

## Original papers

## A leaf chlorophyll vegetation index with reduced LAI effect based on Sentinel-2 multispectral red-edge information

Yuanheng Sun<sup>a,\*</sup>, Qiming Qin<sup>b</sup>, Yao Zhang<sup>c</sup>, Huazhong Ren<sup>b</sup>, Guhuai Han<sup>b</sup>, Zhaoxu Zhang<sup>d</sup>, Tianyuan Zhang<sup>b</sup>, Binyu Wang<sup>a</sup>

<sup>a</sup> Environmental Information Institute, Navigation College, Dalian Maritime University, Dalian, China

<sup>b</sup> School of Earth and Space Sciences, Peking University, Beijing, China

<sup>c</sup> Key Laboratory of Smart Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing, China

<sup>d</sup> School of Environmental Science and Engineering, Tiangong University, Tianjin, China



## ARTICLE INFO

## Keywords:

Leaf chlorophyll content

Red-edge

Sentinel-2

Multispectral remote sensing

Vegetation index

## ABSTRACT

As a pivotal physiological trait influencing a plant's photosynthetic capacity, accurate and efficient characterizing of leaf chlorophyll content (LCC) is crucial for understanding terrestrial ecosystem carbon cycling. The red-edge reflectance spectrum, obtained through remote sensing, offers valuable insights for LCC estimation. The red-edge position (REP) in the vegetation spectral reflectance is a potent proxy for LCC due to its relationship with chlorophyll absorption. However, variations in leaf area index (LAI) can influence the REP and impact its performance in LCC estimation. In this study, we propose the Sentinel-2 leaf chlorophyll index (S2LCI), a novel vegetation index based on REP and a LAI indicator using Sentinel-2 multispectral red-edge bands. This innovative index significantly mitigates the influence of LAI variation and enhances the LCC. The effectiveness of S2LCI has been evaluated through multiple ways, including PROSAIL simulated datasets, ground-measured LCC with canopy spectra, and Sentinel-2 imagery. Our results demonstrate strong agreement between S2LCI and LCC in both simulated and ground-measured datasets ( $R = 0.492$  for ground spectra and  $R = 0.526$  for Sentinel-2 imagery), outperforming classic LCC-related vegetation indices. Furthermore, S2LCI mapping reveals finer spatial details compared to LCC derived using the Sentinel Application Platform (SNAP) biophysical processor. This study highlights the suitability of S2LCI for LCC estimation and offers a promising solution, complementing other LCC retrieval approaches to rapidly generate decameter-scale crop LCC maps using Sentinel-2 imagery.

### 1. Introduction

Vegetation serves as a paramount component in the exchange of carbon, water and energy between the terrestrial ecosystems and the atmosphere, and leaf chlorophyll is the center to this vital process (Green et al., 2017; Xiao et al., 2021). The concentration of leaf chlorophyll content (LCC) is intrinsically linked to a plant's physiological status and photosynthetic capacity (Chen et al., 2022; Croft et al., 2017). Therefore, capturing the spatial and temporal variation in LCC offers crucial insights into global and regional vegetation health and carbon cycling dynamics (Croft et al., 2020).

Traditional methods for LCC measurement typically involve destructive sampling of foliage or non-destructive measurements using contact SPAD chlorophyll meter (Markwell et al., 1995; Uddling et al., 2007). These approaches are not only labor-intensive but also time-

consuming (Houborg and Boegh, 2008). An increasingly promising alternative lies in the utilization of earth observation techniques, where optical remote sensing plays a pivotal role in acquiring large-scale vegetation-related variables. It does so by establishing relationships between sensor-observed signals and the physicochemical parameters of vegetation (Gitelson et al., 2014). Canopy spectral reflectance is influenced by numerous factors, including biochemical parameters (e.g., chlorophyll content, water content), biophysical parameters (e.g., leaf area index (LAI), leaf angle distribution (LAD)), and external factors such as atmosphere conditions, soil background, and illumination and imaging geometry (Berger et al., 2018a; Jacquemoud et al., 2009). Despite its significance, the contribution of LCC to canopy reflectance signals is relatively smaller in comparison to other parameters (Gu et al., 2016; Sun et al., 2022a; Verrelst et al., 2019). Furthermore, the impacts of LCC and other parameters are often intertwined, necessitating a

\* Corresponding author at: Environmental Information Institute, Navigation College, Dalian Maritime University, Dalian 116026, China.  
E-mail address: [yhsun@dlnu.edu.cn](mailto:yhsun@dlnu.edu.cn) (Y. Sun).

means to disentangle the individual influence of LCC within sensor-measured canopy reflectance, which is an essential and challenging task.

The aforementioned challenges amplify the complexity and hinder the precision of LCC estimation using remote sensing data. Hyperspectral images offer a nuanced portrayal of various vegetation biochemical and physiological traits (Sytar et al., 2017; Zhang et al., 2021). Consequently, a multitude of methods has emerged, leveraging hyperspectral remote sensing data for LCC estimation. These methods encompass hyperspectral vegetation index (VI) (Moharana and Dutta, 2016; Wu et al., 2008), spectral feature extraction (Noda, 2017; Zhang et al., 2018), physical model inversion (Berger et al., 2018a; Verrelst et al., 2015), and machine learning or deep learning algorithms (Heckmann et al., 2017; Sonobe et al., 2020; Zhang et al., 2021).

The spectral characteristics of vegetation, particularly within the red-edge, have demonstrated a high sensitivity to variations in LCC (Heckmann et al., 2017; Wu et al., 2008). The red-edge, a distinctive spectral feature located between the red absorption maximum and high reflectance in the near-infrared (NIR) (Curran et al., 1991; Horler et al., 1983), has drawn significant attention. Among the various red-edge spectral features, the red-edge position (REP) has emerged as a valuable indicator for LCC estimation. As the LCC increases, the REP undergoes a substantial shift towards longer wavelengths (Curran et al., 1991; Horler et al., 1983). Quantification of the REP with hyperspectral data is typically achieved by determining the point of maximum slope along the red-edge spectral interval.

While hyperspectral imagery excels in its ability to capture extensive information for leaf-level biochemical parameter retrieval, it is not without its limitations, which constrain its capacity for large-scale LCC mapping. These drawbacks encompass constrained spatial coverage, infrequent temporal acquisition, elevated costs, and the need for a more intricate and time-consuming process due to the wealth of redundant spectral band (Gu et al., 2021; Lu et al., 2020). In contrast, multispectral imagery is deliberately designed to encompass specific wavelengths pertinent to the absorption characteristics of vegetation, enabling a more precise and computationally efficient approach to LCC estimation. In essence, multispectral data strikes a harmonious balance between spectral informativeness and operational practicality, rendering it a pragmatic choice for a multitude of applications.

One crucial asset within the Copernicus program of the European Space Agency (ESA) is the Sentinel-2 satellite constellation. These satellites furnish imagery with three 20-m resolution multispectral narrow red-edge bands at a 5-day revisit interval, presenting an invaluable resource for vegetation monitoring (Drusch et al., 2012). By leveraging its capability to capture narrow-band data across the visible to short-wave infrared spectra, global coverage, and an open data policy, Sentinel-2 has contributed immense value to a diverse array of applications, including precision agriculture (Sun et al., 2022b), forest monitoring (Waldeland et al., 2022), land cover mapping (Phiri et al., 2020), and ecosystem assessment (del Río-Mena et al., 2020).

Amidst the advent of narrow-band multispectral red-edge sensors like Sentinel-2/MSI, a wealth of vegetation indices (VIs) have emerged, capitalizing on the detailed red-edge spectral information they capture, with representative examples including the red-edge chlorophyll index (Cire) (Gitelson et al., 2005) and the red-edge normalized difference vegetation index (NDVIre) (Gitelson and Merzlyak, 1994). Additionally, the Medium Resolution Imaging Spectrometer (MERIS) terrestrial chlorophyll index (MTCI) stands out as the pioneering red-edge chlorophyll index meticulously tailored for a specific multispectral red-edge sensor, and it has been operationally implemented as a standard level 2 global product derived from the ENVISAT MERIS dataset (Dash and Curran, 2004).

In contrast to previous multispectral satellite data, which often featured just a single red-edge band, as exemplified by RapidEye or WorldView, the Sentinel-2 has ushered in the potential to discern shifts in REP with its three narrow red-edge bands. Nonetheless, those methods for REP determination specifically designed for hyperspectral

data may not be directly applicable to Sentinel-2 data (Ollinger, 2011). To tackle this challenge, the Sentinel-2 red-edge position (S2REP), as proposed by Frampton et al. (2013), employs a linear interpolation technique with the advantage of a limited number of spectral bands. It estimates the REP using the first two Sentinel-2 red-edge bands, centered at 705 nm and 740 nm, positioned on the red-edge slope. The S2REP serves as an excellent surrogate for REP in the context of Sentinel-2 data, characterized by its insensitivity to variations in background soil. Nevertheless, the impact of LAI remains substantial. Consequently, it becomes imperative to mitigate the LAI effect and enhance the LCC signal to bolster the accuracy and stability of LCC retrieval, especially in complex scenes.

The primary objective of this study is to formulate a vegetation index that exhibits sensitivity to LCC while remaining resilient to the influence of LAI using the REP in conjunction with Sentinel-2's multispectral red-edge information. Our approach commences with an exploration of vegetation red-edge spectra and their defining characteristics. Subsequently, we introduce the Sentinel-2 leaf chlorophyll index (S2LCI), which hinges on the S2REP and a LAI indicator. The proposed S2LCI is then subjected to comprehensive evaluation and benchmarking against other prominent multispectral chlorophyll vegetation indices, utilizing both simulated and ground-measured datasets. Finally, we engage in a discourse regarding the suitability, advantages, sources of uncertainty, and potential applications of the S2LCI.

## 2. Material

### 2.1. PROSAIL simulations

To explore the spectral characteristics of vegetation and assess the performance of the S2LCI, we employed the PROSAIL model, a well-established tool for simulating canopy reflectance under various conditions. This model integrates the PROSPECT-D leaf optical properties model (Féret et al., 2017) and the Scattering by Arbitrarily Inclined Leaves (SAIL) canopy bidirectional reflectance model (Verhoef et al., 2007). The PROSAIL model is widely recognized for its capability to simulate canopy spectra and conduct sensitivity analyses (Jacquemoud et al., 2009). To comprehensively understand the characteristics of vegetation spectra and Sentinel-2 band reflectance, we conducted a systematic exploration. We selectively varied specific parameter, such as LCC, while keeping other variables constant. The parameter settings used for these simulations are provided in the captions of Fig. 2 and Fig. 3, both of which are presented in Section 4.1. In order to ensure the size and representativeness of the synthetic dataset used for evaluating the S2LCI, we generated a total of 20,000 spectra by randomly generating and combining input parameters within a wide range. This rigorous approach allowed us to capture diverse variations and interactions among the parameters, contributing to a robust and reliable assessment of the S2LCI. Truncated Gaussian or uniform distribution were employed to generate the input variables of the PROSAIL model, and Table 1 provides a comprehensive list of values and distributions for input parameters used in generating the simulated spectral dataset. These parameters are based on those relevant to wheat, the primary vegetation type for our field LCC measurements (Berger et al., 2018b; Zhang et al., 2021). The PROSAIL simulated canopy spectral reflectance was subsequently convolved with the Sentinel-2/MSI spectral filters to align it with the band reflectance.

### 2.2. Ground canopy spectra dataset and LCC field observations

To validate the performance of the S2LCI, we employed *in-situ* datasets collected from two distinct geographical regions: Yangling Agricultural Hi-tech Industries Demonstration Zone and Luohe Experimental Site (as depicted in Fig. 1). Yangling Agricultural Hi-tech Industries Demonstration Zone is geographically situated at 34.3°N, 108.1°E and is located within the Guanzhong Plain of China. This region

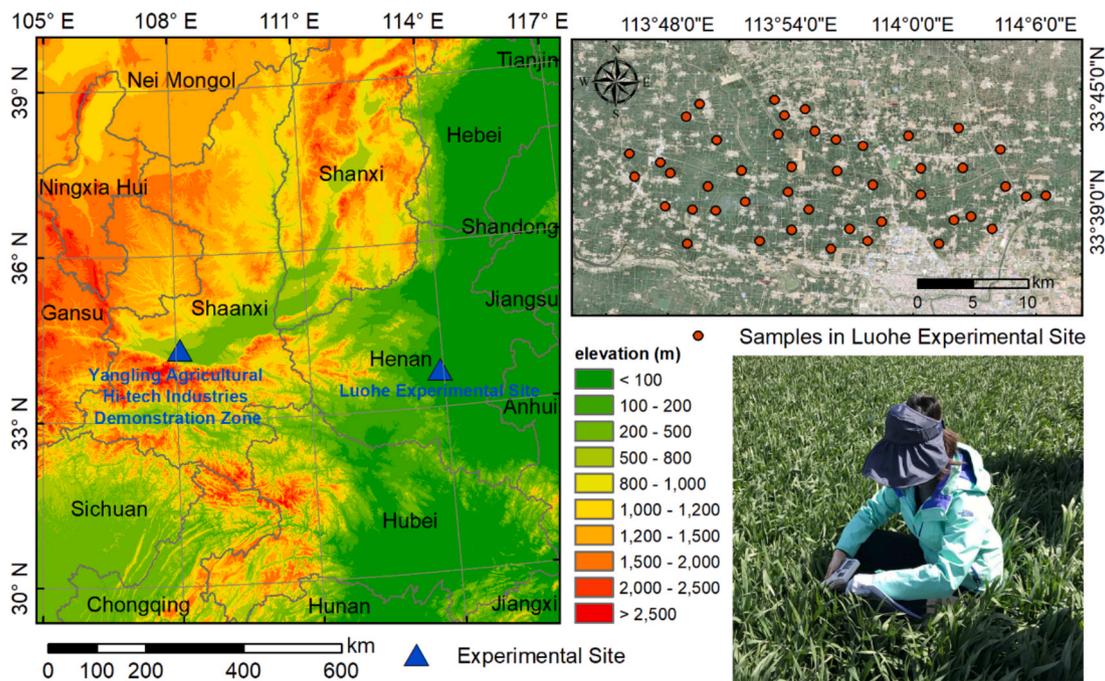


Fig. 1. Geographical locations of experimental sites and sampling for ground canopy spectra and LCC measurements. The upper-right panel shows a true-color composite derived from Sentinel-2 imagery captured on the dates coinciding with the fieldwork in Luohe Experimental Site. The lower-right panel displays a photograph of LCC measurement using a SPAD-502 chlorophyll meter on March 8, 2018.

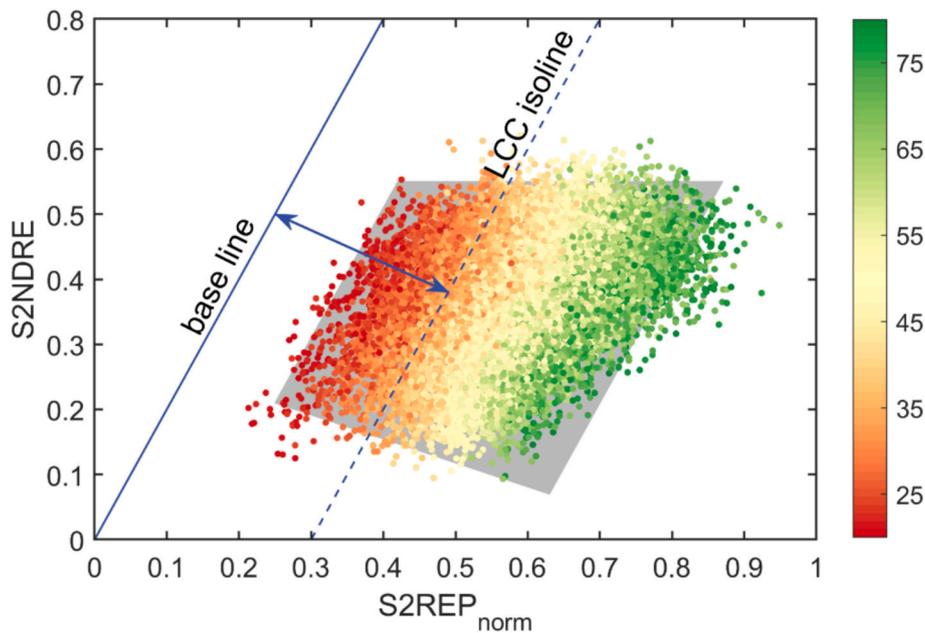
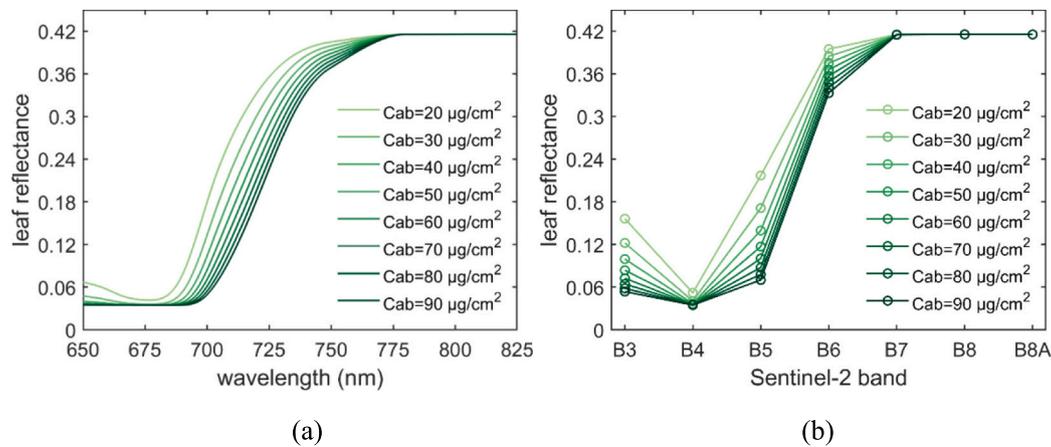


Fig. 2. Conceptual diagram of Sentinel-2 leaf chlorophyll index (S2LCI). The  $S2REP_{norm}$  and  $S2NDRE$  scatters roughly delineate a trapezoid (gray region) in the 2-D feature space, with the color of the scatters indicating LCC values. The LCC isoline is nearly parallel to the bottom edges of the trapezoid, and the distance between the baseline and a specific point is defined as the S2LCI, representing the relative leaf chlorophyll content of vegetation canopy.

exhibits typical continental monsoon climate characteristic, with an annual average temperature of approximately 12.9°C and an average precipitation of around 635 mm. In contrast, the Luohe Experimental Site is positioned at 33.5°N, 114.1°E and experiences a warm temperate monsoon characterized by wet summer and dry winter. The yearly average temperature in this area is approximately 14.6°C, and the average annual precipitation amounts to 805 mm. Both of these regions primarily cultivate winter wheat (*Triticum aestivum* L.) and summer maize (*Zea mays* L.), managed under rotation irrigated practices. Winter

wheat is sown in early October and harvested in early June (Yangling) or late May (Luohe) the following year.

For our study, data collection focused on the jointing stage of wheat in Yangling, with field measurements conducted on March 30, 2013, and March 28, 2014. In Luohe, our field campaign spanned from the tilling stage to the jointing stage of wheat and took place on March 8, 2018, and April 6, 2018. In Yangling, canopy spectra measurements were recorded, while LCC data were collected in both Yangling and Luohe. The canopy spectra and LCC data from Yangling were utilized for ground-scale



**Fig. 3.** Leaf reflectance spectra in the 650–825 nm interval (a) and Sentinel-2 band reflectance (b) simulated by the PROSPECT-D model (Féret et al., 2017). Simulations were performed over a range of leaf chlorophyll content (Cab) from 20  $\mu\text{g cm}^{-2}$  to 90  $\mu\text{g cm}^{-2}$  at a step of 10  $\mu\text{g cm}^{-2}$ . The carotenoid content was set to 10  $\mu\text{g cm}^{-2}$ , the anthocyanin content was 1  $\mu\text{g cm}^{-2}$ , the leaf structure parameter was 1.2, the dry matter content was 0.002  $\text{g cm}^{-2}$  and the equivalent water thickness was 0.005 cm.

**Table 1**  
Parameter settings for PROSAIL model used for canopy reflectance modeling.

Type	Parameters	Values or distributions	Minimum value	Maximum value
Leaf	Leaf chlorophyll content (Cab, $\mu\text{g cm}^{-2}$ )	Gauss(50, 15)	20	80
	Dry matter content (Cm, $\text{g cm}^{-2}$ )	Gauss(0.007, 0.002)	0.003	0.011
	Leaf structure parameter (N)	Gauss(1.5, 0.5)	1	2
	Leaf carotenoid content (Ccar, $\mu\text{g cm}^{-2}$ )	10	–	–
	Brown pigment content (Cbrown, $\mu\text{g cm}^{-2}$ )	0	–	–
	Leaf anthocyanin content (Cant, $\mu\text{g cm}^{-2}$ )	1	–	–
	Equivalent water thickness (Cw, cm)	0.005	–	–
	Canopy	Leaf area index (LAI, $\text{m}^2\text{m}^{-2}$ )	Uniform(1, 6)	1
	Average leaf angle (ALA, degree)	Gauss(50, 10)	30	70
Background	Soil moisture factor ( $\rho_{\text{soil}}$ )	Uniform(0, 1)	0	1
Sun-target-sensor geometry	Solar zenith angle (SZA, degree)	Gauss(30, 10)	0	60
	View zenith angle (VZA, degree)	Gauss(10, 5)	0	20
	Relative azimuth angle (RAA, degree)	0	–	–

Note: The numbers in parenthesis for Gauss distribution represent mean value and standard deviation, and they represent lower bound and upper bound for Uniform distribution.

validation, whereas the data from Luohe were employed to validate LCC results derived from Sentinel-2 imagery. Sample points in Luohe were carefully situated within homogeneous croplands, each having a minimum spatial extent of  $40 \times 40$  m and a sample size of  $20 \times 20$  m. The geographic coordinates of each sample point were meticulously recorded using a portable GPS device, with coordinates formatted in WGS-1984 longitude and latitude. In total, 118 valid samples were collected in Yangling, and an additional 51 samples were gathered in Luohe for the validation process.

Hyperspectral reflectance data from the canopy was acquired using an ASD FieldSpec Pro spectroradiometer (Analytical Spectral Devices, Inc., Boulder, CO, USA). This state-of-the-art instrument boasts an extensive spectral range from 350 to 2,500 nm, with a resampled interval of 1 nm. The spectroradiometer features a fiber optic detector with a field of view of  $25^\circ$ , and it was meticulously positioned at a height of approximately 1 m directly above the wheat canopy. In order to mitigate the potential influence of varying environmental conditions, five distinct positions within each sample point were randomly selected for hyperspectral measurement. The resultant canopy reflectance spectra were calculated by dividing the canopy’s spectrum by that of a corresponding standard white panel. To obtain a representative spectral reflectance for each sample point, the average value of the five reflectance measurements was employed. Subsequently, the measured canopy spectral reflectance was harmoniously integrated into Sentinel-2 band reflectance, utilizing the spectral response functions specific to the Sentinel-2 sensor.

To ascertain LCC, the SPAD-502 chlorophyll meter (Konica Minolta, Inc.) was utilized. In a given sample point, five positions were randomly selected for LCC measurements. To enhance accuracy, four replicate measurements were conducted at varying levels within the wheat canopy—upper, middle, and lower. This approach generated a total of twelve readings, which were then averaged to yield the representative SPAD value for that particular position. The average of the five SPAD readings within each sample point was taken as the LCC proxy for that point. It is important to note that the SPAD values were employed directly as the ground proxy for LCC, as no local calibration for the SPAD-to-LCC conversion relationship was conducted following the initial measurements.

### 2.3. Sentinel-2 images

The Sentinel-2 images utilized in this study coincide with the timing of the Luohe field campaign, captured on March 9 and April 8, 2018. These images provided spatial coverage over the Luohe Experimental

Sites. Our data source for these images was the ESA Sentinels Scientific Data Hub. We obtained these images in the form of Level-1C orthorectified top of atmosphere (TOA) reflectance data. Subsequently, an atmospheric correction process was applied to convert them into top of canopy (TOC) reflectance data, using the Sen2Cor atmosphere correction toolbox (version 2.5.5) which is embedded in the Sentinel Application Platform (SNAP) software (version 6.0.0). To ensure the preservation of the red-edge bands in the atmospherically corrected images, the spatial resolutions of all bands were uniformly set to 20 m, enabling the simultaneous use of the green, red and NIR bands (Drusch et al., 2012).

### 3. Methods

#### 3.1. A leaf chlorophyll index based on Sentinel-2 red-edge position and LAI indicators

Since the REP is influenced by both leaf chlorophyll content and leaf area in most scenario, the necessity arises to disentangle and eliminate the influence of LAI on REP, as our primary focus is the accurate representation and retrieval of leaf chlorophyll content using REP. In this study, we introduce a novel vegetation index for the estimation of leaf chlorophyll content by leveraging the multispectral red-edge reflectance data from Sentinel-2. Table 2 provides an overview of the spectral band configuration of Sentinel-2/MSI. Among these bands, Band 5 and Band 6 of Sentinel-2/MSI play a crucial role in characterizing the distinctive increase in reflectance at the red-edge, offering finer spectral details compared to earlier multispectral satellite data, which typically feature only a single red-edge band. However, it is important to note that Band 7 of Sentinel-2 falls outside the spectral range associated with the sharp red-edge increase, functioning more like a NIR band, despite being labeled as a red-edge band by ESA (Sun et al., 2020).

In this study, we adopt the Sentinel-2 red-edge position (S2REP) as proposed by Frampton et al. (2013) to serve as our proxy of the REP. S2REP is determined using a linear interpolation technique, where the reflectance at the inflexion point is estimated, thus retrieving the REP through interpolation of the Sentinel-2's first two red-edge bands (Band 5 and Band 6), as outlined in Eq. (1). It has been widely recognized as an effective alternative to REP and an excellent indicator for quantifying LCC in numerous studies (Qian et al., 2022; Xie et al., 2019).

$$S2REP = 705 + 35 \times \frac{(B7 + B4)/2 - B5}{B6 - B5} \quad (1)$$

The proposed Sentinel-2 leaf chlorophyll index (S2LCI) is constructed using the S2REP and a LAI indicator, aiming to minimize the impact of LAI on REP data. We employ a simplified version of S2REP (Eq. (2)), with values typically confined between 0 and 1 for most vegetation conditions, which is subsequently combined with the LAI indicator for a unified scale. In this case, we utilize the Sentinel-2 normalized difference red-edge vegetation index (S2NDRE) as the LAI

indicator (Eq. (3)), a metric that exhibits a strong positive correlation with LAI while being minimally affected by LCC (Sun et al., 2023).

$$S2REP_{norm} = \frac{(B7 + B4)/2 - B5}{B6 - B5} \quad (2)$$

$$S2NDRE = \frac{B6 - B4}{B6 + B4} \times B7 \quad (3)$$

The S2REP<sub>norm</sub> and S2NDRE, when combined in the S2REP<sub>norm</sub>/S2NDRE 2-dimensional feature space, offer valuable information about vegetation LCC and LAI. While various factors influence the precise location of a vegetation pixel within this feature space, LCC and LAI are the primary determinants. A trapezoid shape (gray area in Fig. 2) representing different combinations of input parameters from the PROSAIL simulated dataset is evident in the S2REP<sub>norm</sub>/S2NDRE space. Ideally, if S2REP<sub>norm</sub> is entirely and exclusively influenced by LCC, vegetated pixels with the same LCC should have an equal distance between them and the Y-axis. However, the LCC isoline does not run perfectly vertical to the X-axis, indicating that factors other than LCC systematically impact S2REP<sub>norm</sub>, with LAI being the most influential among these factors. Conversely, the S2NDRE is nearly exclusively influenced by LAI and exhibits a near-vertical LAI isoline to the Y-axis (Fig. S1).

To explain the process of calculating S2LCI, we first establish a baseline, which runs parallel to the LCC isoline and crosses the origin, in the S2REP<sub>norm</sub>/S2NDRE space (Fig. 2). The slope of this baseline, denoted as 'k', is primarily determined by the type of vegetation. By default, we set 'k' to 2.0 based on extensive PROSAIL simulations of wheat corresponding to ground LCC measurement (Section 2.2). The distance between the baseline and the pixel, herein defined as the S2LCI (Eq. (4)), serves as a measure of the relative leaf chlorophyll content of the vegetation canopy (Fig. 2).

$$S2LCI = \frac{k \times S2REP_{norm} - S2NDRE}{\sqrt{k^2 + 1}} \quad (4)$$

#### 3.2. Compared vegetation indices and evaluation indicators

We selected a range of representative and state-of-the-art multispectral vegetation indices, many of which were initially developed for the purpose of retrieving leaf chlorophyll content, for comparison with the S2LCI. Detailed formulations of these indices, utilizing Sentinel-2 bands, are provided in Table 3.

To assess the performance of the proposed S2LCI and the aforementioned vegetation indices for LCC retrieval, we employed several key indicators. The coefficient of determination ( $R^2$ ) and root-mean-square error (RMSE) were utilized to assess the effectiveness of regression models for estimating LCC using various vegetation indices, drawing from the simulated dataset. In addition, the coefficient of correlation ( $R$ ) was employed to gauge the goodness-of-fit between these vegetation indices and the ground-measured SPAD values. Our approach involved a five-fold cross-calibration methodology, in which the dataset was randomly divided into five equally sized, mutually exclusive groups. Four groups were designated as the calibration dataset, while the remaining group served as the validation dataset. This five-fold cross-calibration process was repeated to ensure each group had an opportunity to serve as both the calibration and validation datasets. This approach, which avoids relying on a single random partition for calibration and validation, bolsters the robustness and reliability of our calibration and validation procedures. Additionally, to investigate the accuracy of VI-based LCC retrievals across various LAI levels, we employed a bias calculation, representing the difference between the estimated LCC based on a VI and the model-input LCC.

**Table 2**  
Spectral band setting of Sentinel-2/MSI.

Band	Central wavelength (nm)	Band width (nm)	Spatial resolution (m)
B1	443	20	60
B2	490	65	10
B3	560	35	10
B4	665	30	10
B5	705	15	20
B6	740	15	20
B7	783	20	20
B8	842	115	10
B8A	865	20	20
B9	945	20	60
B10	1375	30	60
B11	1610	90	20
B12	2190	180	20

**Table 3**

Leaf chlorophyll-related vegetation indices used for comparison in this research, along with the corresponding Sentinel-2/MSI band configuration.

Vegetation index	Formulation	reference
NDVI	$\frac{B8A - B4}{B8A + B4}$	(Rouse et al., 1974)
Red-edge NDVI 1 (NDRE1)	$\frac{B6 - B5}{B6 + B5}$	(Gitelson and Merzlyak, 1994)
Red-edge NDVI 2 (NDRE2)	$\frac{B8A - B5}{B8A + B5}$	(Gitelson and Merzlyak, 1997)
Modified chlorophyll absorption ratio index (MCARI)	$[(B5 - B4) - 0.2 \times (B5 - B3)] \times \frac{B5}{B4}$	(Daughtry et al., 2000)
Transformed chlorophyll absorption ratio index/optimized soil-adjusted vegetation index (TCARI/OSAVI)	$\frac{3 \times [(B5 - B4) - 0.2 \times (B5 - B3)] \times B5/B4}{(1 + 0.16) \times (B8A - B4)/(B8A + B4 + 0.16)}$	(Haboudane et al., 2002)
MERIS terrestrial chlorophyll index (MTCI)	$\frac{B6 - B5}{B5 - B4}$	(Dash and Curran, 2004)
Red-edge chlorophyll index (Clre)	$\frac{B8A - 1}{B5}$	(Gitelson et al., 2005)
Modified chlorophyll absorption ratio index /optimized soil-adjusted vegetation index (MCARI/OSAVI [705, 750])	$\frac{[(B6 - B5) - 0.2 \times (B6 - B3)] \times B6/B5}{(1 + 0.16) \times (B6 - B5)/(B6 + B5 + 0.16)}$	(Wu et al., 2008)
Transformed chlorophyll absorption ratio index/optimized soil-adjusted vegetation index (TCARI/OSAVI [705, 750])	$\frac{3 \times [(B6 - B5) - 0.2 \times (B6 - B3)] \times B6/B5}{(1 + 0.16) \times (B6 - B5)/(B6 + B5 + 0.16)}$	(Wu et al., 2008)
Sentinel-2 red-edge position (S2REP)	$705 + 35 \times \frac{(B7 + B4)/2 - B5}{B6 - B5}$	(Frampton et al., 2013)
Sentinel-2 triangular vegetation index (STVInorm)	$\frac{S_{RT} - S_{AT}}{S_{RT} + S_{AT}}$ $S_{AT} = 0.5 \times [105 \times (B5 - B2) - 145 \times (B4 - B2)]$ $S_{RT} = 0.5 \times [125 \times (B7 - B6) - 145 \times (B8A - B6)]$	(Qian et al., 2022)

## 4. Results

### 4.1. The REP characteristics of vegetation

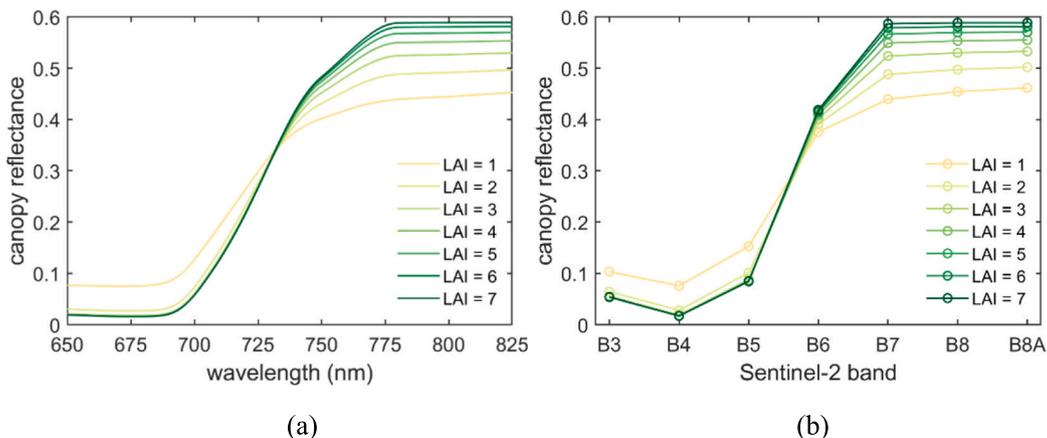
The shape and magnitude of vegetation red-edge spectra are influenced by multiple factors, making it essential to distinguish the contribution of each factor to these red-edge spectral characteristics, especially the red-edge position, which plays a significant role in delineating LCC, as previous studies have indicated (Curran et al., 1991; Gitelson et al., 2005; Wu et al., 2008).

A prominent absorption feature in the red band reflectance of Sentinel-2 is observed in the leaf spectrum, originating from the strong absorption by chlorophyll at around 675–681 nm. In the NIR spectral domain (778–800 nm), a relatively flat plateau is evident in the leaf spectrum (Fig. 3 and Fig. 4), and leaf chlorophyll does not exert a significant impact in the NIR band reflectance (Fig. 3). Positioned between the red valley and the NIR plateau is the red-edge spectral interval, and the variation in chlorophyll content influences its starting and ending points (Fig. 3). As leaf chlorophyll content increases, both the starting and ending points of the red-edge shift towards longer wavelengths, resulting in a red shift of the REP (Fig. 3). This shift occurs because there

is minimal absorption of most other leaf biochemical constituents within this spectral interval, and the absorption coefficient of chlorophyll gradually decreases with increasing wavelength (Fig. S2).

Leaf area index characterizes the leaf density within the vegetation canopy (Chen and Black, 1992), and it significantly influences the shape of the red-edge spectrum at the canopy scale. Simulations using the PROSAIL model indicate that as LAI increases from 1 to 7, the value of the NIR band reflectance of Sentinel-2 sharply rises (Fig. 4). Additionally, the ending point of the red-edge tends to shift towards longer wavelength as LAI increases, resulting in a mild red shift of the REP (Fig. 4).

The magnitude of the NIR plateau is affected by the degree of absorption from dry matter and the multi-scattering from leaf structure (Fig. S3a and Fig. S3c). Variations in the leaf structure parameter change the starting point of the red-edge (Fig. S3c), contributing to the overall REP variation alongside chlorophyll content. The variation in leaf inclination angle impacts the value of NIR plateau and the slope of the red-edge but has nearly no effect on the REP (Fig. S3d). As the proportion of dry soil, i.e., soil moisture content, in the background soil increases, the magnitude of canopy reflectance spectra enhances in the NIR plateau domain, resulting in a slight steepening of the red-edge

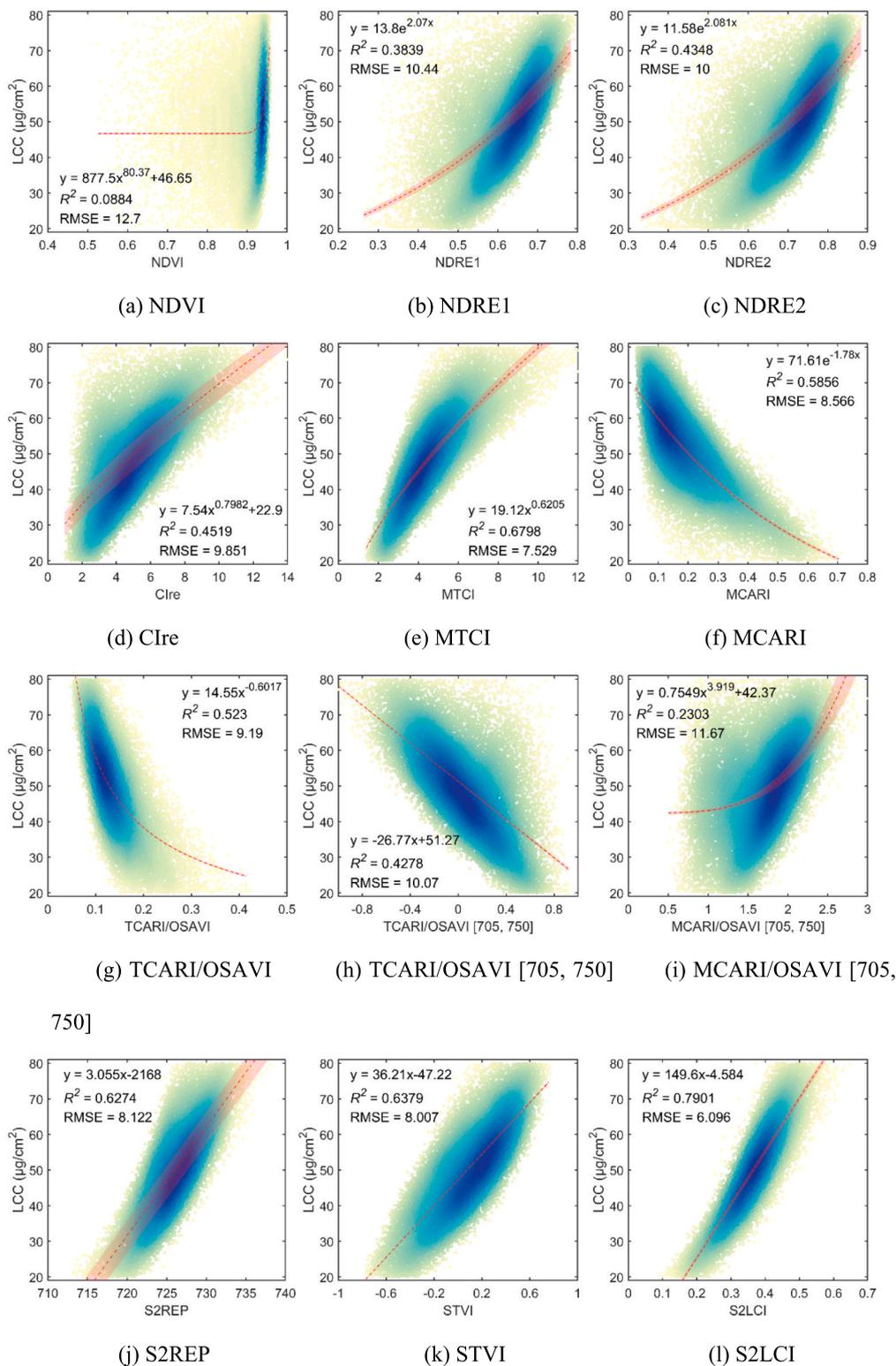


**Fig. 4.** Canopy-scale reflectance spectra in the 650–825 nm interval (a) and Sentinel-2 band reflectance (b) simulated by the PROSAIL model (Jacquemoud et al., 2009). Simulations were performed over a range of increasing LAI from 1 to 7 at a step of 1. The average leaf inclination angle was set to 30°.

curve (Fig. S2b). However, the REP remains unaffected by changes in soil moisture, underscoring its suitability as an index to eliminate the influence of soil characteristics.

In summary, changes in leaf chlorophyll content significantly impact the REP, resulting in a red shift with higher LCC. LAI has a minor

influence on the ending point position of the red-edge and, consequently, a slight variation in the REP. While a higher leaf structure parameter may cause a red shift in the starting point position of the red-edge, its impact is limited compared to LCC and LAI, especially in homogeneous regions or specific vegetation types. Other parameters, such



**Fig. 5.** Relationship between chlorophyll-related vegetation indices and leaf chlorophyll content with the PROSAIL simulated dataset (n = 20,000). The dashed red line represents the best-fitting curve (linear or quadratic polynomial, power, or exponential) for estimating LCC with each vegetation index, and the shaded red area represents the 95 % confidence bounds of the regression equation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as leaf dry matter content, leaf inclination angle of canopy, and background soil properties, have minimal effect on the REP.

#### 4.2. S2LCI evaluation with PROSAIL simulation

The performance of the S2LCI in LCC estimation was first evaluated, and a comparison was made with other chlorophyll vegetation indices using the PROSAIL simulated dataset. Multiple regression models, encompassing linear, quadratic polynomial, power, and exponential regressions, were employed to formulate predictive models for LCC based on each vegetation index. The model that yielded the lowest RMSE was designated as the optimal fit for the respective vegetation index. Generally, all the chlorophyll vegetation indices, except for the NDVI, provided reliable LCC estimates, with  $R^2$  for the regression models ranging from 0.3839 to 0.6798 and RMSE from 7.529 to 10.44  $\mu\text{g cm}^{-2}$  (Fig. 5).

Excluding the NDVI, the MCARI/OSAVI [705, 750], TCARI/OSAVI [705, 750], NDRE1 and NDRE2 exhibited the worst performance in LCC estimation among all the indices, with RMSE values exceeding 10  $\mu\text{g cm}^{-2}$  (Fig. 5b, c, h, i). These indices displayed significant scatter in the low LCC range, indicating sensitivity to underlying soil properties, especially for low LAIs. The MCARI/OSAVI [705, 750] was originally designed for hyperspectral sensor like Hyperion (Wu et al., 2008), which likely limits its adaptation to the characteristics of the Sentinel-2 multispectral sensor. The MCARI, TCARI/OSAVI, and CIre showed moderate goodness-of-fit with LCC, achieving  $R^2$  values for the regression model between 0.4519 and 0.5856, and RMSE values ranging from 8.566 to 9.851  $\mu\text{g cm}^{-2}$  (Fig. 5d, f, g). However, these VIs still exhibited substantial dispersion in some LCC intervals, and their relationship with

LCC were somewhat nonlinear. For instance, CIre showed higher uncertainty in the estimation of large LCC values compared to small ones, while MCARI and TCARI/OSAVI demonstrated the opposite trend.

The MTCI, S2REP and STVI achieved the highest retrieval accuracy and goodness-of-fit with LCC among all the VIs considered for comparison, with  $R^2$  values for the regression model ranging from 0.6274 to 0.6798, and RMSE values between 7.529 and 8.122  $\mu\text{g cm}^{-2}$  (Fig. 5e, j, k). These VIs displayed better linearity with LCC and were less susceptible to background effects. Notably, the MTCI was particularly effective in LCC estimation in smaller value intervals.

The S2LCI proposed in this study outperformed all other VIs in LCC estimation, with an  $R^2$  value for the regression model of 0.7901 and an RMSE of 6.096  $\mu\text{g cm}^{-2}$  (Fig. 5l). This represented a significant improvement compared to the other VIs. Moreover, the linear relationship between S2LCI and LCC remained stable across the entire range of LCC values, and the uncertainty did not increase for extremely high or low LCC values. This indicated that there were no systematic over- or underestimation in LCC retrieval application, highlighting the robustness and reliability of S2LCI for diverse LCC ranges.

It's important to note that some VIs, including NDVI and MCARI, showed saturation effects when LCC exceeds 40  $\mu\text{g cm}^{-2}$ . This saturation phenomenon can limit their performance in accurately estimating LCC. The S2LCI effectively mitigated this issue, leading to enhanced stability and reliability in LCC estimation, even within larger LCC intervals. This characteristic underscores the superiority of S2LCI over traditional VIs in handling high LCC values, emphasizing its potential for robust LCC assessment.

The effect of LAI on the retrieval accuracy and sensitivity of different vegetation indices, including the S2LCI, MTCI, S2REP, and STVI, was

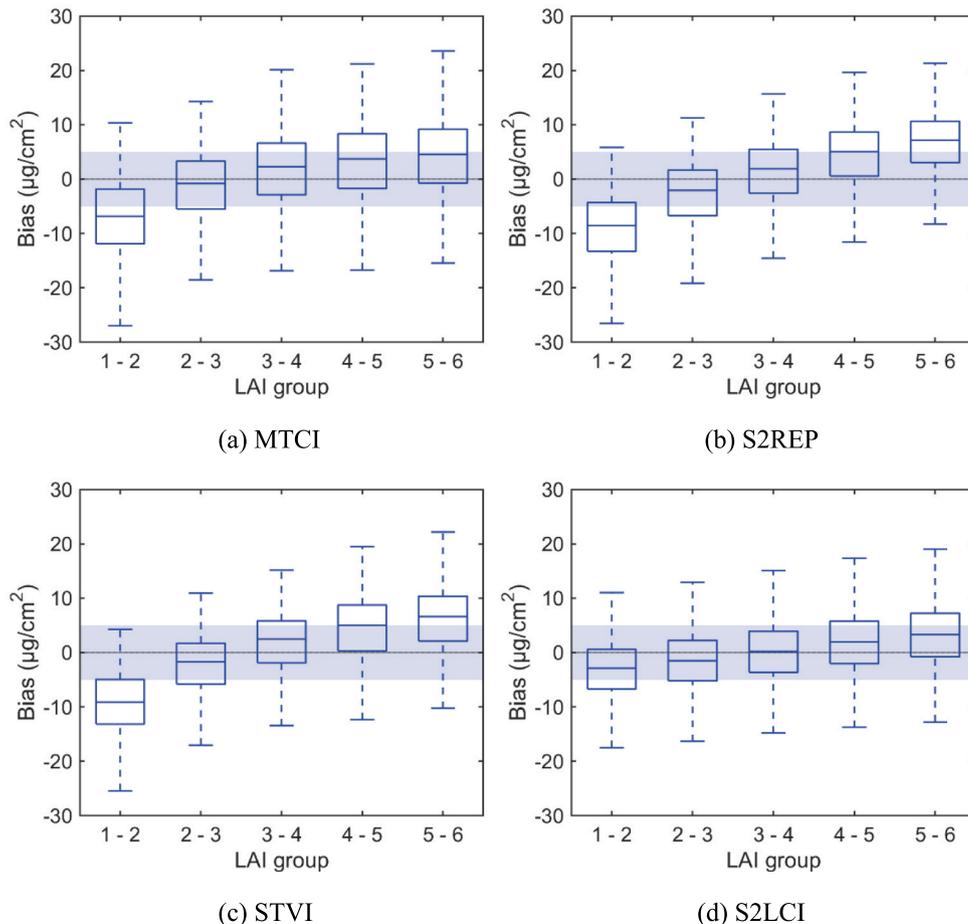


Fig. 6. Whisker boxplots of LCC estimation bias for different LAI intervals using (a) MTCI, (b) S2REP, (c) STVI, and (d) S2LCI. The shaded area represents  $\pm 5 \mu\text{g/cm}^2$  of the LCC retrieval error.

further explored. The PROSAIL simulation dataset was grouped based on LAI values, and the bias in LCC estimation were calculated for each LAI group. Fig. 6 presents boxplots of the LCC estimation bias for different LAI intervals, with the shaded area representing a range of  $\pm 5 \mu \text{g cm}^{-2}$  of LCC retrieval error.

Overall, the S2LCI exhibited the highest independence from variations in LAI, with minimal estimation bias across different LAI levels. In contrast, the MTCI, S2REP, and STVI showed noticeable positive or negative estimation biases in response to large or small LAI values. For moderate LAI conditions, particularly with LAI values between 3 and 4, the bias was small and close to 0 for all the VIs. However, as LAI increased or decreased, the S2REP exhibited a corresponding red or blue shift, resulting in an overestimation or underestimation of LCC. A similar pattern was observed for the MTCI and STVI, both of which were significantly influenced by LAI. It's worth noting that the estimating uncertainty, represented by the length of the box whisker, was relatively higher for the MTCI across all LAI intervals compared to other VIs.

These results emphasize the advantage of the S2LCI in terms of its reduced sensitivity to variations in LAI, making it more robust for LCC estimation across different canopy structures. In contrast, traditional VIs like MTCI, S2REP, and STVI demonstrated greater sensitivity to LAI, which could lead to biased LCC estimates, especially under extreme LAI conditions. This underscores the capability of S2LCI to provide reliable and consistent LCC estimations under diverse vegetation canopy densities.

### 4.3. S2LCI evaluation with ground canopy spectra

The study tested the applicability of the S2LCI at both the ground scale using hyperspectral canopy spectra datasets and satellite scale using Sentinel-2 images. In the ground-scale testing with winter wheat canopy spectra collected at the Yangling experimental sites, the hyperspectral data were converted to multispectral information using the Sentinel-2/MSI spectral response function. The S2LCI, as well as three well-performing comparative VIs (MTCI, S2REP and STVI), were calculated and compared with ground SPAD measurements. Unlike the results from simulated datasets, relationship between VIs and LCC at the ground scale demonstrated increased instability and retrieval uncertainty. Nevertheless, the S2LCI exhibited the highest goodness-of-fit with the SPAD value ( $R = 0.492$ ), and its  $R$  value was 16.7 % to 23.6 % higher than that of the three comparative VIs (Fig. 7), indicating acceptable model accuracy.

### 4.4. S2LCI evaluation with Sentinel-2 images

#### 4.4.1. Ground measured LCC validation

The S2LCI and the three comparative VIs were then calculated using Sentinel-2 images and validated with ground SPAD measurements of winter wheat in the Luohe experimental area. These measurements were conducted during the tilling and jointing stages of winter wheat. The validation of the VIs demonstrated a performance similar to that of the ground canopy spectra dataset, with generally lower goodness-of-fits compared to the PROSAIL simulations. This difference might be

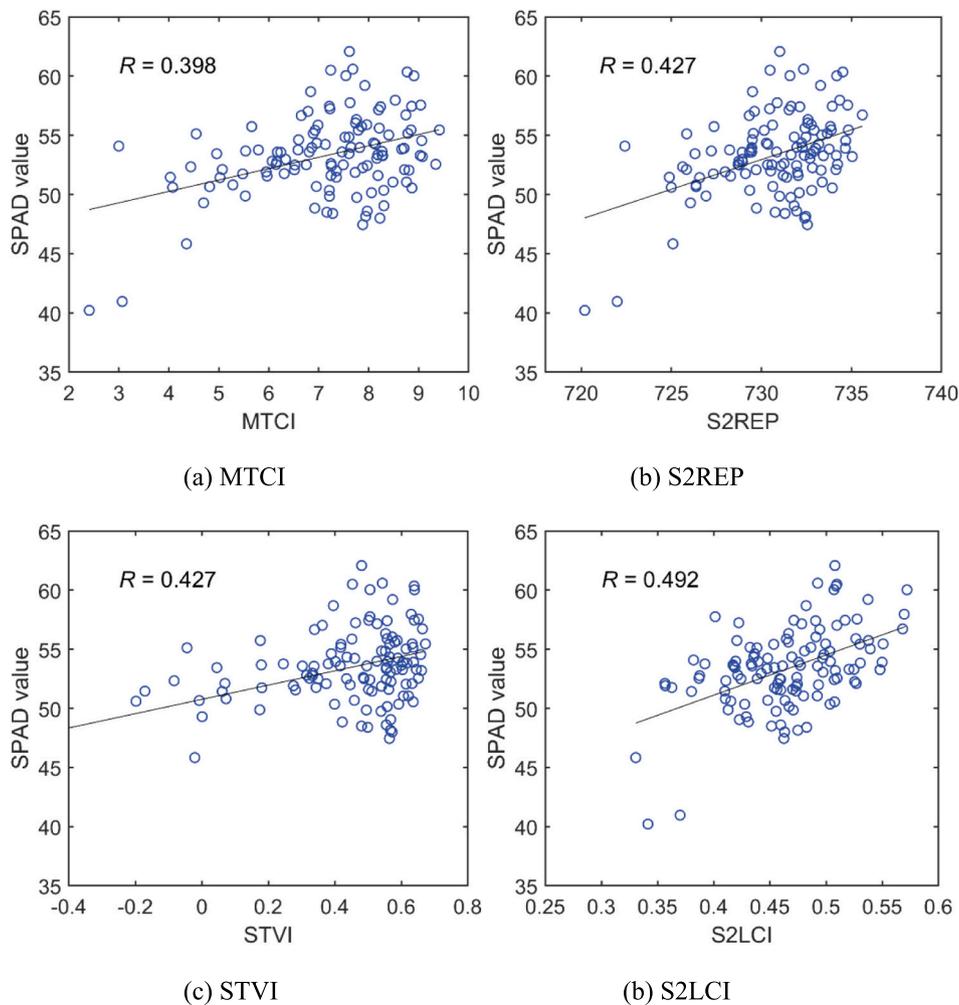


Fig. 7. Relationship between VIs derived from ground canopy spectra dataset and measured leaf SPAD value of winter wheat in the Yangling experimental sites.

attributed to various factors such as the geometry of the Sentinel-2/MSI, atmospheric effect, geo-registration error, and mixed pixels.

The S2LCI consistently showed the highest goodness-of-fit with the SPAD values in both the ground-scale and satellite image-scale validations ( $R = 0.526$ ), while the S2REP and STVI had slightly lower but very similar accuracy to the S2LCI (Fig. 8). In contrast, the MTCI showed the poorest performance, which could be related to spectral band differences between Sentinel-2/MSI and MERIS. In summary, the S2LCI demonstrated a high goodness-of-fit with the SPAD value in both ground-scale and satellite image-scale applications. This characteristic is valuable for regional-level estimation of LCC, highlighting the potential of the S2LCI in practical applications.

#### 4.4.2. Comparison with SNAP biophysical processor derived LCC

The study compared the LCC retrieval results obtained from the S2LCI and the SNAP biophysical processor in a small region near Luohe City, which was where the field LCC measurement were conducted. Despite the performance of SNAP LCC may varied from crop type to region, it could achieve comparable accuracy especially for homogeneous winter wheat (Xie et al., 2019). Thus, we adopted it as a benchmark to explore its consistency with proposed S2LCI. The comparison was conducted for two specific dates, March 9 and April 8, 2018, corresponding to the tilling and jointing stages of winter wheat.

Fig. 9 and Fig. 10 depict the spatial distribution of S2LCI and SNAP-derived LCC over the study region. A notable difference in LCC between the tilling and jointing stage was observed in the SNAP-derived result (Fig. 11). In contrast, more spatial details were observed in the S2LCI maps for both stages. The coefficient of variation, derived as the

standard variation normalized by the mean value, was higher for S2LCI (0.079 and 0.102) compared to SNAP LCC (0.072 and 0.091) within the region (Fig. 11).

The correlation coefficient between S2LCI and SNAP-derived LCC was larger during the tilling stage compared to the jointing stage (Fig. 12). In other words, the consistency between S2LCI and SNAP-derived LCC was higher for smaller LCC values than for larger LCC values. However, it's worth noting that the LCC values between these two growing stages did not significantly differ based on the field LCC measurements. This resulted in noticeable under- or overestimation in the tilling or jointing stages of SNAP results (Fig. S4). As a result, the relatively small change in S2LCI histograms was closer to the actual situation compared to the significant shift in SNAP LCC histograms between two winter wheat growing stages (Fig. 11). These findings indicate that the S2LCI demonstrated more spatial detail and greater consistency between the two growing stages, which makes it a promising tool for tracking changes in LCC during different crop growth stages.

## 5. Discussions

Quantitatively assessing leaf chlorophyll content is of paramount importance in ecological, environmental modeling, and agricultural applications (Chen et al., 2022; Croft et al., 2017; Huang et al., 2019). Nonetheless, accurately retrieving LCC from optical remotely sensed data has posed a formidable challenge due to its relatively subtle signal within canopy spectra compared to other vegetation parameters. Researchers have been diligently working to disentangle the LCC signal

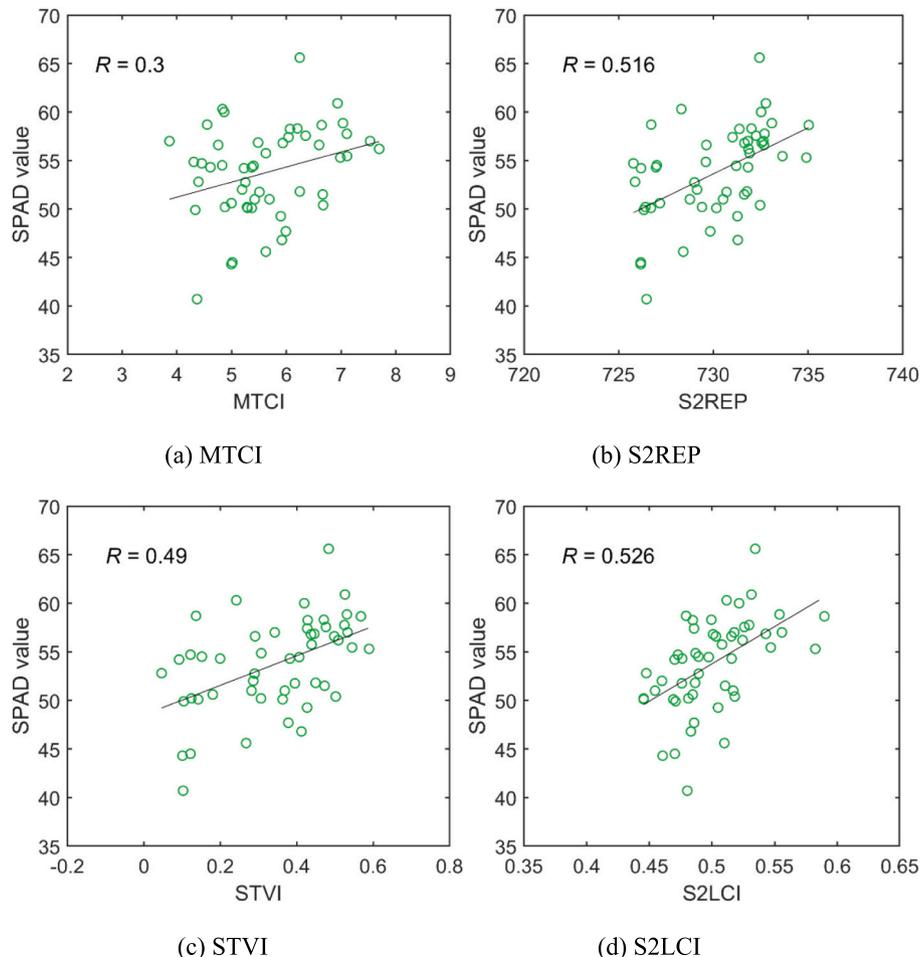
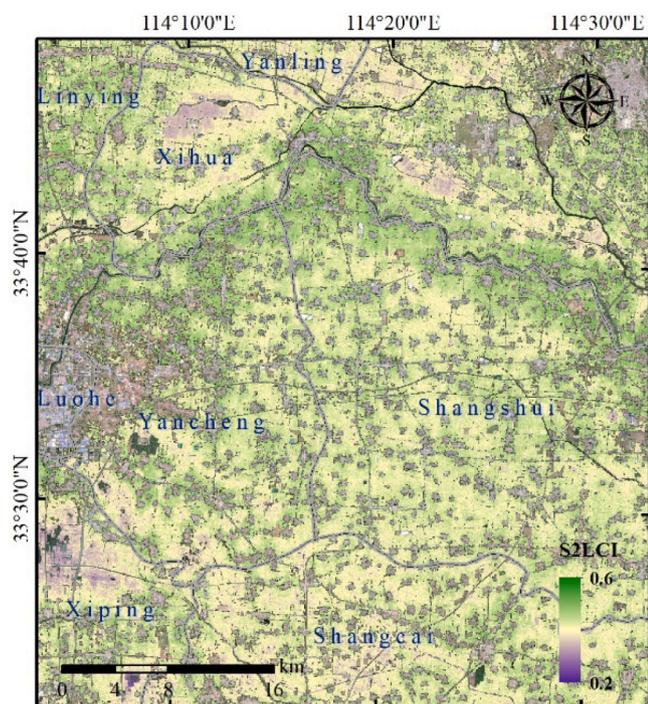
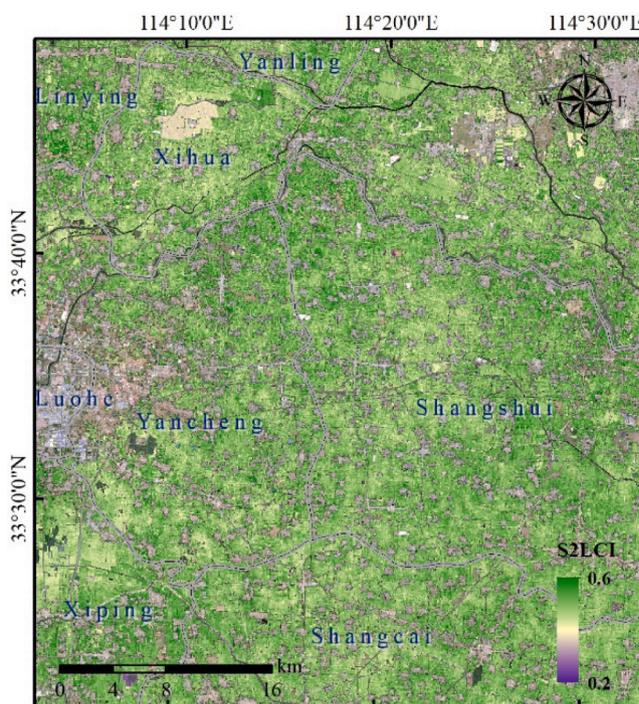


Fig. 8. Relationship between VIs derived from Sentinel-2 images and ground measured leaf SPAD value of winter wheat in the Luohe experimental area.

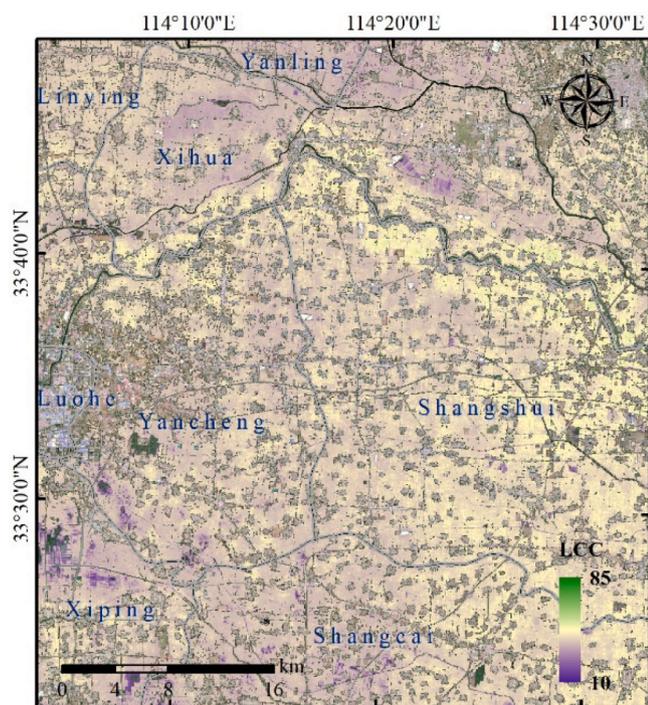


(a) March 9<sup>th</sup>, 2018



(b) April 8<sup>th</sup>, 2018

Fig. 9. S2LCI maps of a small region near Luohe City derived from Sentinel-2/MSI on (a) March 9th, 2018 and (b) April 8th, 2018. The background images are the true color composites from Sentinel-2/MSI.

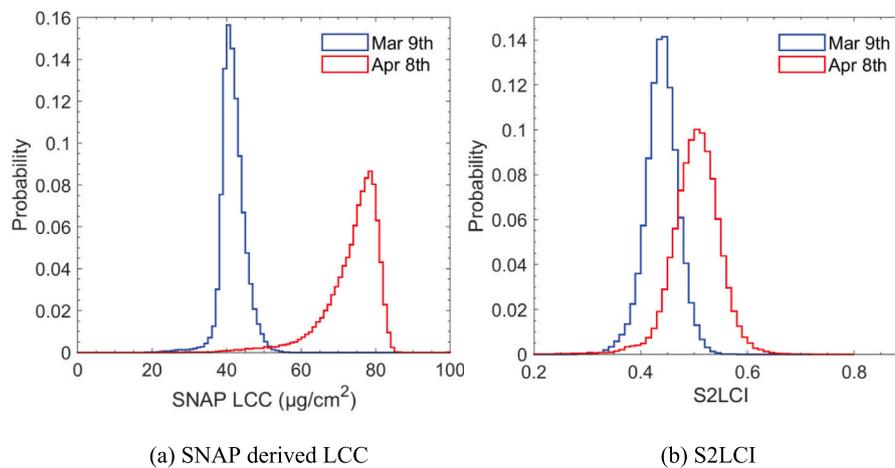


(a) March 9<sup>th</sup>, 2018

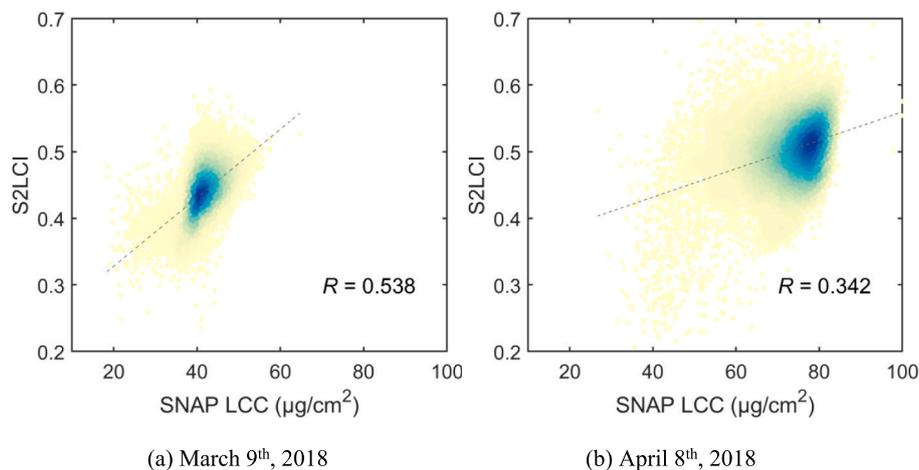


(b) April 8<sup>th</sup>, 2018

Fig. 10. LCC maps of a small region near Luohe City derived from Sentinel-2/MSI with SNAP on (a) March 9th, 2018 and (b) April 8th, 2018. The background images are the true color composites from Sentinel-2/MSI.



**Fig. 11.** Histograms of (a) SNAP-derived leaf chlorophyll content and (b) S2LCI probability over the small region near Luohe City (Fig. 9 and Fig. 10) for March 9th, 2018 and April 8th, 2018. The coefficient of variations for S2LCI are 0.079 (March) and 0.102 (April), while they are 0.071 (March) and 0.091 (April) for SNAP LCC.



**Fig. 12.** Comparison of S2LCI and SNAP-derived leaf chlorophyll content over the small region near Luohe City (Fig. 9 and Fig. 10) for (a) March 9th, 2018 and (b) April 8th, 2018.

from the overall vegetation response, thereby enhancing retrieval accuracy. These efforts span both hyperspectral and multispectral data (Gitelson et al., 2014; Wu et al., 2008; Zhang et al., 2021). While the red band is acutely affected by chlorophyll, coinciding with its peak absorption, it becomes saturated at higher chlorophyll concentrations, causing a shift in the red-edge domain (Curran et al., 1991; Horler et al., 1983). Leveraging the potential of the Sentinel-2 sensor's additional three red-edge multispectral bands has become a promising avenue for cost-effective and precise LCC estimation.

To harness red-edge information in LCC retrieval, the ESA's official algorithm, integrated into the SNAP software, relies on a combination of physical radiative transfer modeling and neural network retrieval strategy (Weiss et al., 2020). Nevertheless, this method is often time-intensive and not conducive to swift, large-scale LCC mapping. Furthermore, the accuracy of LCC estimates derived from SNAP's algorithm may not consistently meet the demands of various scenarios (Xie et al., 2019). Alternatively, employing vegetation indices offers a practical and efficient means of characterizing relative LCC values. This research endeavors to introduce a novel chlorophyll index, leveraging the multispectral red-edge information from the Sentinel-2 platform, aimed at mitigating interference from factors like LAI and other confounding variables.

### 5.1. Suitability and advantages of the S2LCI

In the context of previous research (Curran et al., 1991; Frampton et al., 2013; Gitelson et al., 2005; Wu et al., 2008) and our own analyses based on PROSPECT-D and SAIL simulations, it's well-established that the REP of the vegetation spectrum is predominantly influenced by leaf chlorophyll content, with relatively minor contributions from canopy leaf area in most scenarios. The foundation of the proposed Sentinel-2 leaf chlorophyll index (S2LCI) lies in the innovative S2REP, which effectively estimates the red-edge position by harnessing information from the Sentinel-2 red and red-edge multispectral bands, along with the S2NDRE. Consequently, the S2LCI shares with the S2REP the valuable attribute of robustness against variations in background soil spectra, as affirmed by our sensitivity analysis, which even indicates that the S2LCI exhibits superior and more consistent performance under different soil moisture conditions (Fig. S5). This makes it particularly well-suited for LCC assessments where background soil characteristics may vary.

Notably, the S2LCI distinguishes itself from other prominent chlorophyll related vegetation indices like the MTCI and STVI. While these indices are more influenced by LAI (Dash and Curran, 2004; Qian et al., 2022), S2LCI significantly reduces the impact of it (Fig. 6). This pivotal advantage renders it a valuable tool for comparing LCC across various phenological stages of vegetation when LAIs may exhibit different.

While hyperspectral vegetation indices are generally acknowledged

for their superior correlation with LCC, the limited spatial coverage and restricted access to airborne or satellite hyperspectral data can hinder large-scale regional LCC mapping. Achieving a balance between accuracy and cost-effectiveness is paramount. In this context, the Sentinel-2 multispectral-based S2LCI, which has demonstrated its accuracy through evaluations with simulated and *in-situ* datasets, can serve as a valuable addition to the suite of Sentinel-2 vegetation indices (Frampton et al., 2013; Qian et al., 2022) or approaches (Li et al., 2021) for LCC estimation. Its ability to provide accurate and cost-effective LCC estimates makes it particularly suitable for applications at regional and larger scales.

### 5.2. Sources of uncertainty and limitations

Despite the impressive performance of the S2LCI in LCC estimation, it is crucial to acknowledge certain sources of uncertainty and limitations that should be considered when applying this index. One key limitation arises from the assumption that the slope of the LCC isoline in the  $S2REP_{norm}/S2NDRE$  space is constant across all LCC levels. In reality, this slope experiences slight changes with varying LCC values, particularly becoming smaller under extremely low or high LCC conditions. This inherent variability implies that the relationship between the S2LCI and LCC may exhibit increased uncertainty in scenarios of extreme LCC values. One way to mitigate this limitation could involve implementing a variable slope for the LCC isoline, tailored to different LCC ranges.

The S2LCI relies on two fundamental VIs, S2REP and S2NDRE, and any errors introduced into these constituent VIs can propagate into the S2LCI. For instance, while S2NDRE is highly sensitive to LAI and insensitive to the LCC, it can be influenced by factors such as leaf angle inclination (Sun et al., 2023). The impact of leaf angle inclination on S2LCI is evident in the increasing bias of S2LCI in LCC estimation as the average leaf angle enlarges (Fig. S6). Therefore, research efforts should continue to identify more suitable vegetation indices serve as the proxies of LAI and LCC, minimizing their susceptibility to the influence of other factors. The concept of an error ellipse, inspired by the standard deviation within a small elliptical region in the 2-dimensional space, as utilized in the MODIS LAI/FPAR product (Knyazikhin et al., 1998) and certain matrix-based vegetation indices (Xu et al., 2019), could be integrated into the S2LCI calculation. This would provide a quality flag, indicating the degree of uncertainty associated with the S2LCI value.

Addressing mixed pixel, especially those covering sparse vegetation, remains a challenge due to the dilution of the canopies signal by background soil. While the S2LCI is unbiased in LCC estimation with respect to soil (Fig. S5b), the uncertainty increases as the proportion of vegetation in the mixed pixel decreases. Disentangling the contributions of vegetation and soil in mixed pixel and calculating the S2LCI with the vegetation component alone is essential but challenging with multispectral data (Bioucas-Dias et al., 2012). Further research should explore methods to better account for mixed pixels in S2LCI applications.

The relationship between the S2LCI and LCC, as established in this work, is based on the SAIL canopy radiative transfer model, and only validated on wheat. It's important to recognize that this relationship may vary depending on the vegetation species and the specific characteristics of heterogeneous canopies. Consequently, future work should delve into biome-specific relationships between the S2LCI and LCC to account for this variability. Expanding the dataset with additional field measurements is paramount to provide a more comprehensive evaluation of the S2LCI. This should encompass a broader range of crop types, temporal periods, and geographical regions to enhance the validity and generalizability of the proposed index. In doing so, the S2LCI can be further validated and refined for broader agricultural contexts.

### 5.3. Potential applications

Despite the inherent limitations of the vegetation index approach in remote sensing for vegetation variables, such as regional dependency, the rapid expansion of satellite sensor capabilities, encompassing both broad and narrow spectral bands, continues to drive the development of VIs (Zeng et al., 2022). One of the notable challenges with existing leaf chlorophyll products is their spatial resolution, typically ranging from kilometer to hectometer (Croft et al., 2020; Dash et al., 2010; Xu et al., 2022). The introduction of the S2LCI presents a compelling solution to rapidly attain decameter-scale LCC estimates using the Sentinel-2 imagery, particularly when coupled with cloud-computing platforms such as the Google Earth Engine (GEE) (Gorelick et al., 2017). This improvement in spatial resolution is particularly advantageous for precision agriculture and other applications that require fine-grained LCC spatial information.

The concept underpinning the S2LCI holds promise for extension to other multispectral or hyperspectral sensors equipped with two or more red-edge configurations, such as China's GaoFen-6/WFV (Yang et al., 2020), ESA's Sentinel-5P/TROPOMI (Borsdorff et al., 2018; Guanter et al., 2021), and the upcoming FLEX mission, which covers a broader spectral range from 500 to 780 nm (Drusch et al., 2017). However, it's important to recognize that adapting the specific formulation and coefficients of the new leaf chlorophyll index to these sensors will be necessary to ensure optimal performance.

Moreover, the S2LCI operates within a 2-dimensional space that primarily accounts for leaf area in its calculation. Expanding this framework to a higher-dimensional space could facilitate the inclusion of additional factors in LCC estimation, potentially enhancing the index's versatility and precision. This expansion aligns with the increasing availability of spectral data from advanced sensors and presents opportunities for more comprehensive approaches to vegetation variable retrieval.

## 6. Conclusions

In this study, we introduce a novel leaf chlorophyll sensitive vegetation index, the Sentinel-2 leaf chlorophyll index (S2LCI), utilizing the Sentinel-2 multispectral red-edge information. This index offers a valuable tool for rapid and high-resolution monitoring of leaf chlorophyll content over large geographic regions. The S2LCI is designed based on a 2-dimensional space defined by the red-edge position (REP) and a LAI indicator, effectively mitigating the influence of variations in leaf area. Thorough validation and comparison with existing leaf chlorophyll-related vegetation indices demonstrate that the S2LCI outperforms alternative indices, showcasing the strongest correlation with LCC and the lowest RMSE in LCC estimation across both simulated and ground-based datasets. Furthermore, a comparative assessment with LCC estimates obtained through the SNAP biophysical processor reveals that the S2LCI consistently delivers more reliable results, particularly across diverse crop growth stages and with enhanced spatial details. The S2LCI excels in its resilience to fluctuations in LAI and background soil properties, underpinning its suitability for monitoring leaf chlorophyll content across different phenological and vegetation states. Nevertheless, the S2LCI is not immune to certain sources of uncertainty, such as error accumulation associated with the underlying REP and S2NDRE, and the mixed-pixel effect. Future research should focus on addressing these challenges and enhancing the robustness of the S2LCI in LCC estimation. Moreover, the concept of the S2LCI has the potential to be extended to other multispectral sensors equipped with two or more red-edge bands. This study serves as a foundation for future investigations that may consider the integration of additional factors using higher-dimensional spaces, thereby further refining the accuracy and scope of leaf chlorophyll content estimation in the realm of remote sensing.

## CRedit authorship contribution statement

**Yuanheng Sun:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Qiming Qin:** Conceptualization, Resources, Supervision, Writing – original draft. **Yao Zhang:** Data curation, Investigation, Validation. **Huazhong Ren:** Formal analysis, Investigation, Validation. **Guhuai Han:** Data curation, Formal analysis, Software. **Zhaoxu Zhang:** Investigation, Visualization. **Tianyuan Zhang:** Data curation, Formal analysis. **Binyu Wang:** Investigation, Software.

## Declaration of competing interest

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

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (42301391) and Fundamental Research Funds for the Central Universities of China (3132024120).

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compag.2025.110500>.

## Data availability

Data will be made available on request.

## References

- Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., Hank, T., 2018a. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: a review study. *Remote Sens. (Basel)* 10, 85.
- Berger, K., Atzberger, C., Danner, M., Woher, M., Mauser, W., Hank, T., 2018b. Model-Based Optimization of Spectral Sampling for the Retrieval of Crop Variables with the PROSAIL Model. *Remote Sens. (Basel)* 10, 2063.
- Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., Chanussot, J., 2012. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5, 354–379.
- Borsdorff, T., Aan de Brugh, J., Hu, H., Aben, I., Hasekamp, O., Landgraf, J., 2018. Measuring carbon monoxide with TROPOMI: first results and a comparison with ECMWF-IFS analysis data. *Geophys. Res. Lett.* 45, 2826–2832.
- Chen, J.M., Black, T.A., 1992. Defining leaf area index for non-flat leaves. *Plant Cell Environ.* 15, 421–429.
- Chen, J.M., Wang, R., Liu, Y., He, L., Croft, H., Luo, X., Wang, H., Smith, N.G., Keenan, T. F., Prentice, I.C., Zhang, Y., Ju, W., Dong, N., 2022. Global datasets of leaf photosynthetic capacity for ecological and earth system research. *Earth Syst. Sci. Data* 14, 4077–4093.
- Croft, H., Chen, J.M., Luo, X., Bartlett, P., Chen, B., Staebler, R.M., 2017. Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. *Glob. Chang. Biol.* 23, 3513–3524.
- Croft, H., Chen, J.M., Wang, R., Mo, G., Luo, S., Luo, X., He, L., Gonsamo, A., Arabian, J., Zhang, Y., Simic-Milas, A., Noland, T.L., He, Y., Homolová, L., Malenovsky, Z., Yi, Q., Beringer, J., Amiri, R., Hutley, L., Arellano, P., Stahl, C., Bonal, D., 2020. The global distribution of leaf chlorophyll content. *Remote Sens. Environ.* 236, 111479.
- Curran, P.J., Dungan, J.L., Macler, B.A., Plummer, S.E., 1991. The effect of a red leaf pigment on the relationship between red edge and chlorophyll concentration. *Remote Sensing of Environment* 35, 69–76.
- Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* 25, 5403–5413.
- Dash, J., Curran, P.J., Tallis, M.J., Llewellyn, G.M., Taylor, G., Snoeij, P., 2010. Validating the MERIS Terrestrial Chlorophyll Index (MTCI) with ground chlorophyll content data at MERIS spatial resolution. *Int. J. Remote Sens.* 31, 5513–5532.
- Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtry, J.E., 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sens. Environ.* 74, 229–239.
- del Río-Mena, T., Willemen, L., Tesfamariam, G.T., Beukes, O., Nelson, A., 2020. Remote sensing for mapping ecosystem services to support evaluation of ecological restoration interventions in an arid landscape. *Ecol. Ind.* 113, 106182.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36.
- Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., Kraft, S., Middleton, E.M., Miglietta, F., Mohammed, G., Nedbal, L., Rascher, U., Schittemeyer, D., Verhoef, W., 2017. The Fluorescence Explorer Mission Concept-ESA's Earth Explorer 8. *IEEE Trans. Geosci. Remote Sens.* 55, 1273–1284.
- Féret, J.B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: towards modeling leaf optical properties through a complete lifecycle. *Remote Sens. Environ.* 193, 204–215.
- Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* 82, 83–92.
- Gitelson, A., Merzlyak, M.N., 1994. Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J. Plant Physiol.* 143, 286–292.
- Gitelson, A.A., Merzlyak, M.N., 1997. Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.* 18, 2691–2697.
- Gitelson, A.A., Peng, Y., Arkebauer, T.J., Schepers, J., 2014. Relationships between gross primary production, green LAI, and canopy chlorophyll content in maize: implications for remote sensing of primary production. *Remote Sens. Environ.* 144, 65–72.
- Gitelson, A.A., Viña, A., Gignanda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 32.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27.
- Green, J.K., Konings, A.G., Alemohammad, S.H., Berry, J., Entekhabi, D., Kolassa, J., Lee, J.-E., Gentine, P., 2017. Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nat. Geosci.* 10, 410–414.
- Gu, C., Du, H., Mao, F., Han, N., Zhou, G., Xu, X., Sun, S., Gao, G., 2016. Global sensitivity analysis of PROSAIL model parameters when simulating Moso bamboo forest canopy reflectance. *Int. J. Remote Sens.* 37, 5270–5286.
- Gu, Y., Liu, T., Gao, G., Ren, G., Ma, Y., Chanussot, J., Jia, X., 2021. Multimodal hyperspectral remote sensing: an overview and perspective. *SCIENCE CHINA Inf. Sci.* 64, 121301.
- Guanter, L., Bacour, C., Schneider, A., Aben, I., van Kempen, T.A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., Zhang, Y., 2021. The TROPISIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission. *Earth Syst. Sci. Data* 13, 5423–5440.
- Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* 81, 416–426.
- Heckmann, D., Schlüter, U., Weber, A.P.M., 2017. Machine Learning Techniques for Predicting Crop Photosynthetic Capacity from Leaf Reflectance Spectra. *Mol. Plant* 10, 878–890.
- Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4, 273–288.
- Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sens. Environ.* 112, 186–202.
- Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., Xie, X., 2019. Assimilation of remote sensing into crop growth models: current status and perspectives. *Agric. For. Meteorol.* 276–277, 107609.
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+SAIL models: a review of use for vegetation characterization. *Remote Sens. Environ.* 113, S56–S66.
- Knyazikhin, Y., Martonchik, J.V., Myneni, R.B., Diner, D.J., Running, S.W., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *J. Geophys. Res. Atmos.* 103, 32257–32275.
- Li, Y., Ma, Q., Chen, J.M., Croft, H., Luo, X., Zheng, T., Rogers, C., Liu, J., 2021. Fine-scale leaf chlorophyll distribution across a deciduous forest through two-step model inversion from Sentinel-2 data. *Remote Sens. Environ.* 264, 112618.
- Lu, B., Dao, P.D., Liu, J., He, Y., Shang, J., 2020. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sens. (Basel)* 12, 2659.
- Markwell, J., Osterman, J.C., Mitchell, J.L., 1995. Calibration of the Minolta SPAD-502 leaf chlorophyll meter. *Photosynth. Res.* 46, 467–472.
- Moharana, S., Dutta, S., 2016. Spatial variability of chlorophyll and nitrogen content of rice from hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* 122, 17–29.
- Noda, I., 2017. Vibrational two-dimensional correlation spectroscopy (2DCOS) study of proteins. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 187, 119–129.
- Ollinger, S.V., 2011. Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytol.* 189, 375–394.
- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y., Ranagalage, M., 2020. Sentinel-2 data for land cover/use mapping: a review. *Remote Sens. (Basel)* 12, 2291.
- Qian, B., Ye, H., Huang, W., Xie, Q., Pan, Y., Xing, N., Ren, Y., Guo, A., Jiao, Q., Lan, Y., 2022. A sentinel-2-based triangular vegetation index for chlorophyll content estimation. *Agric. For. Meteorol.* 322, 109000.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the great plains with ERTS.
- Sonobe, R., Hirono, Y., Oi, A., 2020. Quantifying chlorophyll-a and b content in tea leaves using hyperspectral reflectance and deep learning. *Remote Sens. Lett.* 11, 933–942.

- Sun, Y., Knyazikhin, Y., She, X., Ni, X., Chen, C., Ren, H., Myneni, R.B., 2022a. Seasonal and long-term variations in leaf area of Congolese rainforest. *Remote Sens. Environ.* 268, 112762.
- Sun, Y., Qin, Q., Ren, H., Zhang, T., Chen, S., 2020. Red-Edge Band Vegetation Indices for Leaf Area Index Estimation From Sentinel-2/MSI Imagery. *IEEE Trans. Geosci. Remote Sens.* 58, 826–840.
- Sun, Y., Qin, Q., Ren, H., Zhang, Y., 2022b. Decameter Cropland LAI/FPAR Estimation From Sentinel-2 Imagery Using Google Earth Engine. *IEEE Trans. Geosci. Remote Sens.* 60, 4400614.
- Sun, Y., Wang, B., Zhang, Z., 2023. Improving Leaf Area Index Estimation With Chlorophyll Insensitive Multispectral Red-Edge Vegetation Indices. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 16, 3568–3582.
- Sytar, O., Brestic, M., Zivcak, M., Olsovska, K., Kovar, M., Shao, H., He, X., 2017. Applying hyperspectral imaging to explore natural plant diversity towards improving salt stress tolerance. *Sci. Total Environ.* 578, 90–99.
- Uddling, J., Gelang-Alfredsson, J., Piikki, K., Pleijel, H., 2007. Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings. *Photosynth. Res.* 91, 37–46.
- Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified Optical-Thermal Four-Stream Radiative Transfer Theory for Homogeneous Vegetation Canopies. *IEEE Trans. Geosci. Remote Sens.* 45, 1808–1822.
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS J. Photogramm. Remote Sens.* 108, 273–290.
- Verrelst, J., Vicent, J., Rivera-Caicedo, J.P., Lumbierres, M., Morcillo-Pallarés, P., Moreno, J., 2019. Global sensitivity analysis of leaf-canopy-atmosphere RTMs: implications for biophysical variables retrieval from top-of-atmosphere radiance data. *Remote Sens. (Basel)* 11, 1923.
- Waldeland, A.U., Trier, Ø.D., Salberg, A.-B., 2022. Forest mapping and monitoring in Africa using Sentinel-2 data and deep learning. *Int. J. Appl. Earth Obs. Geoinf.* 111, 102840.
- Weiss, M., Baret, F., Jay, S., 2020. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER. Wu, C., Niu, Z., Tang, Q., Huang, W., 2008. Estimating chlorophyll content from hyperspectral vegetation indices: modeling and validation. *Agric. For. Meteorol.* 148, 1230–1241.
- Xiao, J., Fisher, J.B., Hashimoto, H., Ichii, K., Parazoo, N.C., 2021. Emerging satellite observations for diurnal cycling of ecosystem processes. *Nature Plants* 7, 877–887.
- Xie, Q., Dash, J., Huete, A., Jiang, A., Yin, G., Ding, Y., Peng, D., Hall, C.C., Brown, L., Shi, Y., Ye, H., Dong, Y., Huang, W., 2019. Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* 80, 187–195.
- Xu, M., Liu, R., Chen, J.M., Liu, Y., Shang, R., Ju, W., Wu, C., Huang, W., 2019. Retrieving leaf chlorophyll content using a matrix-based vegetation index combination approach. *Remote Sens. Environ.* 224, 60–73.
- Xu, M., Liu, R., Chen, J.M., Liu, Y., Wolanin, A., Croft, H., He, L., Shang, R., Ju, W., Zhang, Y., He, Y., Wang, R., 2022. A 21-year time-series of global leaf chlorophyll content maps from MODIS imagery. *IEEE Trans. Geosci. Remote Sens.* 1.
- Yang, A., Zhong, B., Hu, L., Wu, S., Xu, Z., Wu, H., Wu, J., Gong, X., Wang, H., Liu, Q., 2020. Radiometric Cross-Calibration of the Wide Field View Camera Onboard GaoFen-6 in Multispectral Bands. *Remote Sens. (Basel)* 12, 1037.
- Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J.M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., Xiao, J., Asrar, G.R., Chen, M., 2022. Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nat. Rev. Earth Environ.* 3, 477–493.
- Zhang, Y., Hui, J., Qin, Q., Sun, Y., Zhang, T., Sun, H., Li, M., 2021. Transfer-learning-based approach for leaf chlorophyll content estimation of winter wheat from hyperspectral data. *Remote Sens. Environ.* 267, 112724.
- Zhang, Y., Qin, Q., Ren, H., Sun, Y., Li, M., Zhang, T., Ren, S., 2018. Optimal Hyperspectral Characteristics Determination for Winter Wheat Yield Prediction. *Remote Sens. (Basel)* 10, 2015.