A study on the effects of tree spatial distribution on the accuracy of canopy clumping index retrieved by remote sensing

Teng Zhang, Jun Geng*, Jingsong Xu, Yuping Wang, Hong Hong

College of Civil Engineering, Hefei University of Technology, Hefei, China, 230009

*Corresponding author: gengj@hfut.edu.cn

Abstract. Clumping Index (CI) is a key parameter for estimating the leaf area index (LAI) in forest canopy and distinguishing between "sunlit foliage" and "shaded foliage" in forest canopy. It is of great significance for studying total primary productivity and evapotranspiration of forests. The method of canopy CI retrieved by multi-angle satellite data has been widely used. However, it is unclear whether the method of using the normalized difference between hotspot and darkspot (NDHD) to retrieve CI by remote sensing is suitable for forests with different tree spatial distributions (TSDs). In this study, the accuracy of canopy CI retrieved by remote sensing under different TSDs was investigated by using 1 km × 1 km within the Saihanba Forest Farm as the study area, and using the DART model to calculate the multi-angle gap fractions and reflectance under different TSDs. The results showed that: (1) The CI results of retrieval of the plots with clustered TSDs were generally larger than the CI results of validation, while the CI results of retrieval of the plots with regular TSDs were generally smaller than the CI results of validation. (2) The accuracies of the CI results of retrieval were small for plots with strong clustered TSDs or strong regular TSDs. The results of this study play an important role in understanding the influence of TSDs in canopy CI retrieved by remote sensing, as well as in assessing forest productivity and studying carbon cycling in terrestrial ecosystems.

Keywords: Clumping Index, Tree Spatial Distribution, Retrieved by Remote Sensing, Forest, Saihanba Forest Farm.

1. Introduction

Clumping Index (CI) is defined as the ratio of the effective leaf area index (LAI_e) to the leaf area index (LAI), which is a key parameter for estimating LAI and distinguishing between "sunlit foliage" and "shaded foliage" in the forest canopy. Failure to consider CI will result in LAI errors of up to 70% [1]. Therefore, CI is important for the study of total primary productivity and evapotranspiration of forests [2]. CI is an important parameter of canopy structure, which is used to describe the degree of nonrandom spatial distribution of foliage elements in a canopy. When the foliage in a forest canopy show a clustered distribution, the CI is smaller than 1; when the foliage shows a random distribution, the CI is equal to 1; and when the foliage shows a regular distribution, the CI is larger than 1. The smaller the CI value, the higher the degree of clustering of foliage in the canopy.

Canopy CI retrieved by remote sensing mainly uses satellite sensors to retrieve canopy structural parameters by observing the structural information on the vegetation surface through the bidirectional reflectance distribution function (BRDF). Chen et al. [3] proposed a normalized difference between hotspot and darkspot (NDHD) and produced a global CI map using POLDER. With the release of higher-resolution satellite data, the method of retrieving CI by NDHD are widely used. For example, He et al. [4] generated the first 500 m resolution global CI map using MODIS. Yin et al. [5] retrieved the global CI based on the updated MODIS BRDF, and other scholars derived 8-day and monthly global time series CI maps since 2000 based on the MODIS BRDF [6-7]. In verifying the accuracy of CI retrieved by remote sensing, digital hemispherical photography (DHP), LAI-2000, LAI-2200, and tracing radiation and architecture of canopies (TRAC) are often used to validate.

The above scholars have made important contributions to the study of canopy CI retrieved by remote sensing. However, canopy CI is closely related to the tree spatial distributions (TSDs).

Different TSDs profoundly affects the degree of foliage aggregation in the canopy, which in turn affects the accuracy of canopy CI retrieved by remote sensing. In addition, the TSDs are the main factor contributing to the spatial heterogeneity of the canopy, which affects the foliage spatial distribution in the canopy, and thus contributes to the scale effects of canopy CI. Neglecting the influence of CI scale effects would lead to large uncertainties in water and carbon cycle simulations, as well as erroneous estimates of total vegetation primary productivity and evapotranspiration. However, it is unclear whether the method of CI retrieved by NDHD is suitable for forests with different TSDs. In this study, the Saihanba Forest Farm is taken as the research object, and utilize the DART model to calculate the multi-angle gap fractions (GFs) and reflectance under different TSDs and then validate and retrieve the canopy CI. The correlation analysis was conducted with the CI of validation and CI of retrieval, in order to explore the accuracy of the canopy CI retrieved by remote sensing under different TSDs. The results of the study play an important role in understanding the influence of TSDs in the CI retrieved by remote sensing, as well as assessing the productivity of forests and studying the carbon cycle of terrestrial ecosystems.

2. Materials

2.1. Study area

The study area is located in a 1 km × 1 km area within the Saihanba Forest Farm (SFF) in Chengde City, Hebei Province, China (Figure 1). It is located in the northern part of Hebei Province, which is the intersection of Yanshan Mountains, the residual of the Great Xing'an Mountains, and Hunshandake Sandy Land. The altitude of the SFF is 1010 m~1939.9 m, the average annual temperature is about -1.3℃, the average annual precipitation is 479 mm, and the average annual snow accumulation is 7 months, which belongs to the cold-temperate continental climate of the monsoon zone. The soil types of SFF are dominated by mountain brown loam, gray forest soils, and windy sandy soils. The area of SFF has reached 76 730 ha, with the forest coverage rate exceeding 80% and the forest stock reaching 10.368 million m^3 (For more information, please refer to the website: http://lycy.hebei.gov.cn/shb/show_article.php?id=5876).

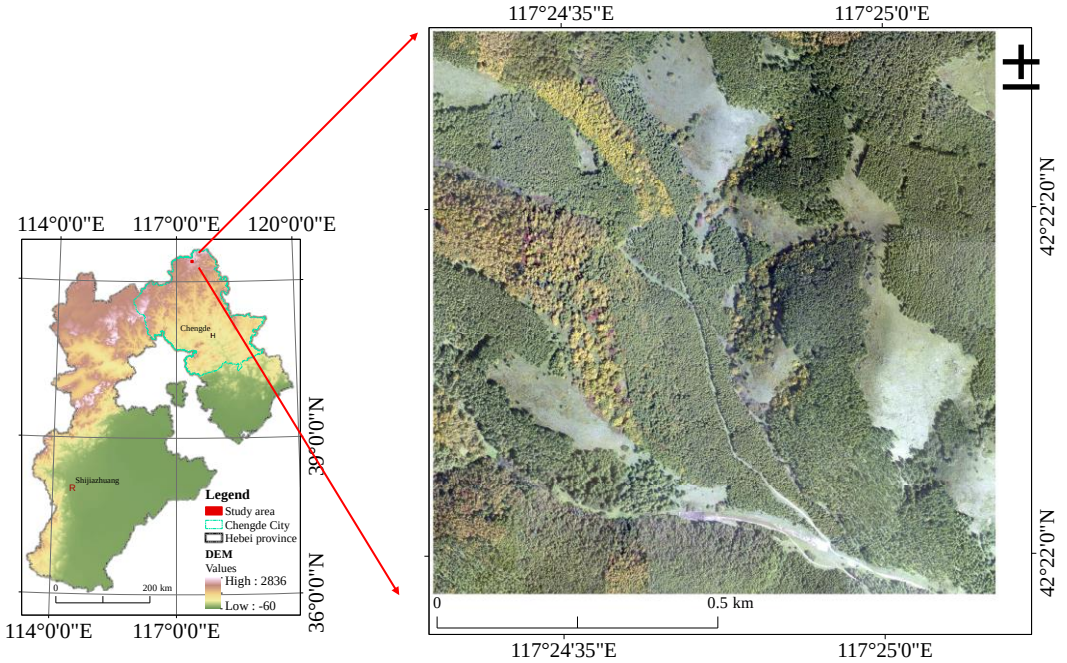


Figure 1. Diagram of study area location

2.2. Data source

The data used in this study was obtained by LiCHy (LiDAR, CCD, and Hyperspectral) Airborne Observation system of Chinese Academy of Forestry Sciences from September 5th to 17th, 2018, and the projection coordinate system is WGS84-UTM50. The single-tree segmentation algorithm is based on the Nyström spectral clustering with the K-nearest neighbour sampling method (https://github.com/limingado/NSC/tree/v1.0.0). For specific information, please refer to reference [8].

2.3. Theories and methods

2.3.1 Indexes

(1) Aggregation index (AI)

AI, which is based on the combination of expectation and variance and does not take into account the radius and area of the crown, is simple way to analyze the TSD [9]. AI is calculated as follows:

$$AI = s^2 / \bar{x} - 1 \tag{1}$$

where s^2 is the variance and $\bar{x}$ is the expectation. When the points in the space are randomly distributed, then $AI = 0$; if the TSD is the cluster, it follows $AI > 0$; finally, if the TSD is regular, one gets $AI < 0$. In this study, the plots were divided by the optimal quadrat size method. The side length of the optimal quadrat is determined based on the number of trees contained in the plot and the area of the plot [10]:

$$Q = \sqrt{2A/n} \tag{2}$$

where Q is the side length of the optimal quadrat; A is the area of plot; n is the number of trees in the plot.

(2) Coefficient of variation (CV)

Duyckaerts et al. [11] proposed a CV based on the Voronoi principle. The CV can be specifically expressed as:

$$CV = \sqrt{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \Big/ N} \Big/ \bar{x} \times 100\% \tag{3}$$

where N is the number of polygons; x_i is the area of the i-th Voronoi polygon; $\bar{x}$ is the average area of all polygons. If the forest is constituted of the regular TSD pattern, then $CV < 33\%$; $33\% \le CV \le 64\%$ means the random whereas $CV > 64\%$ means the cluster.

2.3.2 Calculation of multi-angle canopy GFs and reflectance based on DART model

Computer simulations are an effective alternative to field measurements, providing a highly controlled environment for measuring canopy GFs and reflectance, and avoiding possible confounding factors (e.g., undergrowth, light conditions, and complex topography) that can affect the measurement results. In recent years, three-dimensional radiative transfer models have been used to generate a variety of remote sensing images and applied in the fields of remote sensing model validation, analysis of radiative transfer processes, and retrieval of biochemical parameters [12]. The DART model is one of the most comprehensive three-dimensional radiative transfer models developed by the French laboratory CESBIO. In recent years, due to its strong 3D scene reconstruction and high-precision inversion ability, the DART model has been widely applied in research on reflectance and LAI [13-14]. Therefore, in this study, the DART model was used to calculate the canopy GFs and reflectance of the different TSDs at the 100 m scale in a 1 km × 1 km study area without taking into account various complex factors such as topography, tree branching, and differences in crown types, so as to provide a data basis for the validation and retrieval of canopy CI.

(1) Calculation of multi-angle canopy GFs

Multi-angle canopy GFs can reveal vertical and horizontal heterogeneity of forest. In this study, the LAI_e of the canopy is calculated according to the hemispherical spatial GFs of the canopy, and then validate the canopy CI. Based on the average values of tree height and crown width obtained from the point cloud single tree segmentation results, the single tree obj height was set to 9 m, the crown diameter was set to 4 m, and the leaves in the crown was randomly distributed. The DART model was used to calculate 0.5 m high-resolution four components data at multiple view zenith angles and azimuth angles in each plot. Multi-angle canopy GFs for the plots were calculated based on the sunlit ground and shaded ground in the four components at different view angles. In order to calculate the hemispherical spatial GFs of the plot, this study sets the view azimuth range to 45° to 180° and the step size to 45° during the calculation of the four components. The view zenith angle is from -70° to 70°, with a step size of 10°. The obj parameters of a single tree crown are shown in Table.1.

Table.1. Single tree crown obj parameter

Scene parameters	Parameter settings
Crown type	ellipsoid
Leaf angle distribution	spherical
Crown height (m)	9
Crown diameter (m)	4
Single tree LAI (m^2/ m^2)	3

(2) Calculation of multi-angle canopy reflectance

The multi-angle canopy reflectance reflects the complexity of the canopy structure and changes in physiological characteristics. When calculating canopy reflectance using the DART model, in order to ensure that the view azimuth and solar azimuth are in the same principal plane, the view azimuth and solar azimuth are both set at 180°. The view zenith angle set is from -70° to 70°, with a step size of 10°. The obj parameter information of the tree crown used is shown in Table.1; The reflectance information parameters of the scene, foliage, and ground are shown in Table.2.

Table.2. Setting of reflectance information parameters of scene, foliage, and ground

Structure parameters	Parameter settings (Red, NIR)
Band (nm)	670, 865
Foliage reflectance	0.04, 0.47
Foliage transmittance	0.01, 0.48
Ground reflectance	0.25, 0.34

2.3.3 Estimation methods of canopy CI

(1) Validation method of canopy CI

In this study, the canopy GFs of the plots at different azimuth and zenith angles were calculated by the DART model. The discrete GFs were fitted to the continuous GFs in hemispherical space using curve fitting, and then the LAI_e of the plots was calculated by Miller formula (equation 4). The validation of the canopy CI was tested by the ratio of LAI_e to the LAI of the plots:

$$LAI_e = 2\int_0^{\pi/2} -\ln P(\theta)\cos\theta\sin\theta d\theta \tag{4}$$

$$CI = LAI_e / LAI \tag{5}$$

where θ is the view zenith angle; $P(\theta)$ is the canopy GF when the view zenith angle is θ; LAI_e is the effective leaf area index and LAI is the leaf area index.

(2) Canopy CI retrieved method by remote sensing

In this study, the linear relationship between NDHD and CI is used to retrieve canopy CI:

$$NDHD = (\rho_h - \rho_d)/(\rho_h + \rho_d) \tag{6}$$

$$CI = a(\theta) \times NDHD + b(\theta) \tag{7}$$

where ρ_h and ρ_d are the reflectance values at the hotspot and darkspot in the principal plane, respectively; θ is the sun zenith angle; a and b are coefficients that vary with θ. In this study, according to the parameter settings of the DART model and reference [3], a is taken to be -1.2 and b is taken to be 1.28 when retrieving CI through the red band, and a is taken to be -1.34 and b is taken to be 1.20 when retrieving CI through the NIR band.

3. Results

3.1. Results of the judgment of the TSDs

Figure 2 shows the results of analyzing the TSDs in 1 km study area using AI and CV at the 100 m scale. As can be seen from Figure 2, the plots judged to be clustered distribution by AI and CV at the 100 m scale have similar spatial locations and are mainly located in plots with strong spatial heterogeneity, i.e., mostly located in plots at the edges of the forest stand. For plots with weak spatial heterogeneity, i.e., inside the forest stand, the AI was mostly judged as regular distribution, while the CV was mostly judged as random distribution. It should be noted that the judgment condition for the random distribution of CV is interval, while AI is a fixed value (see Section 2.3.1). The interval value condition is easier to achieve compared to fixed values.



(a) TSD results of AI (b) TSD results of CV

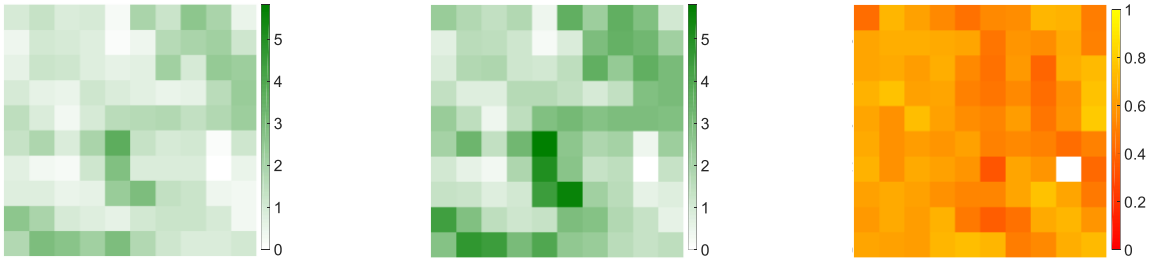
Figure 2. TSDs at 100 m scales in 1 km × 1 km study area

3.2. Estimation of canopy CI

3.2.1 Validation of canopy CI

The canopy GF for the corresponding angle is calculated by four components at multiple angles. Based on the multi-angle GFs results, LAI_e in hemispheric space is calculated with the help of equation (4), and then the validation of canopy CI is calculated by equation (5). Figure 3 is the results of LAI_e, LAI, and validation of canopy CI at 100 m scale.

As can be seen from Figure 2 and Figure 3, the spatial location characteristics of LAI_e and LAI as well as the canopy CI results of validation in the plots were closely related to the distribution characteristics of the TSDs in the plots at the 100 m scale. The TSDs in the areas with smaller CI values often showed clustered distributions. As can be seen from Figure 2 (a) and Figure 3 (c), the crowns in the plots with smaller CI values (i.e., higher degree of foliage aggregation) of the validation at the 100 m scale were mostly clustered distributions. The high spatial heterogeneity of these plots resulted in the phenomenon of aggregation between leaves in the canopy, leading to a larger difference between LAI_e and LAI. For example, the minimum value of CI is 0.3317, corresponding to LAI_e of 0.8945 and LAI of 2.6969. For the plots with regularly distributed crowns, the degree of aggregation of foliage in the canopy is weaker, resulting in small differences between LAI_e and LAI, and large CI results of validation. For example, the maximum value of the CI for the validation was 0.7843, corresponding LAI_e was 2.2579 and LAI was 2.8788.



(a) LAI_e (b) LAI (c) canopy CI results of validation

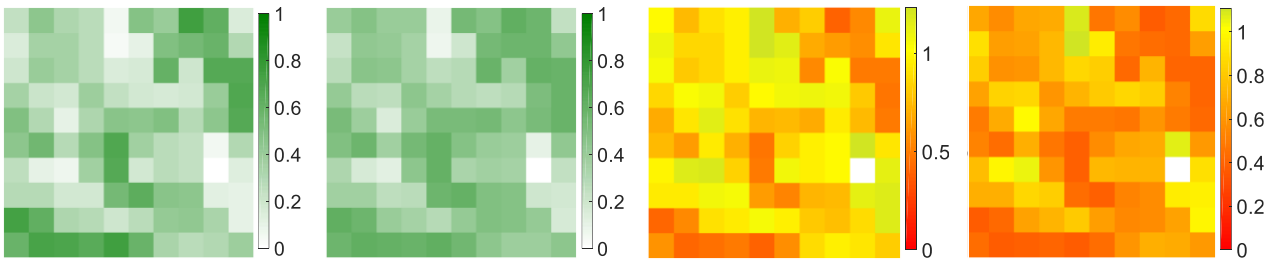
Figure 3. LAI_e, LAI, and validation of canopy CI at 100 m scale

3.2.2 Canopy CI of retrieval

The NDHDs for red band and NIR band were calculated from the multi-angle reflectance in the red band and NIR band in the principal plane. Based on the NDHD results, the CI of the different plots at 100 m scales was calculated by equation (7). Since it is not clear how to represent the CI for plots that do not contain trees, the null value is used in this study. Figure 4 shows the spatial distribution of NDHD and CI based on NDHD retrieval at 100 m scale for red band and NIR band.

From Figure 2 and Figure 4, it can be seen that NDHD values is related to the TSDs. For plots with strong clustering TSDs, the corresponding NDHD values are generally small. The NDHD values for plots with random or regular TSDs are generally higher. Mainly due to the low canopy coverage and strong spatial heterogeneity in plots with strong clustering, the corresponding differences in reflectance between "hotspot" and "darkspot" are small, resulting in a smaller calculated NDHD values. The NDHD values of the red band and the NDHD values of the NIR band have spatial positional similarity. The regions with larger NDHD values in the red band are also larger in the NIR band. The maximum NDHD in the red band is 0.7423, and the maximum NDHD in the NIR band is 0.6202.

As shown in equation (7), the CI of retrieval is negatively and linearly correlated with NDHD, i.e., plots with smaller CI values (high aggregation) corresponded to large NDHD values. From Figure 2 and Figure 4 (c, d), the CI results of retrieval based on the red band and NIR band, the CI results of retrieval for some plots with high aggregation degree are larger than 1. The reason may be due to the low canopy coverage caused by the high degree of canopy aggregation in these plots, resulting in inaccurate CI results based on NDHD retrieval. The CI values based on red band retrieval are generally larger than that in the NIR band. The number of plots with CI larger than 1 in the red band retrieval at the 100m scale is more than that in the NIR band (Figure 4 (c, d)).



(a) NDHD of the red band (b) NDHD of the NIR band (c) CI retrieval based on NDHD of the red band (d) CI retrieval based on NDHD of the NIR band

Figure 4. Retrieval of CI based on NDHD in the red and NIR bands at 100 m scale

3.3. Correlation analysis of TSDs with CI of validation and CI of retrieval

In this section, scatter diagrams of the results of the TSDs judged by AI and CV, with the CI results of validation and CI results of retrieval were plotted respectively. Linear regression models were used

for linear fitting to explore the linear relationship between the TSDs and the CI results of validation and CI results of retrieval. Figure 5 (1) shows the scatter diagram of the CI results of validation with the results of the TSDs. Figure 5 (2, 3) shows the scatter diagrams of the CI results of retrieval by red band and NIR band with the TSDs, respectively.

According to the linear fitting results of the AI and CV in Figure 5 (1) and Section 2.3.1, it can be seen that the CI results of validation are generally small for the clustered TSD plots. For the regular TSD plots, the CI results of validation are generally higher. From the R^2 of the scatter diagram in Figure 5 (1), it can be seen that there are some differences in the linear relationship between the TSDs judged by different indexes and the CI results of validation. The linear relationship between the TSDs of CV and the CI results of validation were stronger than AI. The R^2 of the TSDs judged by CV with the CI results of validation was 0.6377 (Figure 5 (b1)), and the R^2 of the AI was only 0.1992 (Figure 5 (a1)). The reason may be that the AI only consider the number of trees when analyzing the TSDs, while ignoring the impact of crown overlap on canopy CI. As can be seen from Figure 5 (1), it is more appropriate to judge the TSDs by CV rather than AI, and then analyze the effects of TSDs on canopy CI.

According to the linear fitting results of the AI and CV in Figure 5 (2, 3) and Section 2.3.1, it can be seen that for clustered TSD plots, the CI results of retrieval are large, while for regular TSD plots, the CI results of retrieval results are generally small. The linear relationships between the CI results of retrieval and the TSDs of AI and CV are weak, and R^2 are small. However, combined with Figure 5 (b1), it can be seen that there are certain differences in the linear relationship between CV and the CI results of validation, as well as the linear relationship between CV and the CI results of retrieval. For the same plot, especially for plots with significant clustering or regular of TSD, the results of the CI of validation and the CI of retrieval may be different, indicating that the accuracy of the CI results of retrieval are small.

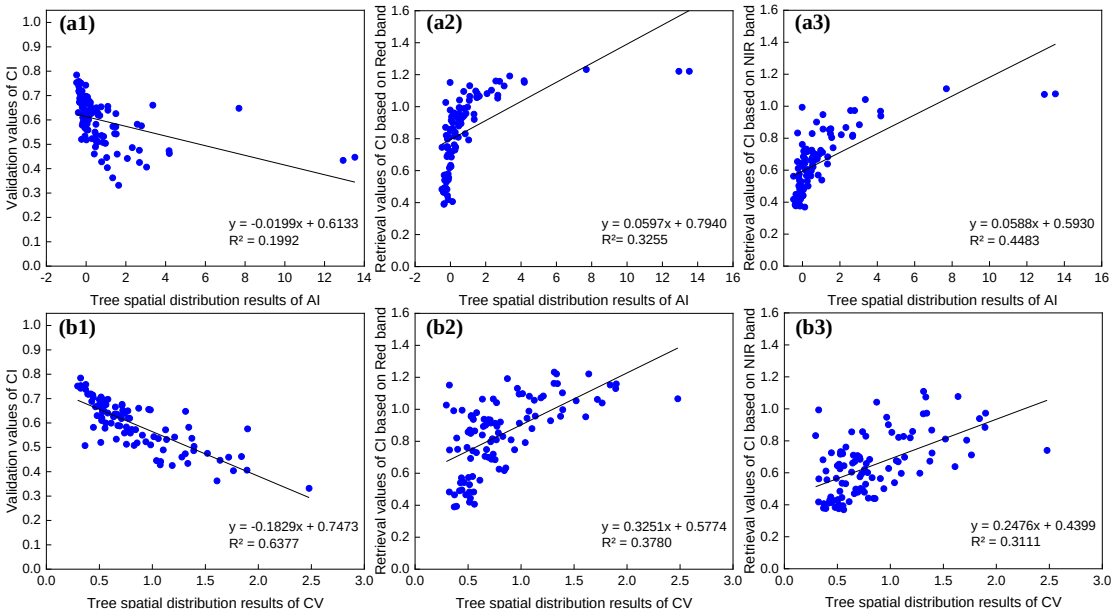


Figure 5. Scatter diagrams of TSDs with the CI results of validation and the CI results of retrieval

4. Conclusions

The tree spatial distribution (TSD) affects the absorption of light in the canopy, which in turn affects photosynthesis in the forest canopy, and is one of the sources of canopy spatial heterogeneity. A full understanding of the TSDs in the forest is of great significance to the study of canopy CI retrieved by remote sensing. The DART model was used in this study to validate and retrieve canopy CI, and explored the accuracies of canopy CI results of retrieval at different TSDs. The main findings of this study are as follows: the CI results of validation were generally small but the CI results of

retrieval were large for the clustered TSDs, and the CI results of validation were generally large but the CI results of retrieval were small for the regular TSDs. For plots with strong clustering or regular TSDs, the accuracies of the CI results of retrieval were small.

In this study, only the effect of TSDs on the canopy CI results of retrieval was considered, and it was found that the accuracies of the CI results of retrieval were poor in the plots where the TSDs were strongly clustered or regular. The accuracies of canopy CI retrieved by remote sensing are affected by various factors such as topography and canopy species. The subsequent studies will incorporate more influencing factors and improve the existing retrieval algorithm to provide a basis for more accurate retrieval of canopy CI.

References

[1] Fang H. Canopy clumping index (CI): A review of methods, characteristics, and applications[J]. Agricultural and Forest Meteorology, 2021,303:108374.

[2] Chen J M, Liu J X, Luo X Z. Improving the Penman-Monteith evapotranspiration model based on the coupling principle of carbon and water fluxes[J]. Transactions of Atmospheric Sciences, 2020,43(1):59-75.

[3] Chen J M, Menges C H, Leblanc S G. Global mapping of foliage clumping index using multi-angular satellite data[J]. Remote Sensing of Environment, 2005,97(4):447-457.

[4] He L, Chen J M, Pisek J, et al. Global clumping index map derived from the MODIS BRDF product[J]. Remote Sensing of Environment, 2012,119:118-130.

[5] Yin S, Jiao Z, Dong Y, et al. Evaluation of the Consistency of the Vegetation Clumping Index Retrieved from Updated MODIS BRDF Data[J]. Remote Sensing, 2022,14(16):3997.

[6] Jiao Z, Dong Y, Schaaf C B, 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[J]. Remote Sensing of Environment, 2018,209:594-611.

[7] Wei S, Fang H, Schaaf C B, et al. Global 500 m clumping index product derived from MODIS BRDF data (2001–2017) [J]. Remote Sensing of Environment, 2019,232:111296.

[8] Pang Y, Wang W, Du L, et al. Nyström-based spectral clustering using airborne LiDAR point cloud data for individual tree segmentation[J]. International journal of digital earth, 2021,14(10):1452-1476.

[9] de Freitas Alves G, de Santana D G. Why do traditional dispersion indices used for analysis of spatial distribution of plants tend to become obsolete? [J]. Population Ecology, 2021:1-13.

[10] Wong D W, Wang F. Spatial Analysis Methods[M]//Comprehensive Geographic Information Systems. 2018:125-147.

[11] Duyckaerts C, Godefroy G. Voronoi tessellation to study the numerical density and the spatial distribution of neurones[J]. Journal of Chemical Neuroanatomy, 2000,20(1):83-92.

[12] Cao B, Roujean J, Gastellu-Etchegorry J, et al. A general framework of kernel-driven modeling in the thermal infrared domain[J]. Remote Sensing of Environment, 2021,252:112157.

[13] Ling X L, Fu C B, Guo W D, et al. Assimilation of Remotely Sensed LAI Into CLM4CN Using DART[J]. Journal of Advances in Modeling Earth Systems, 2019,11(8):2768-2786.

[14] Cao B, Gastellu-Etchegorry J, Yin T, et al. Optimizing the Protocol of Near-Surface Remote Sensing Experiments Over Heterogeneous Canopy Using DART Simulated Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023,61:1-16.