (Fig. 10 (left panel)) were zoomed, with results of the proposed method and benchmarks being illustrated in Fig. 10 (right panel). The GWCCI was generally able to classify the soybean fields accurately with precise field boundary information, delivering results in relatively good consistency with the reference data, as illustrated by the green circles in Fig. 10. In contrary, obvious classification mistakes were found in some benchmark methods (MLC, SVM and RF), as marked by the red circles. For example, in Hailun, a large piece of soybean field identified by GWCCI (see Fig. 10A (a3)) was erroneously omitted by all three benchmarks. In addition, linear soybean fields (e.g., Fig. 10A (a1, a2 and a4)) and the boundary of a large soybean field (e.g., Fig. 10A (a3)) captured by GWCCI were undetected by the other methods. For the county in the North Hemisphere, the soybean field marked with a green circle (Fig. 10B (b1)) in De Witt was exactly detected by the GWCCI, but some of its soybean pixels were underestimated by MLC, as marked within the red circle. Meanwhile, compared with the benchmarks, a more accurate classification was achieved for soybean-dominated pixels (soybean pixels mixed with a small ratio of other land covers) by GWCCI (Fig. 10B (b2)), with clearly less omission of soybean pixels. Besides, soybean fields with relatively large areas in Bonfinopolis were identified accurately with high geometric fidelity by the GWCCI (Fig. 10C (c1, c2, c3)), but they were partially omitted by SVM and RF, or nearly completely omitted by MLC (Fig. 10C (c2, c3). Moreover, small and fragmented soybean fields in Rio Segundo were detected entirely by the GWCCI (Fig. 10D (d1, d2, d3)), whereas most of these soybean pixels were more or less undetected by the three comparators. In short, the GWCCI acquired desirable results in which soybean fields were accurately separated from other land cover types with not only little salt-and-pepper noise, but also high geometric fidelity.

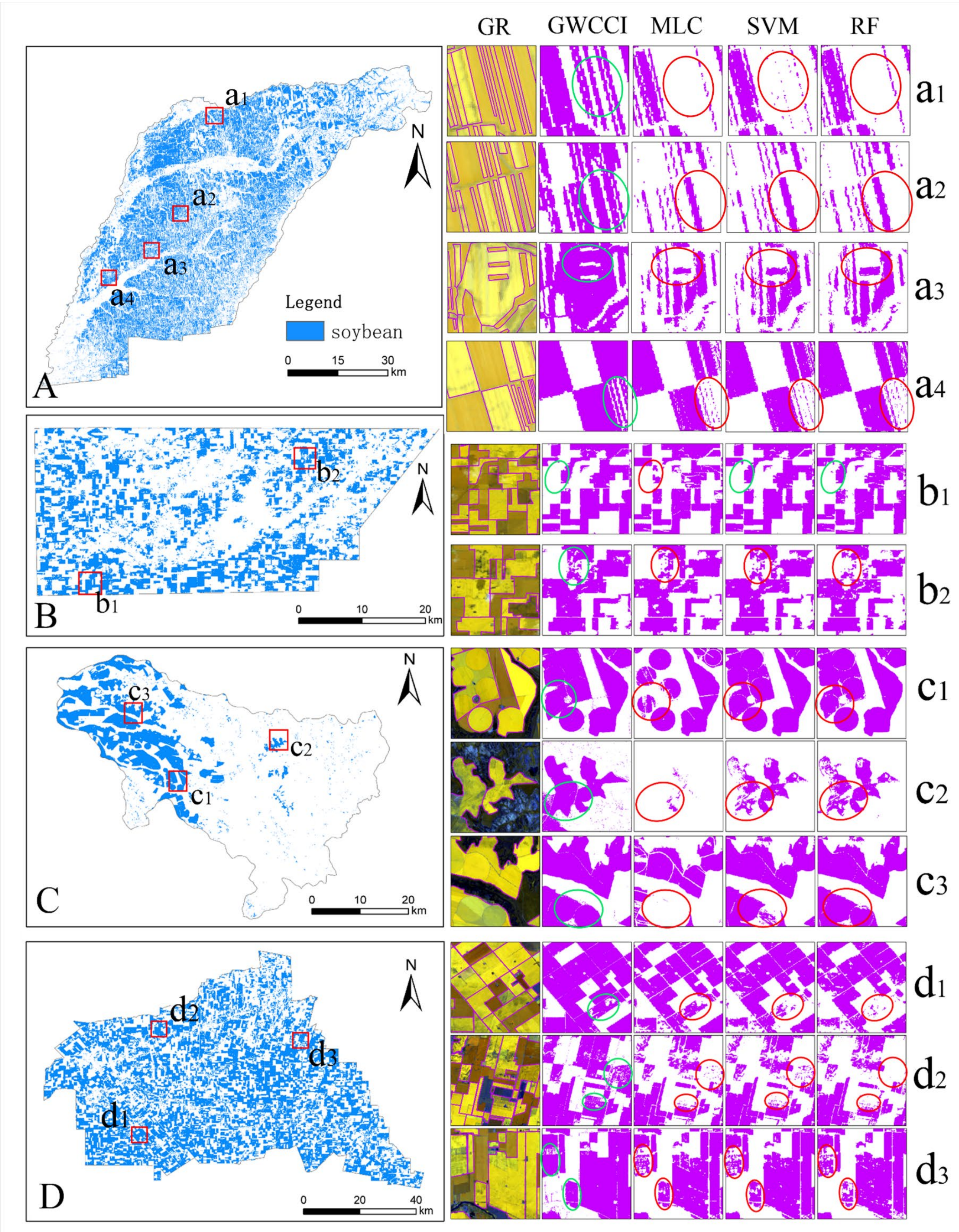

Fig. 10. Soybean maps (in the left panel) achieved by GWCCI for the Hailun (A), De Witt (B), Bonfinopolis (C) and Rio Segundo (D). Zooms of classification results (in the right panel) obtained by the four methods (from left to right: Ground reference (GR) soybean fields delineated using purple lines on the false colour S2 images (R: B7, G: B6, B: B3), GWCCI, MLC, SVM and RF).

4.3 Accuracy assessment

The accuracy of the soybean mapping based on the GWCCI was assessed quantitatively and compared with the MLC, SVM and RF classifiers, by using confusion matrices constructed with the testing data and the corresponding classification results (Table 4). The GWCCI achieved consistently the highest mapping accuracy in four study areas, with an average OA up to 96.29% (2.82%-to-7.23% higher than the benchmarks) and an average Kappa coefficient up to 0.92 (0.06-to-0.15 higher than the benchmarks). Amongst the three benchmarks, the accuracy of SVM ranked first, with an average OA of 93.47% and an average Kappa coefficient of 0.86, followed by the RF, and MLC delivered the least accuracy, with an average OA of 89.06% and an average Kappa coefficient of 0.77.

Table 1 Quantitative assessment of the classification accuracy of the GWCCI method and the three benchmarks, with the optimum result (in bold) of each column for each county.

Counties	Methods	OA (%)	Kappa (k)	PA (%)	UA (%)
Hailun	MLC	95.79	0.88	94.89	99.76
	SVM	96.84	0.91	96.19	99.81
	RF	96.36	0.90	95.57	**99.81**
	GWCCI	**97.80**	**0.93**	**99.71**	97.56
De Witt	MLC	81.40	0.63	63.50	98.91
	SVM	96.00	0.92	92.00	**99.90**
	RF	92.65	0.85	91.10	94.01
	GWCCI	**96.15**	**0.92**	**92.30**	99.00

Bonfinopolis	MLC	88.40	0.77	76.80	**99.90**
	SVM	91.20	0.82	82.50	99.88
	RF	94.75	0.90	90.70	98.69
	GWCCI	**96.65**	**0.93**	**99.00**	93.72
Rio Segundo	MLC	90.65	0.81	81.30	**99.90**
	SVM	89.85	0.80	79.70	99.00
	RF	86.55	0.73	73.20	99.86
	GWCCI	**94.55**	**0.89**	**94.00**	95.14

4.4 Robustness of the GWCCI approach

The robustness of the GWCCI approach over multiple years was investigated first. In addition to the year of 2021, the approach was also implemented from 2017 to 2020 over the three selected counties in the American continent, and the soybean accuracies are presented in Table 5. As illustrated by the table, the OA of the extended four years (2017-2020) is consistent with that of year 2021 for each of the three counties. That is, the GWCCI performed well, and relatively stable, across the five years. For example, the OA of the De Witt from 2017 to 2021 fluctuated around 96%, with a standard deviation (SD) of merely 0.73%. Similarly, consistently accurate and steady OAs (>93%) were also acquired in other two South Hemisphere counites over the study period, especially for the Bonfinopolis with only a difference of 1.10% between the greatest and the lowest OA. These desirable results demonstrate the robustness of the proposed approach over multiple years.

Table 5 OA achieved by GWCCI across the selected counties over the period 2017-2021. Numbers followed by a star (*) indicated that the corresponding soybean maps achieved by the GWCCI were based on the images dated within time window (i.e., PGS).

Counties	Overall accuracy (%)				
	2021	2020	2019	2018	2017
De Witt	96.15*	97.25*	96.90*	96.25*	95.10*
Bonfinopolis	96.65	\	95.55	96.45*	96.60*
Rio Segundo	94.55	93.30	93.25	94.45*	94.60*

Robustness of the GWCCI over time window (TW) was further investigated. Since optical remotely sensed images are vulnerable to cloud, shadow, haze and fog (Ashourloo et al., 2020), S2 images dated within the optimal TW (OTW) might not always be available, especially for the shorter OTW counties in Brazil (30 days) and Argentina (20 days). Rather, using alternative remotely sensed images, the actual time window (ATW) for the county of Brazil and Argentina was 58 and 30 days, respectively (Fig. 11). However, such a significant difference between ATW and OTW had no substantial impact on the classification accuracy, as shown in Table 5 (where an accuracy value marked by "*" was from OTW, the rest from ATW). For example, the difference in OA between the two periods 2017-2018 (using OTW marked by "*" (Table 5)) and 2019 to 2021 (using images beyond time window) for Rio Segundo is only 0.83%.

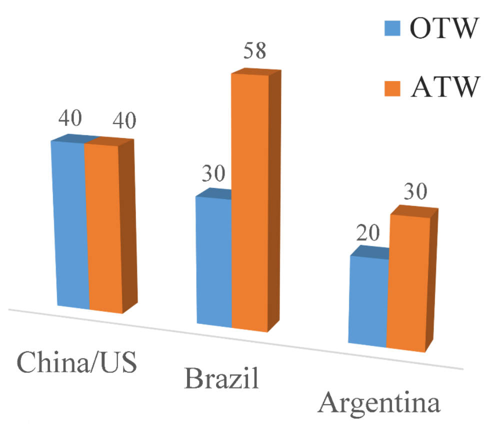

Fig. 11. The length (day) of Optimal time window (OTW) and actual time window (ATW) for China/US, Brazil, and Argentina.

5. Discussion

Timely and accurate mapping of soybean has long been an important, but complex challenge in remote sensing due to the considerable spectral overlap between soybean and other major crops (e.g., corn and rice). In this research, a novel crop index (named GWCCI) was proposed for mapping soybean using Sentinel-2 data. The GWCCI was constructed with the NDVI and SWIR bands, based on analysis of the spectral dynamics of crops. The proposed method was applied for soybean classification in four major soybean-producing countries in the world that differed in geography, soybean field size, phenology cropping system and crop management. Capitalizing on the feature dimensions of NDVI and SWIR, the GWCCI method was able to highlight the information differences between soybean and other land cover types, and provide a relatively long-time window for soybean information separation. The classification results indicate that the GWCCI method can effectively and accurately map soybean fields by using a single remotely sensed multispectral image, and it achieved the highest classification accuracy (with an average OA of 96.29% and k of 0.92) in comparison with the three traditional benchmark approaches (MLC, SVM and RF). The major advantages of the GWCCI method include the following two aspects: 1) The proposed concise, robust, and convenient GWCCI approach can be employed for mapping soybean automatically independent of training datasets, which saves expensive labor and computational resources that are required for traditional machine learning data training and sampling collection processes using time-series imagery. (2) The relatively long-time window for the computation of GWCCI may facilitate large-scale cloud-free data collection, given the fact that the soybean growing stage is commensurate with rainfall, promoting practical application of the GWCCI. Moreover, it may allow soybean detection ahead of time, which can be carried out as early as mid-July, up to three months before the soybean harvest, the time normally required by time-series based mapping approaches.

In this research, the GWCCI method incorporates the feature dimension benefits of the NDVI and SWIR bands by choosing the period that is most discriminative between soybean and other crops, that is, when soybean greenness and canopy water content are most different to those of other crops (e.g., corn). NDVI is used typically to represent the greenness of vegetation (He et al., 2021; Chen et al., 2021). During the growing stage, the greenness of soybean increases steadily with the gradual emergence of more lush foliage, and reaches a peak in the PGS. Meanwhile, the plant height of corn increases sharply while soybean grows more slowly relatively. The height difference between soybean and corn also peaks in the PGS with a significant difference in canopy water content, as captured by the SWIR band (Zhong et al., 2016a; Zhang et al., 2020). Therefore, both soybean and corn reach not only high greenness, but also high water content differences. The multiplication of NDVI and SWIR, as adopted by the proposed GWCCI, reflects comprehensively the spectral differences amongst the different land cover types, as illustrated by Table 6. For example, with both large values of NDVI and SWIR, soybean achieves a large GWCCI value, while corn has a medium GWCCI value due to the large NDVI, but small SWIR. Moreover, the construction of GWCCI depending on soybean's biophysical mechanism of its correlated development of the greenness and canopy water content provides a new strategy for detecting other similar vegetation from remotely sensed images.

Table 2 The effect of GWCCI as the product of NDVI and SWIR for different land cover types (H=High; M=Medium; L=Low). Soybean is the only crop with a high value of the GWCCI.

Land cover types	NDVI	SWIR	GWCCI
Soybean	H	H	**H**
Corn	H	L	M
Rice	H	L	M
Other crops	L	H	M
Woodland	H	L	M
Built-up area	L	H	M
Water	L	L	L

Spectral differences at certain phenological stages are used commonly to classify crops, as was done in the current research (Konduri et al., 2020; Ajadi et al., 2021). Diverse spectral bands carry a variety of spectral information, which can reflect the greenness, brightness, or water content (Liu et al., 2020; Li et al., 2021a) of crops, amongst others, such as to support classification. However, data redundancy may appear in some spectral bands, which can impact the classification accuracy negatively (Zhong et al., 2016a; Zhang et al., 2021). In this paper, a correlation analysis was conducted for the 11 variables utilized (10 spectral bands and NDVI) based on 1000 samples chosen randomly from the Sentinel-2 image (Table 7). It was found that the correlation coefficient between NDVI and SWIR was 0.4, which can be regarded as a weak correlation, and was not statistically significant (Sheugh and Alizadeh, 2015; Zhou et al., 2016). Meanwhile, SWIR produced a large correlation with the remaining nine bands, which were, thus, considered to contain some duplicate information statistically. The relative independence of each variable in the GWCCI (i.e., NDVI and SWIR) is critical for accurate soybean information extraction, as well as the rich (full band) spectral information contained inside the two variables of GWCCI.

Table 7 The correlation (r) of 11 variables from the imagery of Sentinel-2 in mid-July.

	Blue	Green	Red	Red Edge1	Red Edge2	Red Edge3	NIR	Red Edge4	SWIR1 (SWIR)	SWIR2	NDVI
Blue	1	0.90	0.86	0.88	0.85	0.81	0.81	0.81	0.86	0.85	0.17
Green	0.90	1	0.85	0.98	0.96	0.92	0.91	0.91	0.96	0.96	0.31
Red	0.86	0.85	1	0.81	0.77	0.71	0.71	0.71	0.78	0.78	-0.13
Red Edge1	0.88	0.98	0.81	1	0.98	0.93	0.92	0.92	0.99	0.99	0.37
Red Edge2	0.85	0.96	0.77	0.98	1	0.97	0.96	0.96	0.99	0.99	0.46
Red Edge3	0.81	0.92	0.71	0.93	0.97	1	0.99	1.00	0.94	0.92	0.58
NIR	0.81	0.91	0.71	0.92	0.96	0.99	1	0.99	0.92	0.91	0.60

Red Edge4	0.81	0.91	0.71	0.92	0.96	1.00	0.99	1	0.93	0.92	0.58
SWIR1 (SWIR)	0.86	0.96	0.78	0.99	0.99	0.94	0.92	0.93	1	0.99	0.40
SWIR2	0.85	0.96	0.78	0.99	0.99	0.92	0.91	0.92	0.99	1	0.38
NDVI	0.17	0.31	0.13	0.37	0.46	0.58	0.60	0.58	0.40	0.38	1

Note: Correlation is significant at the 0.05 level.

While contributing major advantages over traditional soybean mapping methods, some limitations still exist in the proposed GWCCI. First, the GWCCI was derived with representative agricultural landscape consisting of major crops (soybean, corn, rice) and some local crops, yet there is still room for improvement, especially for handling complex landscapes with greater crop diversity. Second, although the probability of achieving a cloud-free image mosaic for the GWCCI computation is extremely high with different dates of imagery within the GWCCI time window, it is still not guaranteed that it will be possible to acquire complete optical remotely sensed data covering a large region, which may lead to misclassification over cloud covered areas (Kontgis et al., 2015; Picoli et al., 2018; Zhang et al., 2020).

6. Conclusions

We developed a new vegetation index named GWCCI for the soybean mapping and classification, a goal that has been challenging to-date due to the considerable spectral overlap between soybean and other crops. The proposed GWCCI is a simple, reliable and cost-effective approach for mapping soybean. We evaluated extensively the proposed GWCCI across four counties distributed in four major soybean-producing countries in the world during the period from 2017 to 2021. When compared against three benchmark methods (MLC, SVM and RF) the GWCCI produced consistently the most accurate results in terms of the OA, Kappa index and Producer's Accuracy, demonstrating the wide applicability potential of the GWCCI across a variety of agricultural landscapes over multiple years. Moreover, in a further experiment, the GWCCI was shown to be fairly robust to choose of image acquisition date, thus, facilitating computation of the index from a time-series contaminated by clouds. The GWCCI, thus, can produce a high classification accuracy for in-season soybean classification while reducing costs, for example, avoiding the need for the collection of massive training datasets and preprocessing large time-series images. Given that soybean represents approximately 5% of all crops grown globally, the GWCCI has great potential for widespread application in operational settings, for example, as the basis for decision-making in support of economic production and to ensure local food security.

Acknowledgments

This research supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA28070500), the National Key Research and Development Program of China (2021YFD1500100), and the Capital Construction Fund of Jilin Province (2021C045-2).

References

Ajadi O A, Barr J, Liang S Z, et al. Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 97: 102294.

Al-Mamoori S K, Al-Maliki L A, Al-Sulttani A H, et al. Statistical analysis of the best GIS interpolation method for bearing capacity estimation in An-Najaf City, Iraq[J]. Environmental Earth Sciences, 2021, 80(20): 1-14.

Antonio C, Ovando G G, Díaz G J. ENSO influence on corn and soybean yields as a base of an early warning system for agriculture in Córdoba, Argentina[J]. European Journal of Agronomy, 2021, 129: 126340.

Ashiq W, Vasava H B, Ghimire U, et al. Topography controls N2O emissions differently during early and late corn growing season[J]. Agronomy, 2021, 11(1): 187.

Ashourloo D, Shahrabi H S, Azadbakht M, et al. Automatic canola mapping using time series of sentinel 2 images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 156: 63-76.

Ashourloo D, Shahrabi H S, Azadbakht M, et al. A novel method for automatic potato mapping using time series of Sentinel-2 images[J]. Computers and Electronics in Agriculture, 2020, 175: 105583.

Azadbakht M, Ashourloo D, Aghighi H, et al. Wheat leaf rust detection at canopy scale under different LAI levels using machine learning techniques[J]. Computers and Electronics in Agriculture, 2019, 156: 119-128.

Boryan C, Yang Z, Mueller R, et al. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program[J]. Geocarto International, 2011, 26(5): 341-358.

Boschetti M, Busetto L, Manfron G, et al. PhenoRice: A method for automatic extraction of spatio-temporal information on rice crops using satellite data time series[J]. Remote sensing of environment, 2017, 194: 347-365.

Cai Y, Guan K, Peng J, et al. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach[J]. Remote sensing of environment, 2018, 210: 35-47.

Calderón-Loor M, Hadjikakou M, Bryan B A. High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015[J]. Remote Sensing of Environment, 2021, 252: 112148.

Castro J B, Feitosa R Q, Happ P N. An hybrid recurrent convolutional neural network for crop type recognition based on multitemporal sar image sequences[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 3824-3827.

Chen C H , Tu T M . Computation reduction of the maximum likelihood classifier using the Winograd identity[J]. Pattern Recognition, 1996, 29(7):1213-1220.

Chen Z, Liu H, Xu C, et al. Modeling vegetation greenness and its climate sensitivity with deep‐learning technology[J]. Ecology and Evolution, 2021.

Colombo R, Meroni M, Marchesi A, et al. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling[J]. Remote sensing of environment, 2008, 112(4): 1820-1834.

Dado W T, Deines J M, Patel R, et al. High-resolution soybean yield mapping across the us midwest using subfield harvester data[J]. Remote Sensing, 2020, 12(21): 3471.

de Souza C H W, Mercante E, Johann J A, et al. Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices[J]. International Journal of Remote Sensing, 2015, 36(7): 1809-1824.

da Silva Junior C A, Leonel-Junior A H S, Rossi F S, et al. Mapping soybean planting area in midwest Brazil with remotely sensed images and phenology-based algorithm using the Google Earth Engine platform[J]. Computers and electronics in agriculture, 2020, 169: 105194.

Defourny P, Bontemps S, Bellemans N, et al. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world[J]. Remote sensing of environment, 2019, 221: 551-568.

Domínguez J A, Kumhálová J, Novák P. Winter oilseed rape and winter wheat growth prediction using remote sensing methods[J]. Plant, Soil and Environment, 2015, 61(9): 410-416.

Elsherif A, Gaulton R, Mills J. Estimation of vegetation water content at leaf and canopy level using dual-wavelength commercial terrestrial laser scanners[J]. Interface Focus, 2018, 8(2): 20170041.

Gómez C, White J C, Wulder M A. Optical remotely sensed time series data for land cover classification: A review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 116: 55-72.

Griffiths P, Nendel C, Hostert P. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping[J]. Remote sensing of environment, 2019, 220: 135-151.

Garnot V S F, Landrieu L, Giordano S, et al. Satellite image time series classification with pixel-set encoders and temporal self-attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12325-12334.

Haigh T, Takle E, Andresen J, et al. Mapping the decision points and climate information use of agricultural producers across the US Corn Belt[J]. Climate Risk Management, 2015, 7: 20-30.

He P, Sun Z, Han Z, et al. Dynamic characteristics and driving factors of vegetation greenness under changing environments in Xinjiang, China[J]. Environmental Science and Pollution Research, 2021: 1-17.

Inglada J, Arias M, Tardy B, et al. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery[J]. Remote Sensing, 2015, 7(9): 12356-12379.

Jacquemoud S, Verhoef W, Baret F, et al. PROSPECT+ SAIL models: A review of use for vegetation characterization[J]. Remote sensing of environment, 2009, 113: S56-S66.

Jia M, Wang Z, Wang C, et al. A new vegetation index to detect periodically submerged Mangrove forest using single-tide sentinel-2 imagery[J]. Remote Sensing, 2019, 11(17): 2043.

Johnson D M. An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States[J]. Remote Sensing of Environment, 2014, 141: 116-128.

Konduri V S, Kumar J, Hargrove W W, et al. Mapping crops within the growing season across the United States[J]. Remote Sensing of Environment, 2020, 251: 112048.

Kontgis C, Schneider A, Ozdogan M. Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data[J]. Remote Sensing of Environment, 2015, 169: 255-269.

Li H, Zhang C, Zhang S, et al. Full year crop monitoring and separability assessment with fully-polarimetric L-band UAVSAR: A case study in the Sacramento Valley, California[J]. International Journal of Applied Earth Observation and Geoinformation, 2019, 74: 45-56.

Li H, Zhang C, Zhang S, et al. Crop classification from full-year fully-polarimetric L-band UAVSAR time-series using the Random Forest algorithm[J]. International Journal of Applied Earth Observation and Geoinformation, 2020, 87: 102032.

Li X, Yu L, Peng D, et al. A large-scale, long time-series (1984–2020) of soybean mapping with phenological features: Heilongjiang Province as a test case[J]. International Journal of Remote Sensing, 2021a, 42(19): 7332-7356.

Li H, Zhang C, Zhang Y, et al. A Scale Sequence Object-based Convolutional Neural Network (SS-OCNN) for crop classification from fine spatial resolution remotely sensed imagery[J]. International Journal of Digital Earth, 2021b, 14(11): 1528-1546.

Li H, Zhang C, Zhang S, et al. Iterative Deep Learning (IDL) for agricultural landscape classification using fine spatial resolution remotely sensed imagery[J]. International Journal of Applied Earth Observation and Geoinformation, 2021c, 102: 102437.

Liu J, Feng Q, Gong J, et al. Winter wheat mapping using a random forest classifier combined with multi-temporal and multi-sensor data[J]. International Journal of Digital Earth, 2018, 11(8): 783-802.

Liu X, Yu L, Zhong L, et al. Spatial-temporal patterns of features selected using random forests: A case study of corn and soybeans mapping in the US[J]. International Journal of Remote Sensing, 2019, 40(1): 269-283.

Liu C, Zhang Q, Tao S, et al. A new framework to map fine resolution cropping intensity across the globe: Algorithm, validation, and implication[J]. Remote Sensing of Environment, 2020, 251: 112095.

Liu S, Chen Y, Ma Y, et al. Mapping Ratoon Rice Planting Area in Central China Using Sentinel-2 Time Stacks and The Phenology-Based Algorithm[J]. Remote Sensing, 2020, 12(20): 3400.

Liu Y, Zhi W, Xu B, et al. Detecting high-temperature anomalies from Sentinel-2 MSI images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 177: 174-193.

Lu Z, Qian S, Liu K, et al. Rice cultivation changes and its relationships with geographical factors in Heilongjiang Province, China[J]. Journal of integrative agriculture, 2017, 16(10): 2274-2282.

Marcos D, Volpi M, Kellenberger B, et al. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models[J]. ISPRS journal of photogrammetry and remote sensing, 2018, 145: 96-107.

Masuda T, Goldsmith P D. World soybean production: area harvested, yield, and long-term projections[J]. International food and agribusiness management review, 2009, 12(1030-2016-82753): 1-20.

Matton N, Canto G S, Waldner F, et al. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series[J]. Remote Sensing, 2015, 7(10): 13208-13232.

Nasrallah A, Baghdadi N, Mhawej M, et al. A novel approach for mapping wheat areas using high resolution Sentinel-2 images[J]. Sensors, 2018, 18(7): 2089.

Paul S, Saxena K G, Nagendra H, et al. Tracing land use and land cover change in peri-urban Delhi, India, over 1973–2017 period[J]. Environmental Monitoring and Assessment, 2021, 193(2): 1-12.

Peel M C, Finlayson B L, Mcmahon T A. Updated world map of the Köppen-Geiger climate classification[J]. Hydrology & Earth System Sciences, 2007, 11(3):259-263.

Picoli M C A, Camara G, Sanches I, et al. Big earth observation time series analysis for monitoring Brazilian agriculture[J]. ISPRS journal of photogrammetry and remote sensing, 2018, 145: 328-339.

Peel M C, Finlayson B L, Mcmahon T A. Updated world map of the Köppen-Geiger climate classification[J]. Hydrology & Earth System Sciences, 2007, 11(3):259-263.

Qiu B, Li W, Tang Z, et al. Mapping paddy rice areas based on vegetation phenology and surface moisture conditions[J]. Ecological Indicators, 2015, 56: 79-86.

Qu C, Li P, Zhang C. A spectral index for winter wheat mapping using multi-temporal Landsat NDVI data of key growth stages[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 175: 431-447.

Rußwurm M, Körner M. Self-attention for raw optical satellite time series classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 169: 421-435.

Radočaj D, Jurišić M, Gašparović M, et al. Optimal soybean (Glycine max L.) land suitability using GIS-based multicriteria analysis and Sentinel-2 multitemporal images[J]. Remote Sensing, 2020, 12(9): 1463.

Riccetto S, Davis A S, Guan K, et al. Integrated assessment of crop production and resource use efficiency indicators for the US Corn Belt[J]. Global Food Security, 2020, 24: 100339.

Sayago S, Ovando G, Bocco M. Landsat images and crop model for evaluating water stress of rainfed soybean[J]. Remote Sensing of Environment, 2017, 198: 30-39.

Schwalbert R A, Amado T, Corassa G, et al. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil[J]. Agricultural and Forest Meteorology, 2020, 284: 107886.

Sheugh L, Alizadeh S H. A note on pearson correlation coefficient as a metric of similarity in recommender system[C]//2015 AI & Robotics (IRANOPEN). IEEE, 2015: 1-6.

She B, Yang Y, Zhao Z, et al. Identification and mapping of soybean and maize crops based on Sentinel-2 data[J]. International Journal of Agricultural and Biological Engineering, 2020, 13(6): 171-182.

Shen Y, Zhang X, Yang Z. 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[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 186: 55-69.

Siyal, Altaf, Ali, et al. Rice yield estimation using Landsat ETM plus Data[J]. Journal of Applied Remote Sensing, 2015, 9:095986-1-095986-16.

Sonnentag O, Hufkens K, Teshera-Sterne C, et al. Digital repeat photography for phenological research in forest ecosystems[J]. Agricultural and Forest Meteorology, 2012, 152: 159-177.

Song X P, Potapov P V, Krylov A, et al. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey[J]. Remote sensing of environment, 2017, 190: 383-395.

Song X P, Hansen M C, Potapov P, et al. Massive soybean expansion in South America since 2000 and implications for conservation[J]. Nature sustainability, 2021a, 4(9): 784-792.

Song X P, Huang W, Hansen M C, et al. An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping[J]. Science of Remote Sensing, 2021b, 3: 100018.

Sulik J J, Long D S. Spectral considerations for modeling yield of canola[J]. Remote Sensing of Environment, 2016, 184: 161-174.

Teluguntla P, Thenkabail P S, Oliphant A, et al. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 144: 325-340.

Thenkabail P S, Wu Z. An automated cropland classification algorithm (ACCA) for Tajikistan by combining Landsat, MODIS, and secondary data[J]. Remote Sensing, 2012, 4(10): 2890-2918.

Tian H, Huang N, Niu Z, et al. Mapping winter crops in China with multi-source satellite imagery and phenology-based algorithm[J]. Remote sensing, 2019a, 11(7): 820.

Tian H, Meng M, Wu M, et al. Mapping spring canola and spring wheat using Radarsat-2 and Landsat-8 images with Google Earth Engine[J]. Curr. Sci, 2019b, 116(2): 291-298.

Turkoglu M O, D'Aronco S, Perich G, et al. Crop mapping from image time series: Deep learning with multi-scale label hierarchies[J]. Remote Sensing of Environment, 2021, 264: 112603.

Wang S, Azzari G, Lobell D B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques[J]. Remote sensing of environment, 2019, 222: 303-317.

Wang J, Xiao X, Liu L, et al. Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images[J]. Remote Sensing of Environment, 2020, 247: 111951.

Wilcox J R. World distribution and trade of soybean[J]. Soybeans: improvement, production, and uses, 2004, 16: 1-14-2.

Xiong J, Thenkabail P S, Tilton J C, et al. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine[J]. Remote Sensing, 2017, 9(10): 1065.

Xu J, Zhu Y, Zhong R, et al. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping[J]. Remote Sensing of Environment, 2020, 247: 111946.

Xu J, Yang J, Xiong X, et al. Towards interpreting multi-temporal deep learning models in crop mapping[J]. Remote Sensing of Environment, 2021, 264: 112599.

Yilmaz M T, Hunt Jr E R, Jackson T J. Remote sensing of vegetation water content from equivalent water thickness using satellite imagery[J]. Remote Sensing of Environment, 2008, 112(5): 2514-2522.

Zhang X, Jayavelu S, Liu L, et al. Evaluation of land surface phenology from VIIRS data using time series of PhenoCam imagery[J]. Agricultural and Forest Meteorology, 2018a, 256: 137-149.

Zhang C, Pan X, Li H, et al. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018b, 140: 133-144.

Zhang F, Zhou G. Estimation of vegetation water content using hyperspectral vegetation indices: A comparison of crop water indicators in response to water stress treatments for summer maize[J]. BMC ecology, 2019a, 19(1): 1-12.

Zhang C, Di L, Lin L, et al. Machine-learned prediction of annual crop planting in the US Corn Belt based on historical crop planting maps[J]. Computers and Electronics in Agriculture, 2019b, 166: 104989.

Zhang H, Kang J, Xu X, et al. Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi’an County, Heilongjiang province, China[J]. Computers and Electronics in Agriculture, 2020, 176: 105618.

Zhang H, Du H, Zhang C, et al. An automated early-season method to map winter wheat using time-series Sentinel-2 data: A case study of Shandong, China[J]. Computers and Electronics in Agriculture, 2021, 182: 105962.

Zhang P, Du P, Guo S, et al. A novel index for robust and large-scale mapping of plastic greenhouse from Sentinel-2 images[J]. Remote Sensing of Environment, 2022, 276: 113042.

Zhang C, Dong J, Ge Q. IrriMap_CN: Annual irrigation maps across China in 2000–2019 based on satellite observations, environmental variables, and machine learning[J]. Remote Sensing of Environment, 2022, 280: 113184.

Zhong L, Gong P, Biging G S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery[J]. Remote Sensing of Environment, 2014, 140: 1-13.

Zhong L, Hu L, Yu L, et al. Automated mapping of soybean and corn using phenology[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016a, 119: 151-164.

Zhong L, Yu L, Li X, et al. Rapid corn and soybean mapping in US Corn Belt and neighboring areas[J]. Scientific reports, 2016b, 6(1): 1-14.

Zhong L, Hu L, Zhou H, et al. Deep learning based winter wheat mapping using statistical data as ground references in Kansas and northern Texas, US[J]. Remote Sensing of Environment, 2019, 233: 111411.

Zhou H, Deng Z, Xia Y, et al. A new sampling method in particle filter based on Pearson correlation coefficient[J]. Neurocomputing, 2016, 216: 208-215.

Zhou Y, Luo J, Feng L, et al. Long-short-term-memory-based crop classification using high-resolution optical images and multi-temporal SAR data[J]. GIScience & Remote Sensing, 2019, 56(8): 1170-1191.