Improving the Retrieval of Forest Canopy Chlorophyll Content From MERIS Dataset by Introducing the Vegetation Clumping Index

Qi Sun, Quanjun Jiao, Liangyun Liu, Xinjie Liu, Xiaojin Qian, Xiao Zhang, and Bing Zhang, *Fellow, IEEE*

***Abstract*—The accurate retrieval of canopy chlorophyll content (CCC) is essential to the effective monitoring of forest productivity, and environmental stress. However, the clumping index (CI), a vital canopy structural parameter, affects inaccurate remote sensing of forest CCC. In this article, we proposed a concept of effective CCC (CCC_e) and an integrated CI approach to retrieving forest CCC using empirical regression and random forest regression. First, the PROSPECT-D and four-scale models were used to simulate forest canopy spectra, and the forest CCC_e including CI was found more feasible to be remotely sensed than the CCC. Then, an empirical regression model and random forest model trained using different combinations of the medium resolution imaging spectrometer (MERIS) terrestrial chlorophyll index (MTCI), reflectance, and CI values were used to estimate the CCC. Finally, the proposed approach was tested using satellite-based CI and MERIS product. Using the empirical regression model, the results showed that the retrieval of forest CCC using the MTCI was greatly improved by the inclusion of the CI (RMSE from 63.64 to 36.51 μg cm^{-2} for broadleaf; RMSE from 96.02 to 58.49 μg cm^{-2} for coniferous). Using the random forest approach and the model trained using the reflectance in red and red-edge bands, MTCI, and CI performed best, with RMSE = 27.95 μg cm^{-2} for broadleaf and RMSE = 34.83 μg cm^{-2} for coniferous. Overall, it is concluded that include the CI, particularly the approach using forest random regression, have the potential for satellite-based forest CCC mapping at regional and global scales.**

***Index Terms*—Clumping index (CI), forest chlorophyll, four-Scale, medium resolution imaging spectrometer (MERIS), PROSPECT-D, random forest.**

I. Introduction

CHLOROPHYLL converts solar radiation into stored chemical energy during photosynthesis. It is an important ecological variable and also an indicator of vegetation physiological activity [1]–[3]. Vegetation chlorophyll contains most of the incorporated foliar nitrogen, which is highly correlated with the foliar photosynthetic rubisco capacity [4], [5]; moreover, most of the photosynthetically active radiation absorbed by the chlorophyll in the vegetation canopy is used for photochemical conversion during photosynthesis [6]. The canopy chlorophyll content (CCC) is determined by the leaf area index (LAI) and leaf chlorophyll content (LCC), which are expressed per unit leaf area [7], [8], as

$$\mathrm{CCC} = \mathrm{LCC} \times \mathrm{LAI}. \tag{1}$$

Accurate CCC information is essential to assessments of the gross primary production of vegetation ecosystems [3], [9]–[11].

The use of remote sensing for extracting the spatial and temporal details of vegetation CCC has proved to be superior to ground sampling [1], [12], [13]. In recent decades, empirical regression, in particular linear regression, has been the most popular and simplest technique used to relate CCC to spectral characteristics such as the reflectance and spectral vegetation index. Compared with the red and near-infrared (NIR) regions, the red-edge region of the canopy spectrum is more closely related to the vegetation CCC [12]. Several chlorophyll-related indices, especially one that use red-edge spectral characteristics, have been developed and demonstrated to show promise for making estimates of crop and forest CCC [1], [14]–[16]. In general, the application of linear regression methods is limited to by the underrepresentation of training samples [17], [18], and it is a great challenge to obtain sufficient field samples that include seasonal and spatial changes in canopy chlorophyll and structure for use in the construction of statistical regression models [19].

To increase the representativeness of training data, a hybrid inversion method combing linear regression and canopy spectral simulation based on radiative transfer models (RTMs) has been used for estimating vegetation chlorophyll status [17], [18]. RTMs can better describe complex canopies as well as leaf spectra for different canopy structures, leaf parameters, and soil backgrounds [3]. RTMs of vegetation at the leaf level [20] and canopy level [21] have been used to simulate canopy

Manuscript received February 8, 2021; revised March 26, 2021; accepted May 18, 2021. Date of publication May 21, 2021; date of current version June 8, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFA0603001 and in part by the National Natural Science Foundation of China under Grant 42071330 and Grant 41825002. (*Corresponding author: Quanjun Jiao.*)

Qi Sun is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China (e-mail: tbp150202003z@student.cumtb.edu.cn).

Quanjun Jiao, Liangyun Liu, Xinjie Liu, Xiaojin Qian, Xiao Zhang, and Bing Zhang are with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: jiaoqj@radi.ac.cn; liuly@radi.ac.cn; liuxj@radi.ac.cn; qianxj@radi.ac.cn; zhangxiao01@aircas.ac.cn; zb@radi.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3082621

reflectance and to retrieve vegetation parameters including vegetation chlorophyll [3], [22]–[24]. However, the direct inversion of physical RTMs to retrieve vegetation parameters has the disadvantage of a large workload and being slow, and the look-up tables (LUTs) method suffers from huge amount of high-dimension data in the LUT [17], [18]. The hybrid inversion approach, which has the advantages of including a large amount of training data from RTMs and a high computational efficiency for the empirical regression has been considered [19], [25], [26]. However, as applied to the remote sensing of vegetation parameters, linear regression is still site-specific and does not account for complex situations where the modeled relationship is not linear.

Advanced nonlinear regression methods have a better learning ability than linear regression methods, but require a larger sample size [11]. Therefore, in the hybrid inversion approach, RTMs are used to provide large amounts of simulated spectral data corresponding to different vegetation conditions to train those advanced nonlinear regression models [27]. In recent years, machine-learning methods have been widely used to build models for retrieving physicochemical parameters of vegetation and have improved the prediction accuracy and generalization of these models [28], [29]. Because of its proven accuracy and stability, random forest regression, which is as an important nonparametric ensemble algorithm, has recently been used to retrieve crop CCC [30], [31] and forest CCC [32], [33].

Prior knowledge is increasingly recognized as an important criterion for the quantitative remote sensing of vegetation parameters [34], [35]. The foliage clumping effect, which is characterized by the clumping index (CI), are one important canopy structural factor that influence the interception, absorption, and transmission of photosynthetic-effective radiation in vegetation, especially in the case of forests [36]. Variation in CI strongly affects canopy reflectance and reflectance-based vegetation indices [37], [38]. Therefore, retrieval of the CCC based on the reflectance or vegetation indices should mainly account for the effect of the CI. Early satellite CI products were estimated based on polarization and directionality of the earth's reflectances with limited temporal frequency, and spatial resolution [39]. Several recent global satellite CI products have been derived from the moderate resolution imaging spectroradiometer (MODIS) bidirectional reflectance distribution function (BRDF) product [40]–[42], and these CI products represent not only the differences between different vegetation types, but also temporal and spatial differences for vegetation of the same type [38], [43]. It has been proved that satellite-based CI can improve the estimating accuracy of the LAI [44] and the fractional vegetation cover [45]. It may be useful to use prior knowledge of the CI to improve the remote sensing of forest CCC, however, there has been little research on using satellite CI products for forest CCC estimation.

Given that the medium resolution imaging spectrometer (MERIS) produced the first red-edge dataset with global coverage and long-term series, the objective of this article was to evaluate the effect of the CI on the MERIS terrestrial chlorophyll index (MTCI) related to CCC using simulations from the PROSPECT-D and four-scale geometrical-optical model, and to develop and test hybrid approaches for estimating forest CCC from MERIS data using random forest regression and traditional empirical regression that take into account the foliage clumping effect by including prior knowledge of the CI.

II. Data and Methods

A. Simulated Dataset

To simulate the remotely sensed spectra of forests that are spatially heterogeneous and that have clumped structural characteristics with varying CCC, the leaf RTM PROSPECT-D [20] and the four-scale geometrical-optical model [21] were used in this article. Previous studies have demonstrated that the PROSPECT model [46] is suitable for a large number of vegetation species and plant functional types (PFTs), including coniferous forest (CNF) [47]–[49]. As given in Table I, the PROSPECT-D model adds another input parameter—the leaf anthocyanin content—to the seven parameters in the PROSPECT model that are used to simulate leaf reflectance and transmittance. The four-scale model is a geometric-optical model that takes into account the structural composition of forest canopies at different scales; this includes nonrandomly distributed crowns; foliage clumped into crowns; foliage clumped into branches; and foliage clumped into shoots [21]. As this model encapsulates the clumping phenomenon, it provides a more realistic description of the forest canopy.

First, the PROSPECT-D model was used to simulate leaf reflectance spectra for use as input parameters (leaf optical properties) to the four-scale model. The leaf parameter values used in PROSPECT-D were based on the ranges reported in [47]–[49]; the values used are given in Table I. Next, canopy reflectance were simulated based on the four-scale model using the leaf optical properties, forest background reflectance spectra, and the tree architecture, canopy structure, and imaging geometry as inputs. Table II gives the parameters that were used as inputs to the four-scale model.

Based on the specified input parameters, the 4-Scale model generates the canopy reflectance (ρ) as a linear sum of four components

$$\rho = \rho_{PT\lambda} F_{PT} + \rho_{ZT\lambda} F_{ZT} + \rho_{PG\lambda} F_{PG} + \rho_{ZG\lambda} F_{ZG} \quad (2)$$

where $\rho_{PT\lambda}$, $\rho_{ZT\lambda}$, $\rho_{PG\lambda}$, and $\rho_{ZG\lambda}$ are the reflectance factors corresponding to the sunlit vegetation, shaded vegetation, sunlit ground, and shaded ground; respectively, F_{PT}, F_{PG}, F_{PG}, and F_{ZT} represent the probabilities of viewing the respective scene components [50]. The CI for the landscape can be quantified by a vegetation dispersion parameter using a modification of Beer's Law in the direction of the zenith angle, θ [36], [51]

$$P(\theta) = e^{-G(\theta) L \Omega / \cos\theta} \quad (3)$$

where $P(\theta)$ is the probability that a beam of light at a zenith angle θ will be transmitted through the canopy, i.e., the gap fraction; G is the mean projection of the unit foliage area in the same direction, for which normally the value is 0.5; L is the LAI, and Ω is the CI for the landscape. To avoid the variation of CI with

TABLE I
INPUT PARAMETERS USED IN THE PROSPECT-D MODEL. PFT; LCC; N = STRUCTURAL PARAMETER; CAR = CAROTENOID CONTENT; C_{ANT} = LEAF ANTHOCYANIN CONTENT; C_{b} = BROWN MATTER; C_{m} = DRY MATTER CONTENT; AND C_{w} = EQUIVALENT WATER THICKNESS

PFT	LCC (μg cm^{-2})	N (unitless)	Car (μg cm^{-2})	C_{Ant} (μg cm^{-2})	C_b	C_m (g cm^{-2})	C_w (g cm^{-2})	Reference
Broadleaf forest	10-80	1.2	LCC/7	1	0	0.005	0.01	[47]
Coniferous forest	10-80	2.5	LCC/7	1	0	0.035	0.048	[48], [49]

TABLE II
INPUT PARAMETERS USED IN THE FOUR-SCALE MODEL

Parameters	Broadleaf forest	Coniferous forest
Stand density (trees/ha)	1000, 2000, 3000, 4000, 6000, 8000, 12000	1000, 2000, 3000, 4000, 6000, 8000, 12000
Leaf area index (m^2 m^{-2})	0.5, 1, 2, 4, 6, 8	0.5, 1, 2, 4, 6, 8
Crown base height (m)	1, 5, 10	1, 5, 10
Crown vertical dimension (m)	5, 10, 20	5, 10, 20
Crown radius (m)	0.75, 1, 1.25, 1.5	0.5, 0.75, 1, 1.25
Crown shape	Spheroid	Cone & cylinder
Clumping index of the shoots (ΩE)	0.6, 0.9	0.5, 0.8
Solar zenith angle (°)	10–70°; step 10°	10–70°; step 10°
View zenith angle (°)	0–60°; step 10°	0–60; step 10°
Relative azimuth angle (°)	0°, 180°	0°, 180°
Neyman grouping	1, 2, 3	1, 2, 3
Needle to shoot ratio (γE)	1	1.41
Background composition	Green vegetation and soil	Green vegetation and soil

θ, an average CI [39] can be used

$$\mathrm{CI} = \mathrm{LAI_e}/\mathrm{LAI} \tag{4}$$

where $\mathrm{LAI_e}$ is the effective LAI computed using the following equation [52]:

$$\mathrm{LAI_e} = -2\int_0^{\frac{\pi}{2}} \ln\left[P(\theta)\right]\cos\theta\sin\theta\,\mathrm{d}\theta. \tag{5}$$

The canopy reflectance and CI generated by the four-scale model were used in this article. It should be noted that the canopy reflectance spectra derived from the model were resampled to match the corresponding band of the MERIS based on its spectral response function. In total, 5 778 864 simulated samples were generated for broadleaf forest (BOF) and 8 113 560 samples generated for CNF. Due to the enormous amount of data, one percent of the dataset was randomly selected to train the retrieval model using linear regression and random forest regression.

TABLE III
SPECIFICATIONS OF THE MERIS SENSOR

Band	Band center (nm)	Band width (nm)
B1	412.5	10
B2	442.5	10
B3	490	10
B4	510	10
B5	560	10
B6	620	10
B7	665	10
B8	681.25	7.5
B9	708.75	10
B10	753.75	7.5
B11	760.625	3.75
B12	778.75	15
B13	865	20
B14	885	10
B15	900	10

B. Satellite Data

1) MERIS Surface Reflectance Data: Data from the MERIS sensor carried on the ENVISAT satellite provide multispectral data including red-edge bands that provide the highest available revisit frequency; these data were, therefore, used for the test carried out in this article. MERIS is a medium-spectral resolution imaging spectrometer, which samples the surface reflectance in fifteen spectral bands with a temporal revisit time of two–three days and a spatial resolution of 300 m. Detailed specifications of the MERIS sensor are given in Table III. The full-resolution (FR) surface reflectance product produced by ESA which is a seven-day synthesis of data collected at the original two–three days revisit frequency was used. In the seven-day product, the reflectance is normalized to the nadir view, and the solar zenith angle corresponds to that at 10: 00 local time on the median day of the compositing period. The product consists of 13 bands, removing bands 11 and 15. The MERIS FR surface reflectance product covers the period from 2003 to 2012.

2) LAI Product: The MODIS LAI product (MCD15A3H V6 Level4) is derived from MODIS data and has a spatial resolution of 500 m and a 4-day temporal resolution. The algorithm used to produce the LAI chooses the best pixel available from all the acquisitions made by the two MODIS sensors located on NASA's Terra and Aqua satellites within a given 4-day period

TABLE IV
DETAILS OF THE FIELD MEASUREMENTS USED IN THIS ARTICLE. THE SAMPLING LOCATION (LAT = LATITUDE, LONG = LONGITUDE), SAMPLING YEAR, DOMINANT SPECIES, PFT, CNF, BOF, MEAN LAI VALUE, AND MEAN LCC VALUE ARE SHOWN. THE COMMUNITY-LEVEL CCC CAN BE OBTAINED BY MULTIPLYING THE LCC BY THE LAI

Site name	County	Sampling years	Lat/Long	PFT	Dominant species	Mean LAI	Mean LCC ($\mu g\ cm^{-2}$)	Reference
Sudbury_Zhang	Canada	2003	47.16, -81.74	CNF	Black spruce	3.8*	29.0	[55]
Sudbury_Zhang	Canada	2004	47.16, -81.74	CNF	Black spruce	3.7*	32.1	[55]
Haliburton	Canada	2004	45.24, -78.54	BOF	Sugar maple	2.8*	38.3	[56]
Sudbury_Simic	Canada	2007	47.16, -81.71	BOF	Trembling aspen	2.3*	48.8	[54]
Sudbury_Simic	Canada	2007	47.18, -81.74	CNF	Black spruce	4.3*	29.0	[54]
Borden	Canada	2013	44.32, -79.93	BOF	Red maple	4.0*	36.6	[10]

LAI*: field measurement; LAI*: MODIS LAI.

[53]. The MODIS LAI is defined as the one-sided green leaf area per unit ground area in the case of broadleaf canopies and as half of the total needle surface area per unit ground area in the case of coniferous canopies.

3) CI Product: Depending on a more robust hotspot corrected BRDF model to reconstruct the hotspot effect and a backup algorithm to estimate CI for outlying pixels, a global CI product was derived from MODIS BRDF parameters based on CI–NDHD (the normalized difference between hotspot and dark spot) equations with 500 m spatial resolution and 30-day temporal intervals [41], and was used in this article.

C. Validation Sites and Data

Three article sites located in Sudbury, Haliburton, and Borden, Ontario, Canada were used in this article. Field sampling of black spruce and aspen was carried out in 2003/2004/2007 (black spruce coniferous forest) and 2007 (aspen BOF) at the Sudbury site [54], [55]. The sugar maple (broadleaf) forest at Haliburton was sampled in 2004 [56] and red maple (broadleaf) forest at Borden was sampled in 2013 [10]. The *in situ* LCC and LAI were measured to validate the modeled results based on the MERIS data. It should be noted that the LAI ground-based measurements made at Sudbury in 2003/2004 and Borden site in 2013 were field measurements, while in 2007 at Sudbury and 2004 at Haliburton, MODIS LAI data were used due to the lack of simultaneously measured LAI values matched with the field-measured LCC. Meanwhile, in order to reduce the impact of spatial heterogeneity, the Landsat 5 TM 16-day NDVI composite product was employed to upscale the spatial resolution of the ground-based LAI measurements to that of medium-resolution data (500 m). First, the NDVI values for 30-m and 500-m pixels were obtained and a functional relationship between them was established. The LAI at a resolution of 500 m was then calculated by applying this function to the field-measured LAI data.

Table IV gives details of the validation data acquired at each field location. These data included 20 BOF measurements and 28 coniferous forest (CNF) measurements. The ground-based observations at Borden site in 2013 was used to match MERIS data in the previous year, due to the stability of vegetation seasonal changes at the same site between adjacent years [47]. Because available field-measured forest CCC is relatively scarce, the Borden data is also used in this article. The average 500-m CCC was obtained by multiplying the measured LCC by the corresponding LAI. Using a bilinear resampling method implemented in ENVI software (Exelis VIS, Inc., Boulder, CO, USA), the MERIS data were resampled to spatial resolution of 500-m—the same as the MODIS data—to be consistent with that of the validation data.

D. Algorithm Development

Two distinct approaches were undertaken to retrieve the CCC by including the CI. The workflow of the proposed approaches developed to retrieve forest CCC is shown in Fig. 1. First, the canopy reflectance and CI were generated by the four-scale geometrical-optical model. Next, by using MATLAB software (The Math Works, Natick, MA, USA), empirical regression and the random forest method were employed to build CCC retrieval models involving CI based on the simulated datasets. Finally, the MERIS satellite data were used to validate the feasibility of the models that had been constructed. In order to validate the feasibility of the retrieval models that did and did not include CI, this article used the coefficient of determination (R^2), and the bias (Bias), and the root mean square error (RMSE) as metrics to quantify the accuracy of the models.

1) Empirical Regression Approach Including CI: The commonly used MTCI, which has been proven to be closely related to the CCC [14], [24], was selected for estimating the CCC using an empirical regression approach. The MTCI is calculated from the red and red-edge bands and is quantified by the following equation, where R_{ijk} is the reflectance at the *ijk*th wavelength [14]

$$\mathrm{MTCI} = \frac{R_{753.75} - R_{708.75}}{R_{708.75} - R_{681.25}}. \tag{6}$$

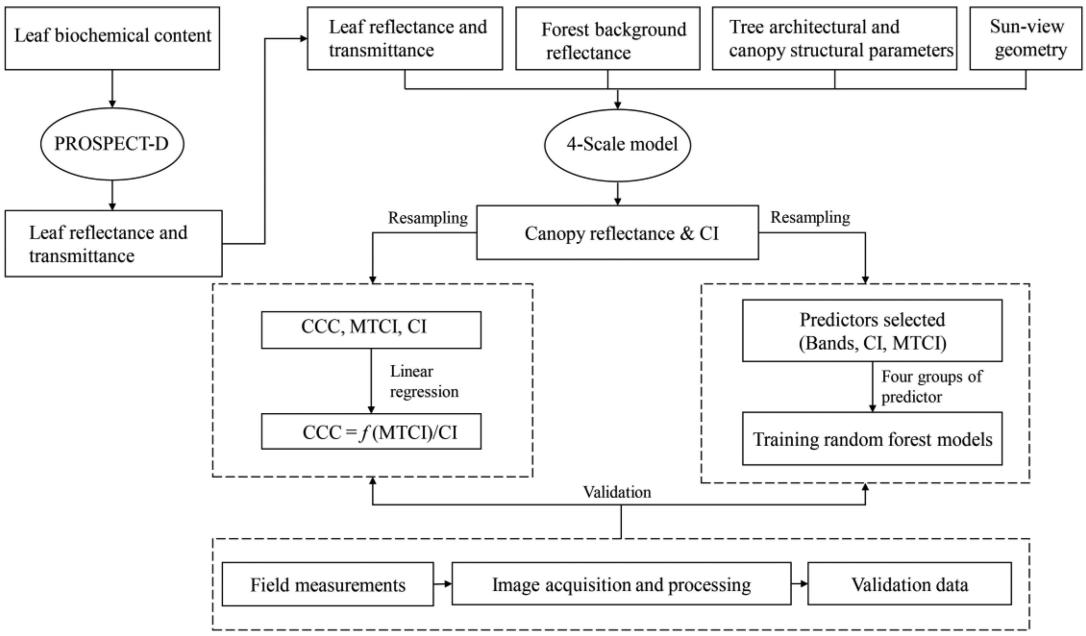


Fig. 1. Workflow of the proposed approach that was developed to predict forest CCC.

The linear regression of the MTCI against the CCC was analyzed. However, the variation in forest canopy CI may mean that there were still variations in the MTCI even for the same CCC value.

To reduce the influence of the vegetation canopy structure on the estimates of forest CCC using only the MTCI, an improved MTCI-based empirical method that included the CI was developed in this article. Inspired by the idea that the effective LAI is strongly related to LAI-related VIs, the concept of the effective CCC (CCC_e), which included the LCC and LAI_e, was introduced. The LAI_e, expressed as (4) and (5), is based on gap-fraction theory, which eliminates the need to assume a random distribution of foliage [57]. We defined CCC_e as

$$\mathrm{CCC_e} = \mathrm{LCC} \times \mathrm{LAI_e} = \mathrm{CCC} \times \mathrm{CI}. \tag{7}$$

CCC_e can also be considered as including CCC and CI. Chlorophyll-related vegetation indices such as the MTCI should be more strongly dependent on CCC_e than on the CCC. Linear regressions were carried out using the MTCI as the independent variable and CCC_e as the dependent variable

$$\mathrm{CCC_e} = f(\mathrm{MTCI}). \tag{8}$$

The estimated CCC_e can be converted into the CCC by using the CI according to (7) and (8). This improved empirical regression model of CCC retrieval that introduces prior knowledge of the CI, and then be expressed as

$$\mathrm{CCC} = \mathrm{CCC_e}/\mathrm{CI} = \frac{f(\mathrm{MTCI})}{\mathrm{CI}}. \tag{9}$$

Implementing the Random Forest Regression Approach: Random forest regression is an advanced machine learning technique and was used to train narrowband MERIS spectral data for making estimates of the CCC in this article. Random forest is a nonparametric nonlinear statistical ensemble bagging method that employs recursive partitioning of data into increasingly homogeneous subsets, called regression trees, and averages the results for all of the trees. In this article, the random forest regressions were trained using 100 trees for the training set. In previous research, red and red-edge bands were selected as predictors for random forest regression model [32], [33]. However, the reflectance of these bands is strongly affected by the canopy structure, which makes it difficult to distinguish the characteristics of some types of vegetation, especially forest. The CI is a vital structural parameter of vegetation canopies and the MTCI is sensitive to chlorophyll at the leaf and canopy levels [14], [24], [39], [41]. Therefore, in this article, an attempt was made to include CI and MTCI to improve the ability of the retrieval model with use of only red and red-edge bands. As given in Table V, four types of predictor variable combinations were selected. Six red and red-edge bands including bands–10 were selected as input features—known as Refl-basic predictors—for the CCC estimation.

III. Results

A. Influence of the CI on the spectral reflectance, the MTCI and the CCC retrieval model

In this article, the correlation between the CI, the MERIS spectral bands (bands 8–10) and the MTCI using simulations produced by the 4-Scale model (see Fig. 2), which were generated for a fixed LCC (LCC = 40 μg cm^{-2}) and LAI (LAI = 4) for nadir observations (solar zenith angle = 30°), was investigated. As shown in Fig. 2, the CI is significantly correlated with band 8 of MERIS and the MTCI. In addition, the correlation between the CI and the MERIS bands appears to more significant for BOF than for coniferous forest. Surprisingly, the CI has relatively little

TABLE V
RANDOM FOREST TRAINING PROGRAMS

PFT	The predictor variables			
Broadleaf forest	Refl	Refl _MTCI	CI_ Refl	CI_ Refl _MTCI
Coniferous forest	Refl	Refl _MTCI	CI_ Refl	CI_ Refl _MTCI

Refl: MERIS reflectance in the six red and red–edge bands (bands 5–10); Refl_MTCI: combined predictors for the six red and red-edge reflectance (bands–10) and the MTCI; CI_Refl: CI, bands 5–10 combined predictors.

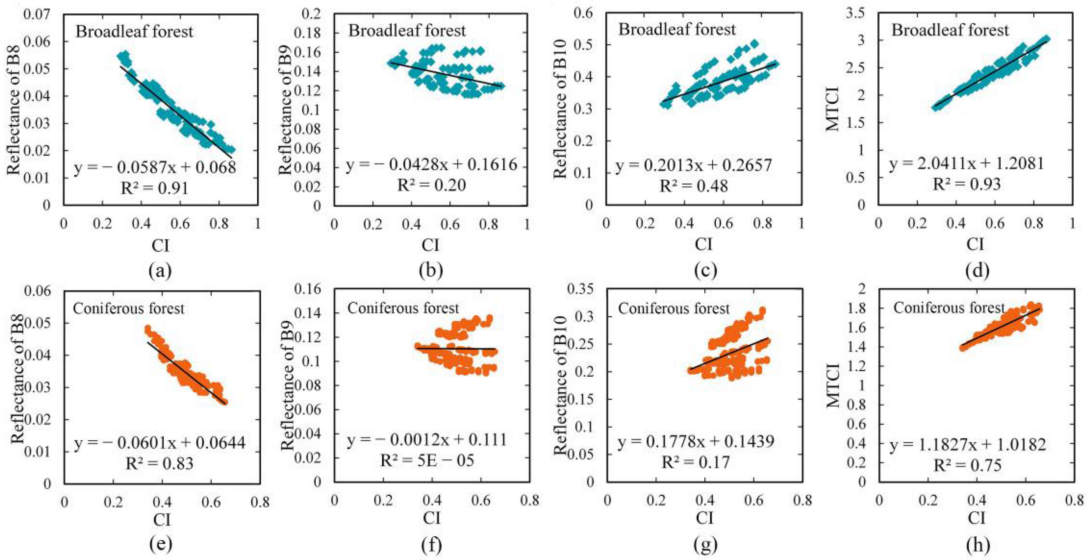


Fig. 2. Relationship between CI and MTCI and CI and the red and red-edge reflectance for broadleaf and coniferous forest types. The red and red-edge bands used to calculate the MTCI were bands 8–10.

effect on bands 9 and 10, especially in the case of coniferous forest.

To explore the role of the CI in the CCC linear regression model, the relationships between the MTCI and CCC with and without the inclusion of the CI were analyzed using a simulated dataset from the 4-Scale model (see Section II-A). Scatter plots of CCC against MTCI and CCC_e against MTCI for BOF [see Fig 3(a) and (c)] and coniferous forest [see Fig. 3(b) and (c)] are shown in Fig. 3. Fig. 3 shows that the correlation between CCC_e and MTCI is obviously better than that between CCC and MTCI, especially for BOF. In detail, the scatter points in Fig. 3(a) and (b) are clearly banded according to the CI values. If the previously established linear regression model had been used, the CCC would be overestimated for points with low CI values and underestimated for high CI values. In Fig. 3(c) and (d), scatter points are closely grouped around the regression line, and the points corresponding to different CI values are clustered together. Taking these results together, it can be seen that if accurate prior knowledge of the CI is integrated into the CCC_e–MTCI relationship, the influence of foliar clumping in the canopy on the MTCI-based chlorophyll estimates can be reduced.

Fig. 4 shows estimates of the importance of the different predictors in the random forest regression model for both broadleaf and coniferous forest. Of the six MERIS bands, band 10 is the most important. With the inclusion of the MTCI, the importance is separated by MTCI, as the MTCI is calculated using bands 8–10. Once the CI is included, the CI is the most important of all the predictors. Therefore, in the random forest regression model, the CI is the most important predictor for CCC estimation

B. Validation of the Improved Empirical Regression Model Using MERIS Satellite Data

The transferability of an algorithm across spatial and temporal scales is essential for large-scale retrieval of the CCC or any ecological variable. MERIS satellite data were used to validate the accuracy of the CCC retrieval models. According to the formulations of the CCC_e–MTCI relationship shown in Fig. 3, the retrieval models that include CI can be expressed as $CCC = (121.9 \times MTCI - 67.34)/CI$ for BOF and $CCC = (61.02 \times MTCI - 33.17)/CI$ for coniferous forest. Fig. 5 shows the results of the validation of the empirical regression models for CCC retrieval using MERIS data. The results show that MTCI-based empirical regression model for CCC estimation has large limitations and generally overestimates the CCC, giving an RMSE of over 60 $\mu g\ cm^{-2}$ for either BOF or coniferous forest. However, when the CI is included, the accuracy of the CCC retrieval is significantly improved. In detail, the MTCI-based regression model with the CI has relatively better performance

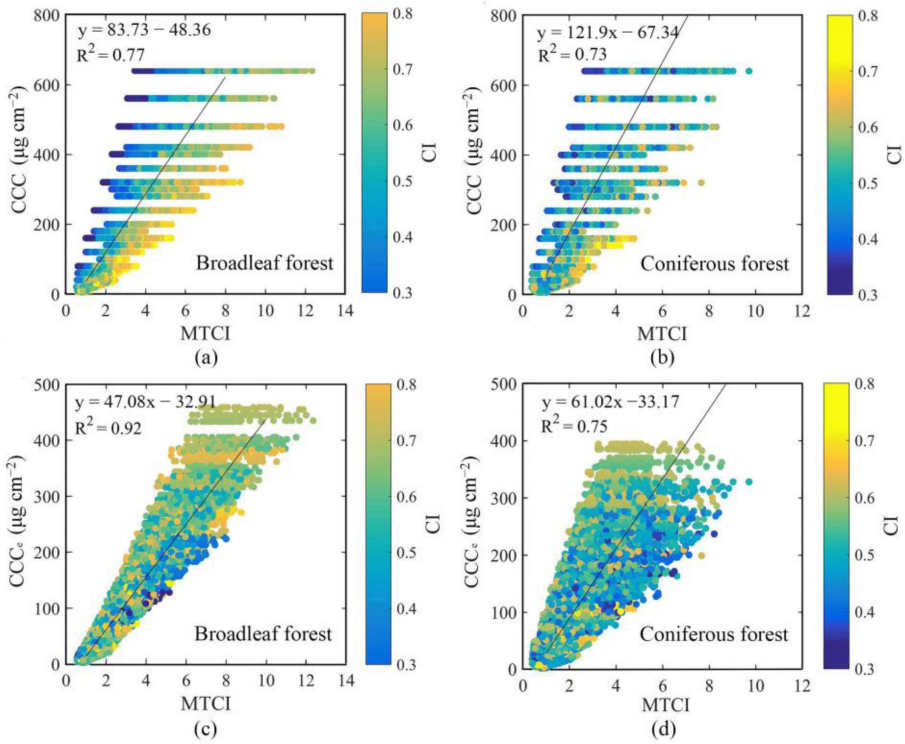


Fig. 3. Influence of variations in the CI on the MTCI–CCC and MTCI–CCC_e relationships for broadleaf and coniferous forest types.

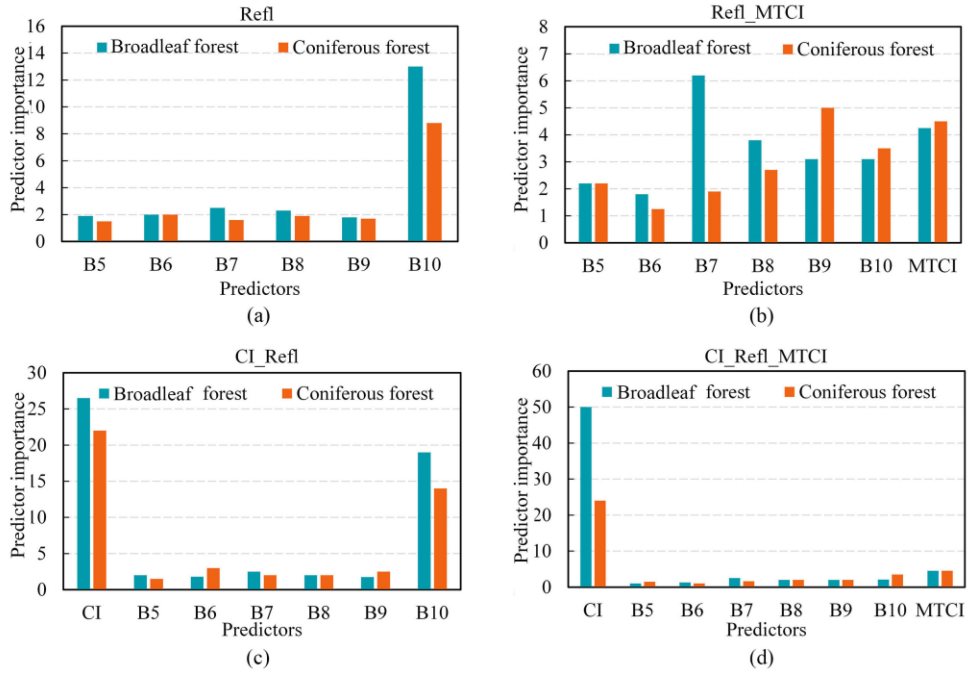


Fig. 4. Predictor importance for CCC estimates using random forest regression ensembles for broadleaf and coniferous forest types.

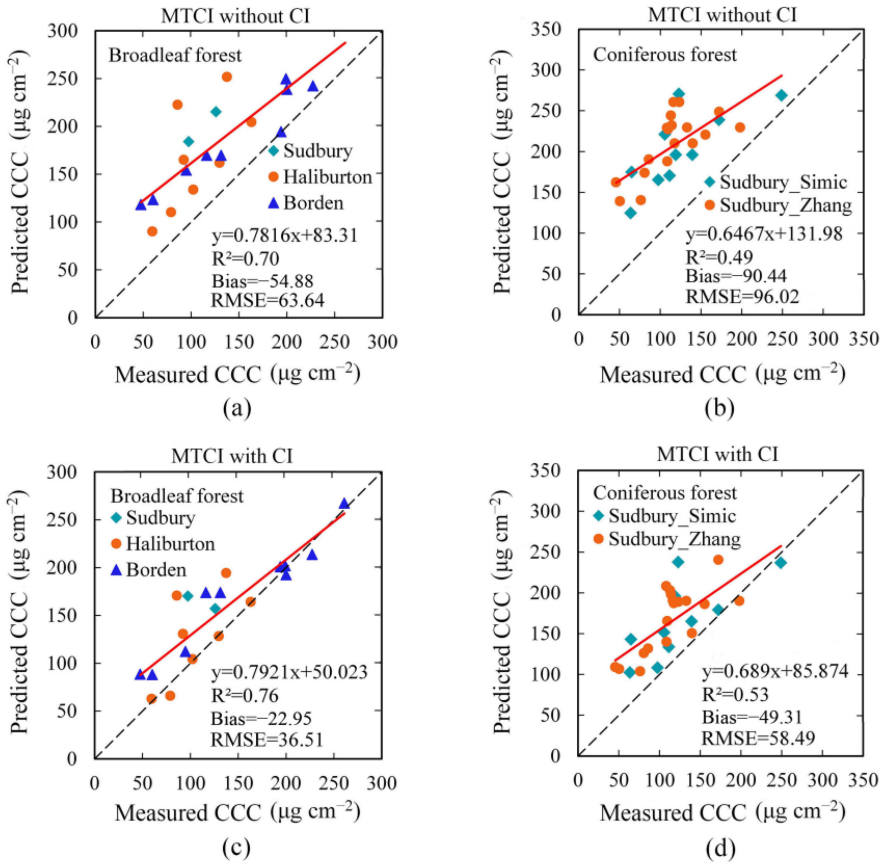


Fig. 5. Validation of empirical regression models for estimating forest CCC using MERIS satellite data. (a) and (b) MTCI-based retrieval models without the CI for broadleaf and coniferous forest, respectively. (c) and (d) MTCI-based retrieval models with the CI for broadleaf and coniferous forest, respectively.

($R^2 = 0.76$, Bias $= -22.95$ μg cm^{-2}, and RMSE $= 36.51$ μg cm^{-2}) compared to the original model without the CI ($R^2 = 0.70$, Bias $= -54.88$ μg cm^{-2}, and RMSE $= 63.64$ μg cm^{-2}) for broadleaf trees, and there are also an improvement ($R^2 = 0.53$, Bias $= -49.31$ μg cm^{-2}, and RMSE $= 58.49$ μg cm^{-2}) compared to the original one ($R^2 = 0.49$, Bias $= -90.44$ μg cm^{-2}, and RMSE $= 96.02$ μg cm^{-2}) for coniferous trees. These results imply that the proposed empirical regression model for retrieving the CCC that includes the CI can be successfully used with MERIS satellite data.

C. Validation of the Random Forest Model Using MERIS Satellite Data

The random forest model was trained and run using six red and red-edge bands, and then the spectral bands combined with the MTCI and CI as input predictors for the CCC. The results for the random forest model, showing the relationships between the predicted CCC and measured CCC for broadleaf and coniferous forest, are presented in Figs. 6 and 7, respectively. As shown in these figures, the performance of the retrieval models trained with predictors that included the CI was generally better than those that did not include the CI. For BOF, the CI_Refl_MTCI-based model performed best and achieved an R^2 value of 0.77, a Bias value of 3.66 μg cm^{-2}, and an RMSE value of 27.95 μg cm^{-2} [see Fig. 6(d)]. The validation results for coniferous forest are similar to those for BOF. The best-performing model for coniferous forest was also the CI_Refl_MTCI-based model, which achieved an R^2 value of 0.61, a Bias value of -21.66 μg cm^{-2}, and an RMSE value of 34.83 μg cm^{-2}. Although the two random forest regression models that did not include the CI had RMSE values of over 30 μg cm^{-2} for broadleaf [see Fig. 6(a) and (b)] and coniferous forest [see Fig. 7(a) and (b)], they still clearly outperformed the empirical regression model based on the MTCI [see Fig. 5(a) and (b)]. Overall, these results suggest that including the CI effectively improves the accuracy of the retrieval models and that the random forest approach has great potential for making predictions of the CCC based on simulations made by the four-scale model.

IV. Discussion

A. Potential of Different Regression Models for Making CCC Estimates by Introducing the CI

The CI describes the random spatial distribution of foliage, which is an important structural parameter, especially for forest canopies [36]. As a result, for the same CCC, the canopy reflectance can be different due to the difference in canopy structure [55]. The relationship between vegetation indices and the effective LAI has been shown to be stronger than that

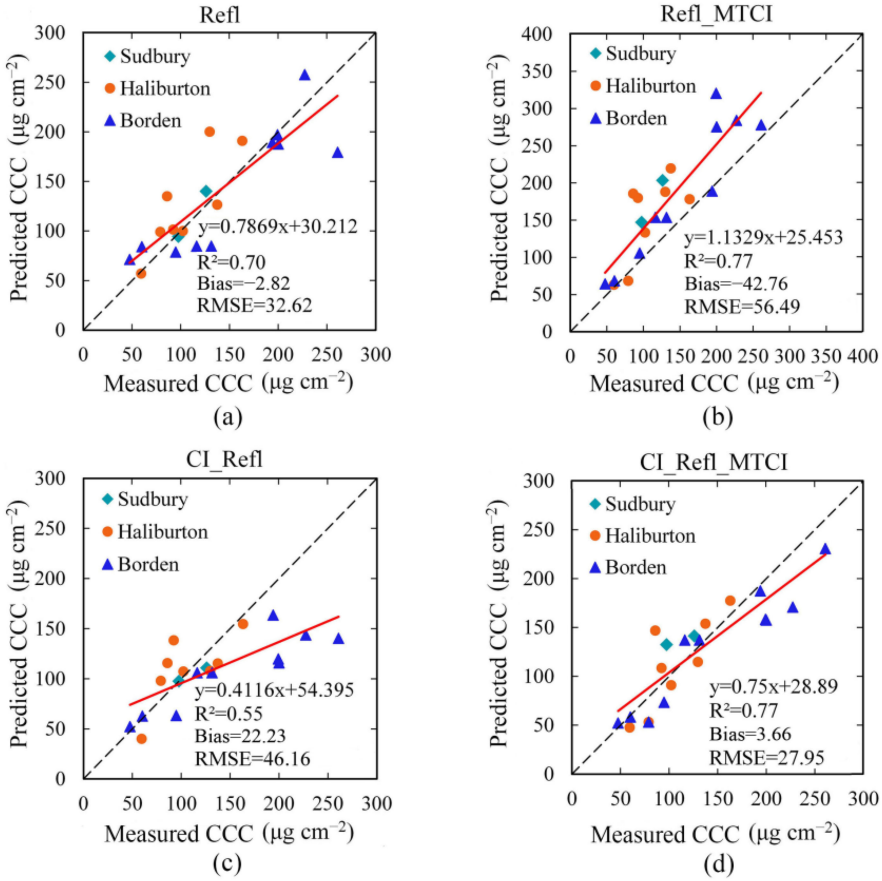


Fig. 6. Validation of CCC modeling based on MERIS satellite data for BOF. (a) Random forest trained using six red and red-edge bands. (b) Random forest trained using six red and red-edge bands and MTCI. (c) Random forest trained using CI and six red and red-edge bands. (d) Random forest trained using CI, six red and red-edge bands, and MTCI.

between vegetation indices and forest LAI [57], [58]. Similarly to these previous studies, in this article, it has been found that the variation in CI greatly affects the chlorophyll-related vegetation index, MTCI (see Fig. 3), and that the MTCI is more closely correlated with CCC_e than with CCC (see Fig. 3). These results show that the effect of the CI on the canopy reflectance and vegetation indices should be considered when making estimates of forest CCC. Previous studies have shown that incorporating the airborne CI minimizes the ill-posed nature of forest LCC estimates made by inverting the RTM using a LUT [55]. Our findings indicate that using satellite CI products as prior knowledge is, indeed, useful for making CCC estimates based on either empirical regression or random forest regression. This proposed MERIS-based CCC retrieval model involving CI also has potential for other red edge satellites, such as the successor of MERIS—ocean and land color instrument on sentinel-3 satellite.

In addition, the empirical regression model and random forest regression model performed differently in terms of precision and accuracy. Although the empirical regression model that included the CI performed better at making CCC estimates (RMSE = 36.51 μg cm^{-2} for BOF and 58.49 μg cm^{-2} for coniferous forest) than when only the MTCI was included, the weak predictive ability of this model resulted in RMSE values of over 35 μg cm^{-2}. The random forest model that included six spectral bands, MTCI and CI produced the best CCC estimates and had the smallest RMSEs for both BOF (27.95 μg cm^{-2}) and coniferous forest (34.83 μg cm^{-2}) due to its stronger learning and generalization ability for complex samples compared to linear regression [27], [33]. Thus, it is suggested that the random forest regression models that include satellite CI products are suitable for retrieving CCC. It is also appropriate to include more inputs in these models, including remote sensing features and prior knowledge.

B. Random Forest Performance Using Different Combinations of MERIS Bands

It is well known that certain, carefully selected spectral bands are more sensitive to the physical and chemical parameters of vegetation than all the bands together [59]–[61]. In this article, forest CCC retrieval models were trained using red and red-edge bands which have proved to be sensitive to the vegetation chlorophyll status at the canopy level [33], [62], [63]. Also, in order to make a comparison with our selected red and red-edge bands from the MERIS SR product, we also assessed the performance of retrieval models that were trained using all 13 available bands ($Refl_{all}$) and another six bands ($Refl_{cor}$: B5, B6, B8, B9, B10,

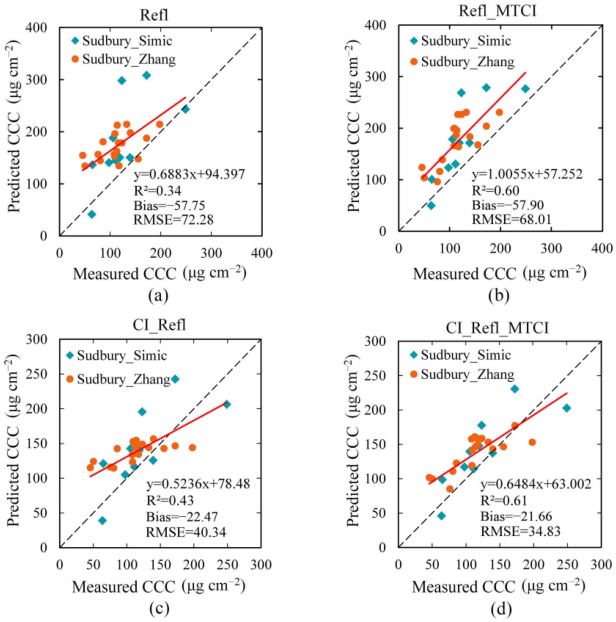


Fig. 7. Validation of CCC modeling based on MERIS satellite data for coniferous forest. (a) Random forest trained using six red and red-edge bands. (b) Random forest trained using six red and red-edge bands and MTCI. (c) Random forest trained using CI and six red and red-edge bands. (d) Random forest trained using CI, six red and red-edge bands, and MTCI.

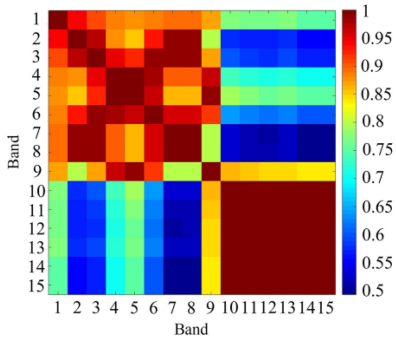


Fig. 8. MERIS band correlation matrix used for band selection.

and B13) that were selected based on the band-to-band correlation matrix (see Fig. 8). The band correlation matrix offers an overview of the statistical analysis of MERIS reflectance data for all possible combinations of wavelengths [64]. Based on these results, undesirable bands can be omitted because highly correlated bands generally contain similar information. Therefore, the Refl_{cor} that had low correlation with other bands, including B5, B6, B8, B9, B10, and B13, were compared with the six red and red-edge bands described in Section III. Table VI gives the performance achieved for broadleaf and coniferous forest using the three different training sets. Because the blue band of satellite images is greatly affected by aerosol, atmosphere correction in blue band is not accurate as red and near infrared band, which can affect the retrieval accuracy of models with the Refl_{all} set of predictor variables. When the Refl_{cor} set were used, it can be seen that the results for BOF were similar to those obtained using the Refl set of predictor variables proposed in this article. However, its performance was far inferior to that of the Refl set of predictor variables in the case of coniferous forest. The biggest discrepancy in the results between the Refl_{cor} and Refl sets occurred in the near infrared. The addition of another NIR band—B13, in the Refl_{cor} did not improve the overall accuracy of the CCC retrieval for the coniferous forest case, and the reason may be that the coniferous forest has more shadows components, which brings great uncertainty to the NIR reflectance [25]. In general, the misspecification of training parameters is the most important source of uncertainty. The use of more bands or combinations of vegetation indexes is something that can be tried in later studies.

C. Uncertainties in Modeling CCC

For our proposed models that included the CI, validation of the results of MERIS-based CCC estimates for BOF (best RMSE = 27.95 μg cm^{-2}) and coniferous forest (best RMSE = 34.83 μg cm^{-2}) were made for a total of 48 sample areas. The few studies for forests based on sentinal-2 data that have been carried out have reported CCC estimation results with an RMSE of 31 μg cm^{-2} for model inversion using an LUT [11] and an RMSE of 34 μg cm^{-2} using random forest [33]. Compared with the sentinal-2 data, which have a resolution of 10 m, MERIS data, which have a resolution of 300 m, contain more mixed pixels, which will increase the difficulty of the inversion. Our results are reasonable and encouraging compared with the accuracy evaluation results reported in the previous studies described above [11], [33]. Our validation results indicate that the proposed algorithms performed well across different sites and forest types although some uncertainty remains. The sources of this uncertainty and ways that this uncertainty might be minimized are discussed in more detail below.

In this article, simulated datasets generated by the PROSPECT-D model and four-scale model were used to estimate the CCC. However, the uncertainty in the simulated data used for modeling can affect the retrieved vegetation parameters [47]. To evaluate the effects of including the CI, in this article, a wider range of canopy structure parameters was used and more than 5 000 000 simulated datasets were produced for each forest type; however, it is still difficult to fully cover all possible real forest types. In the future, more detailed simulation scenarios corresponding to different forest types should be established or more existing data, such as tree heights extracted using Lidar [65], should be used as prior knowledge to improve the ability to describe real forest scenes.

The MERIS product and the CI product also served as an input variable for CCC estimation in this article. In addition to the uncertainty factors of MERIS surface reflectance relating to sensor noise, radiation calibration and atmospheric correction errors [66], the error in the satellite CI product is another important source of uncertainty in our proposed CCC retrieval model. The accuracy of CI products derived from satellite data, especially MODIS BRDF product, is constantly improving [40]–[42]. He *et al.* [40] used the MODIS BRDF parameter that has a 500-m resolution to derive a global CI map and the MODIS-derived

TABLE VI
COMPARISONS BETWEEN THE PERFORMANCES OF CCC RETRIEVAL USING THE RANDOM FOREST MODEL WITH DIFFERENT MERIS TRAINING DATASETS

The predictor variables	Broadleaf forest		Coniferous forest	
	R^2	RMSE (μg cm^{-2})	R^2	RMSE (μg cm^{-2})
Refl	0.70	32.62	0.34	72.28
Refl_MTCI	0.77	56.49	0.60	68.01
CI_Refl	0.55	46.16	0.43	40.34
CI_Refl_MTCI	0.77	27.95	0.61	34.83
$Refl_{all}$	0.69	50.54	0.33	99.25
$Refl_{all}$_MTCI	0.82	52.03	0.48	90.59
CI_$Refl_{all}$	0.57	39.08	0.31	61.40
CI_$Refl_{all}$_MTCI	0.78	30.99	0.60	46.04
$Refl_{cor}$	0.84	33.84	0.39	94.37
$Refl_{cor}$_MTCI	0.70	53.74	0.61	76.43
CI_$Refl_{cor}$	0.64	40.22	0.41	51.33
CI_$Refl_{cor}$_MTCI	0.75	30.19	0.54	38.85

Refl: Reflectance in the red and red-edge bands, including bands 5–10; these had previously been tested as described in Section III; $Refl_{all}$: all 13 available MERIS bands; $Refl_{cor}$: reflectance selected based on the correlation matrix, which includes MERIS band 5, band 6, band 8, band 9, band 10, and band 13; CI: Clumping index; and MTCI: MERIS terrestrial chlorophyll index.

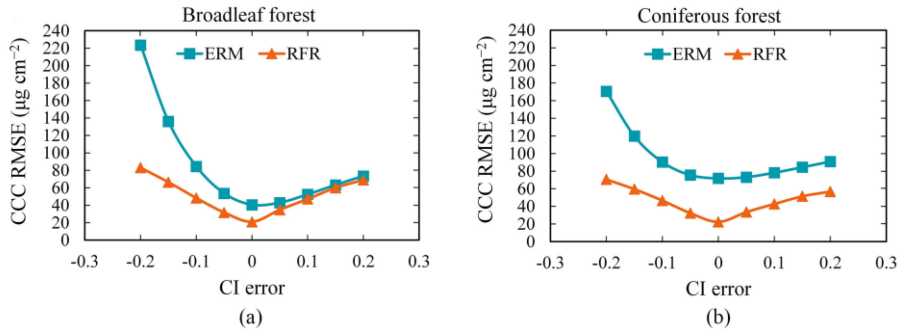


Fig. 9. Relationship between the CI error and the RMSE of the CCC retrieved using the empirical regression and random forest models for (a) BOF and (b) coniferous forest. ERM stands for the empirical regression model based on MTCI and CI, and RFR stands for random forest regression.

CI values were found to be highly correlated ($R^2 = 0.76$) with the field-measured CI data. Jiao *et al.* [41] updated the CI estimation algorithm by reconstructing the hotspot signatures of the MODIS BRDF product; the evaluation results gave an R^2 value of 0.80 and an RMSE of 0.07 for the CI. As satellite CI products improve, more studies are using satellite CI products as prior knowledge in the estimation of vegetation parameters. For example, the MODIS-derived CI data produced by Jiao *et al.* [41] were included to build a robust retrieval model of fractional vegetation cover [44]. This MODIS-derived CI data also have been used to test our approach—an improvement in the results could be seen when this was done. We included the error in the CI in a group of simulated test data and found that this error and the error in the CCC estimates were correlated (see Fig. 9). Taking the current reported error in the CI product (RMSE = 0.07) as an example, the increase in the RMSE for CCC estimates is small (less than 20 μg cm^{-2}), regardless of whether the empirical method or the random forest method is used. In cases where the CI is underestimated, if the absolute error in the CI exceeds 0.1, using the empirical regression method will produce a greater RMSE for the CCC estimates. In general, the likely size of the error in CCC estimates produced by the reported error in the CI product based on MODIS BRDF data is acceptable. In the future, more accurate CI products should help improve the accuracy of CCC estimates retrieved using our proposed approaches.

D. Uncertainties in Validation of the CCC Retrieval Model Based on MERIS Data

The validation of MERIS-based CCC estimates requires field measurements of forest LCC and LAI which are matched with satellite pixels. We used Landsat NDVI data to carry out scaling of the field-measured LAI; however, it is difficult to acquire field measurements of forest LCC and LAI data at the same time and, at present, there are not many simultaneous samples of LCC and LAI available. In order to obtain as many test samples as possible, in this article, for some samples, measured LAI data were used and, for others, MODIS LAI data were used. Compared with the field-measured LAI data, the MODIS LAI product matches the MERIS data better in terms of spatial

scale. In the MODIS LAI retrieval algorithm, the MODIS LAI is designed to give the true LAI and several previous studies have found that the MODIS LAI is closer to the true field-measured values than to the effective LAI [67]–[69]. Although the MODIS LAI C6 product has an RMSE of 0.66 and is a significant improvement on the previous versions [69], the error of the MODIS LAI data increases the uncertainty of the validation of CCC retrieval in this article.

In addition, there was no available higher-resolution red-edge satellite data during the acquisition time of the MERIS data that could be used to scale the *in situ* LCC. In forest areas, the heterogeneity of the LAI is generally higher than that of the LCC [11], [56]. Taking one particular forest area as an example [11], the coefficient of variation for the *in situ* LCC was 8.9%, whereas for the *in situ* LAI it was 20.9%. This means that the heterogeneity of the forest LAI contributes more to the heterogeneity of the CCC in a coarse-resolution pixel than the LCC. In fact, there have been no studies on the scaling of *in situ* LCC data for testing MERIS-based LCC estimates [47]. This lack of scaling of the LCC inevitably produces uncertainties. Naturally, due to the diversity of forest species and the complexity of growth conditions, more ground measurements of forest CCC need to be tested in the future.

V. Conclusion

In this article, the concept of the effective CCC by integrating CCC and CI was proposed. Compared with CCC, CCC_e is closely related to canopy spectral feature (e.g., MTCI), which implies that the CI may be important for remote sensing of forest CCC. We developed a new approach for MERIS-based estimating forest CCC that included the CI using empirical regression and also a random forest approach based on simulations made by the PROSPECT-D and four-scale models. Our findings demonstrate that when the prior knowledge of the CI was included in the CCC estimation using either empirical regression or random forest regression, the accuracy improved markedly compared with the results for models that do not include the CI. First, the results for the empirical regression approach showed that the MTCI-based CCC retrieval model that included CI information performed better than that without the CI and produced a decrease in the RMSE to 36.51 μg cm^{-2} from 63.64 μg cm^{-2} for BOF and to 58.49 μg cm^{-2} from 96.02 μg cm^{-2} for coniferous forest. Next, the validation of the results for the random forest model trained using different predictors suggested that CI_6B_MTCI-based model that used the CI, the red and red-edge bands, and the MTCI performed best, achieving an R^2 value of 0.77, a Bias value of 3.66 μg cm^{-2}, and an RMSE of 27.95 μg cm^{-2} for BOF and an R^2 value of 0.61, a Bias value of -21.66 μg cm^{-2}, and an RMSE of 34.83 μg cm^{-2} for coniferous forest. These results demonstrate the feasibility of making MERIS-based CCC estimates by including the CI, as well as showing that the random forest approach has more potential to predict forest CCC than the empirical regression approach at large scales. In addition, our findings also show that selecting the chlorophyll-related reflectance and utilizing a chlorophyll-related vegetation index like the MTCI is necessary for building a robust random forest model for estimating forest CCC. In future, more diverse validation data and satellite data with red-edge band should be used to further assess the accuracy and robustness of the proposed forest CCC retrieval models. The evaluation of the effectiveness of different satellite-based CI products involved in forest CCC retrieval is also worth exploring further.

Acknowledgment

The authors would like to thank all the authors who published the forest field data cited in this article, and thank Dr. H. Croft from University of Toronto for her kind help during the published field data collection.

References

[1] A. A. Gitelson, Y. Gritz, and M. N. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," *J. Plant Physiol.*, vol. 160, no. 3, pp. 271–282, Mar. 2003.

[2] M. Chen and R. E. Blankenship, "Expanding the solar spectrum used by photosynthesis," *Trends Plant Sci.*, vol. 16, no. 8, pp. 427–431, Aug. 2011.

[3] H. Croft *et al.*, "Leaf chlorophyll content as a proxy for leaf photosynthetic capacity," *Global Chang Biol.,* vol. 23, no. 9, pp. 3513–3524, Sep. 2017.

[4] F. Baret, V. Houles, and M. Guerif, "Quantification of plant stress using remote sensing observations and crop models: The case of nitrogen manage.," *J. Exp. Botany,* vol. 58, no. 4, pp. 869–808, Mar. 2007.

[5] J. G. P. W. Clevers and L. Kooistra, "Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 5, no. 2, pp. 574–583, Apr. 2012.

[6] P. J. Kumar, and R. Gopal, "Laser-induced chlorophyll fluorescence and reflectance spectroscopy of cadmium treated triticum aestivum l. plants," *J. Spectroscopy*, vol. 26, no. 2, pp. 129–139, 2011.

[7] L. A. Brown, B. O. Ogutu, and J. Dash, "Estimating forest leaf area index and canopy chlorophyll content with Sentinel-2: An evaluation of two hybrid retrieval algorithms," *Remote Sens.,* vol. 11, no. 15, pp. 1752–1769, Aug. 2019.

[8] J. Clevers, L. Kooistra, and M. van den Brande, "Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop," *Remote Sens.,* vol. 9, no. 5, pp. 405–420, May. 2017.

[9] J. A. Gamon *et al.*, "A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers," *Proc. Nat. Acad. Sci. USA,* vol. 113, no. 46, pp. 13087–13092, Nov. 2016.

[10] H. Croft, J. M. Chen, N. J. Froelich, R. M. Staebler, "Seasonal controls of canopy chlorophyll content on forest carbon uptake: Implication for GPP model," *J. Geophys. Res. Biogeosci.*, vol. 120, no. 8, pp. 1576–1586, Aug. 2015.

[11] A. M. Ali *et al.*, "Comparing methods for mapping canopy chlorophyll content in a mixed mountain forest using Sentinel-2 data," *Int. J. Appl. Earth Observ. Geoinf.,* vol. 87, 2020, Art. no. 102037.

[12] A. Pinar and P. J. Curran, "Grass chlorophyll and the reflectance red edge," *Int. J. Remote Sens.,* vol. 17, no. 2, pp. 351–357, 1996.

[13] J. G. P. W. Clevers and L. Kooistra, "Using hyperspectral remote sens. data for retrieving canopy chlorophyll and nitrogen content," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 5, no. 2, pp. 574–583, Apr. 2012.

[14] J. Dash and P. J. Curran, "The MERIS terrestrial chlorophyll index," *Int. J. Remote Sens.,* vol. 25, no. 23, pp. 5403–5413, Dec. 2004.

[15] F. Vuolo, J. Dash, P. J. Curran, D. Lajas, and E. Kwiatkowska, "Methodologies and uncertainties in the use of the terr. chlorophyll index for the Sentinel-3 mission," *Remote Sens.,* vol. 4, no. 5, pp. 1112–1133, May. 2012.

[16] P. J. Zarco-Tejada *et al.*, "Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral imagery in the context of forest decline," *Remote Sens. Environ.,* vol. 223, pp. 320–335, Mar. 2019.

[17] D. Haboudane, J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze, "Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture," *Remote Sens. Environ.,* vol. 81, no. 2/3, pp. 416–426, Aug. 2002.

[18] P. J. Zarco-Tejada, J. R. Miller, A. Morales, A. Berjon, and J. Aguera, "Hyperspectral indices and model simuluation for chlorophyll estimation in open-canopy tree crops," *Remote Sens. Environ.,* vol. 90, no. 4, pp. 463–476, Apr. 2004.

[19] Y. Peng, A. Nguy-Robertson, T. Arkebauer, and A. Gitelson, "Assessment of canopy chlorophyll content retrieval in maize and soybean: Implication of hysteresis on the development of generic algorithms," *Remote Sens.,* vol. 9, no. 3, pp. 226–244, Mar. 2017.

[20] J. B. Féret, A. A. Gitelson, S. D. Noble, and S. Jacquemoud, "PROSPECT-D: Towards modeling leaf opticals properties through a complete lifecycle," *Remote Sens. Environ.,* vol. 193, pp. 204–215, May. 2017.

[21] J. M. Chen and S. G. Leblanc, "A four-scale bidirectional reflectance model based on canopy architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1316–1337, Sep. 1997.

[22] S. Jacquemoud, C. Bacour, H. Poilve, and J.-P. Frangi, "Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode," *Remote Sens. Environ.,* vol. 74, no. 3, pp. 471–481, Dec. 2000.

[23] P. J. Zarco-Tejada, J. R. Miller, T. L. Noland, G. H. Mohammed, and P. H. Sampson, "Scaling-up and model inversion methods with narrow-band opt. indices for chlorophyll content estimation in closed forest canopies with hyperspectral data," *IEEE Trans. Geosci. Remote Sens.,* vol. 39, no. 7, pp. 1491–1507, Jul. 2001.

[24] H. Croft, J. M. Chen, and Y. Zhang, "The applicability of empirical vegetation indices for determining leaf chlorophyll content over different leaf and canopy structures," *Ecol. Complex.,* vol. 17, pp. 119–130, Mar. 2014.

[25] G. le Maire, C. François, and E. Dufrêne, "Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements," *Remote Sens. Environ.,* vol. 89, no. 1, pp. 1–28, Jan. 2004.

[26] J. U. H. Eitel, D. S. Long, P. E. Gessler, E. R. Hunt, and D. J. Brown, "Sensitivity of ground-based remote sensing estimates of wheat chlorophyll content to variation in soil reflectance," *Soil Sci. Soc. Amer. J.,* vol. 73, no. 5, pp. 1715–1723, Sep–Oct. 2009.

[27] R. Houborg and M. F. McCabe, "A hybrid training approach for leaf area index estimation via cubist and random forests machine-learning," *ISPRS-J. Photogramm. Remote Sens.,* vol. 135, pp. 173–188, Jan. 2018.

[28] L. Wang, X. Zhou, X. Zhu, Z. Dong, and W. Guo, "Estimation of biomass in wheat using random forest regression algorithm and remote sensing data," *Crop J.,* vol. 4, no. 3, pp. 212–219, Jun. 2016.

[29] J. Wang, Y. Chen, F. Chen, T. Shi, and G. Wu, "Wavelet-based coupling of leaf and canopy reflectance spectra to improve the estimation accuracy of foliar nitrogen concentration," *Agricultural Forest. Meteorol.,* vol. 248, pp. 306–315, Jan. 2018.

[30] L. Liang *et al.*, "Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method," *Int. J. Remote Sens.,* vol. 37, no. 13, pp. 2923–2949, Jun. 2016.

[31] D. Upreti *et al.*, "A comparison of hybrid mach. Learn. algorithms for the retrieval of wheat biophysical variables from Sentinel-2," *Remote Sens.,* vol. 11, no. 5, pp. 481–503, Mar. 2019.

[32] D. P. Cavallo, M. Cefola, B. Pace, A. F. Logrieco, and G. Attolico, "Contactless and non-destructive chlorophyll content prediction by random forest regression: A case study on fresh-cut rocket leaves," *Comput. Electron. Agriculture,* vol. 140, pp. 303–310, Aug. 2017.

[33] A. M. Ali, R. Darvishzadeh, A. Skidmore, T. W. Gara, and M. Heurich, "Machine learning methods' performing in radiative transfer model inversion to retrieve plant traits from Sentinel-2 data of a mixed mountain forest," *Int. J. Digit. Earth,* vol. 14, no. 1, pp. 106–120, Jul. 2020.

[34] W. Dorigo, R. Richter, F. Baret, R. Bamler, and W. Wagner, "Enhanced automated canopy characterization from hyperspectral data by a novel two step radiative transfer model inversion approach," *Remote Sens.,* vol. 1, no. 4, pp. 1139–1170, Dec. 2009.

[35] A. Simic Milas, M. Romanko, P. Reil, T. Abeysinghe, and A. Marambe, "The importance of leaf area index in mapping chlorophyll content of corn under different agricultural treatments using UAV images," *Int. J. Remote Sens.,* vol. 39, no. 15–16, pp. 5415–5431, Mar. 2018.

[36] J. M. Chen and T. A. Black, "Foliage area and architecture of plant canopies from sunfleck size distributions," *Agricultural Forest Meteorol.,* vol. 60, no. 3–4, pp. 249–266, Aug. 1992.

[37] V. Thomas, T. Noland, P. Treitz, and J. H. McCaughey, "Leaf area and clumping indices for a boreal mixed-wood forest: Lidar, hyperspectral, and landsat models," *Int. J. Remote Sens.,* vol. 32, no. 23, pp. 8271–8297, Oct. 2011.

[38] H. Fang, W. Liu, W. Li, and S. Wei, "Estimation of the directional and whole apparent clumping index (ACI) from indirect optical measuremnets," *ISPRS-J. Photogramm. Remote Sens.,* vol. 144, pp. 1–13, Oct. 2018.

[39] J. M. Chen, C. H. Menges, and S. G. Leblanc, "Global mapping of foliage clumping index using multi-angular satellite data," *Remote Sens. Environ.,* vol. 97, no. 4, pp. 447–457, Sep. 2005.

[40] L. He, J. M. Chen, J. Pisek, C. B. Schaaf, and A. H. Strahler, "Global clumping index map derived from the MODIS BRDF product," *Remote Sens. Environ.,* vol. 119, pp. 118–130, Apr. 2012.

[41] Z. Jiao *et al.*, "An algorithm for the retrieval of the clumping index (CI) from the MODIS BRDF product using an adjusted version of the kernel-driven BRDF model," *Remote Sens. Environ.,* vol. 209, pp. 594–611, May. 2018.

[42] S. Wei, H. Fang, C. B. Schaaf, L. He, and J. M. Chen, "Glob. 500 m clumping index product derived from MODIS BRDF data (2001–2017)," *Remote Sens. Environ.,* vol. 232, 2019, Art. no. 111296.

[43] J. Pisek *et al.*, "Intercomparison of clumping index estimates from POLDER, MODIS, and MISR satellite data over reference sites," *ISPRS-J. Photogramm. Remote Sens.,* vol. 101, pp. 47–56, Mar. 2015.

[44] Z. Xiao *et al.*, "Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance," *IEEE Trans. Geosci. Remote Sens.* vol. 52, no. 1, pp. 209–223, Jan. 2015.

[45] J. Zhao *et al.*, "Estimating fractional vegetation cover from leaf area index and clumping index based on the gap probability theory," *Int. J. Appl. Earth Observ. Geoinf.,* vol. 90, 2020, Art. no. 102112.

[46] J. B. Féret *et al.*, "PROSPECT-4 and 5: Advances in the leaf opt. properties model separating photosynthetic pigments," *Remote Sens. Environ.,* vol. 112, no. 6, pp. 3030–3043, Jun. 2008.

[47] H. Croft *et al.*, "The global distribution of leaf chlorophyll content," *Remote Sens. Environ.,* vol. 236, 2020, Art. no. 111479.

[48] B. Kötz *et al.*, "Radiative transfer modeling within a heterogeneous canopy for estimation of forest fire fuel properties," *Remote Sens. Environ.,* vol. 92, no. 3, pp. 332–344, Aug. 2004.

[49] A. De Santis, E. Chuvieco, and P. J. Vaughan, "Short-term assessment of burn severity using the inversion of PROSPECT and geosail models," *Remote Sens. Environ.,* vol. 113, no. 1, pp. 126–136, Jan. 2009.

[50] J. M. Chen and S. G. Leblanc. "Multiple-scattering scheme useful for geometric optical modeling," *IEEE Trans. Geosci. Remote Sens.,* vol. 39, no. 5, pp. 1061–1071, May. 2001.

[51] T. Nilson, "A theoretical analysis of the frequency of gaps in plant stands," *Agricultural. Meteorol.,* vol. 7, pp. 25–38, 1970.

[52] J. M. Chen, "Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands," *Agricultural Forest Meteorol.*, vol. 80, no. 2/4, pp. 135–163, Jul. 1996.

[53] R. Myneni, Y. Knyazikhin, T. Park, "MCD15A3H MODIS/terra+aqua leaf area index/FPAR 4-day L4 global 500m SIN grid V006 [Data set]," NASA EOSDIS Land Processes DAAC, 2015. Accessed: Oct. 13, 2020. [Online]. Available: https://doi.org/10.5067/MODIS/MCD15A3H.006

[54] A. Simic, J. M. Chen, and T. L. Noland, "Retrieval of forest chlorophyll content using canopy structure parameters derived from multi-angle data: The measurement concept of combining nadir hyperspectral and off-nadir multispectral data," *Int. J. Remote Sens.,* vol. 32, no. 20, pp. 5621–5644, Oct. 2011.

[55] Y. Zhang, J. M. Chen, J. R. Miller, and T. L. Noland, "Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery," *Remote Sens. Environ.,* vol. 112, no. 7, pp. 3234–3247, Jul. 2008.

[56] Y. Zhang, J. M. Chen, S. C. Thomas, "Retrieving seasonal variation in chlorophyll content of overstory and understory sugar maple leaves from leaf-level hyperspectral data," *Can. J. Remote Sens.,* vol. 33, no. 5, pp. 406–415, Oct. 2007.

[57] J. M. Chen and J. Cihlar, "Quantifying the effect of canopy architecture on optical measurement of leaf area index using two gap size analysis methods," *IEEE Trans. Geosci. Remote Sens.,* vol. 33, no. 3, pp. 777–787, May. 1995.

[58] Y. Li *et al.*, "Retrieving the gap fraction, element clumping index, and leaf area index of individual trees using single-scan data from a terrestrial laser scanner," *ISPRS-J. Photogramm. Remote Sens.,* vol. 130, pp. 308–316, Aug. 2017.

[59] R. Darvishzadeh, A. Skidmore, M. Schlerf, and C. Atzberger, "Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland," *Remote Sens. Environ.,* vol. 112, no. 5, pp. 2592–2604, May. 2008.

[60] A. Banskota *et al.*, "An LUT-based inversion of DART model to estimate forest LAI from hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 8, no. 6, pp. 3147–3160, Jun. 2015.

[61] X. Qian and L. Liu, "Retrieving crop leaf chlorophyll content using an improved look-up-table approach by combining multiple canopy structures and soil backgrounds," *Remote Sens.,* vol. 12, no. 13, pp. 2139–2156, Jul. 2020.

[62] J. G. P. W. Clevers and A. A. Gitelson, "Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3," *Int. J. Appl. Earth Observ. Geoinf.,* vol. 23, pp. 344–351, Aug. 2013.
[63] D. P. Cavallo, M. Cefola, B. Pace, A. F. Logrieco, and G. Attolico, "Contactless and non-destructive chlorophyll content prediction by random forest regression: A case study on fresh-cut rocket leaves," *Comput. Electron. Agricultural,* vol. 140, pp. 303–310, Aug. 2017.
[64] Y. Inoue, J. Penuelas, A. MiyataI, and M. Mano, "Normalized difference spectral indices for estimating photosynthetic efficiency and capacity at a canopy scale derived from hyperspectral and CO_2 flux measurements in rice," *Remote Sens. Environ.,* vol. 112, no. 1, pp. 156–172, Jan. 2008.
[65] S. Ganz, Y. Käber, and P. Adler, "Measuring tree height with remote sensing–A comparison of photogrammetric and LiDAR data with different field meas.," *Forests,* vol. 10, no. 8, pp. 694–708, Aug. 2019.
[66] S. Garrigues *et al.*, "Validation and intercomparison of global leaf area index products derived from remote sensing data," *J. Geophys. Res.-Biogeosci.,* vol. 113, no. G2, 2008, Art. no. G02028.
[67] H. Fang, S. Wei, C. Jiang, and K. Scipal, "Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method," *Remote Sens. Environ.,* vol. 124, pp. 610–621, Sep. 2012.
[68] A. Verger, F. Baret, and M. Weiss, "Performances of neural network for deriving LAI estimates from existing CYCLOPES and MODIS products," *Remote Sens. Environ.,* vol. 112, no. 6, pp. 2789–2803, Jun. 2008.
[69] K. Yan *et al.*, "Evaluation of MODIS LAI/FPAR product collection 6. Part 2: Validation and intercomparison," *Remote Sens.,* vol. 8, no. 6, pp. 460–486, Jun. 2016.

Qi Sun received the B.S. degree in surveying and mapping engineering from Henan Polytechnic University, Henan, China, in 2015. He is currently working toward the Ph.D. degree at the College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing, China.

His research interest is vegetation quantitative remote sensing.

Quanjun Jiao received the B.S. degree in cartography and geographic information system from Peking University, Beijing, China, in 2003, and the Ph.D. degree in cartography and geographic information system with the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS) , Beijing, China, in 2008.

He is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research is mainly in the field of optical remote sensing techniques in vegetation parameters retrieval and its applications.

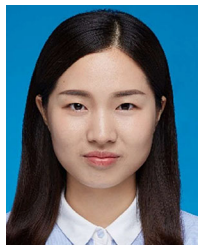
Liangyun Liu received the B.S. degree in physics from Huaibei Normal University, Huaibei, China, in 1996, and the Ph.D. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 2000.

He is currently a Research Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. He has authored or coauthored more than 200 journal papers in vegetation quantitative remote sensing. He is currently a Research Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. He has authored or coauthored more than 200 journal papers in vegetation quantitative remote sensing.

Dr. Liu was the recipient of five government's prizes, including two Second-Class National Scientific and Technological Progress Award, and also honored by National Science Fund for Distinguished Young Scholars in 2018.

Xinjie Liu received the B.S. degree from the Nanjing University of Information Science and Technology, Nanjing, China, in 2011 and the Ph.D. degree from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2016.

He is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include quantitative remote sensing of vegetation and solar-induced chlorophyll fluorescence.



Xiaojin Qian received the B.S. degree in remote sensing science and technology from Jiangsu Normal University, Xuzhou, China, in 2016. She is currently working toward the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

Her research interest is vegetation quantitative remote sensing.

Xiao Zhang received the B.S. degree in remote sensing science and technology from China University of Geosciences, Wuhan, China, in 2015 and the Ph.D. degree from the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2020.

He is currently an Assistant Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include image processing and time series land cover mapping.

Bing Zhang (Fellow, IEEE) received the B.S. degree in geography from Peking University, Beijing, China, in 1991, and the M.S. and Ph.D. degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 1994 and 2003, respectively.

He is currently a Full Professor and the Deputy Director of the Aerospace Information Research Institute, CAS, where he has been leading lots of key scientific projects in the area of hyperspectral remote sensing for more than 25 years.. He has developed five software systems in the image processing and applications. He has authored more than 300 publications, including more than 170 journal articles. He has edited six books/contributed book chapters on hyperspectral image processing and subsequent applications. His research interests include the development of mathematical and physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas.

Dr. Zhang's creative achievements were rewarded ten important prizes from Chinese government, and special government allowances of the Chinese State Council. He was awarded the National Science Foundation for Distinguished Young Scholars of China in 2013, and was the recipient of the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, the highest level of Awards for the CAS scholars. He is currently an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He has been serving as a Technical Committee Member of IEEE Workshop on Hyperspectral Image and Signal Processing since 2011, and as the president of hyperspectral remote sensing committee of China National Committee of International Society for Digital Earth since 2012, and as the Standing Director of Chinese Society of Space Research (CSSR) since 2016. He is the Student Paper Competition Committee member in IGARSS, from 2015 to 2019.