Retrieval of 30-m-Resolution Leaf Area Index From China HJ-1 CCD Data and MODIS Products Through a Dynamic Bayesian Network

Yonghua Qu, Yuzhen Zhang, and Huazhu Xue

Abstract—The leaf area index (LAI) is a characteristic parameter of vegetation canopies. This parameter is significant in research on global climate change and ecological environments. The China HJ-1 satellite has a revisit cycle of four days and provides CCD (HJ-1 CCD) data with a resolution of 30 m. However, the HJ-1 CCD is incapable of obtaining observations at multiple angles. This is problematic because single-angle observations provide insufficient data for determining the LAI. This article proposes a new method for determining the LAI using the HJ-1 CCD data. The proposed method uses background knowledge of the dynamic land surface processes that is extracted from MODerate resolution Imaging Spectroradiometer (MODIS) LAI data with a resolution of 1 km. The proposed method was implemented in a dynamic Bayesian network scheme by integrating an LAI dynamic process model and a canopy reflectance model with the remotely sensed data. The validation was conducted using field LAI data collected in the Guantao County of the Hebei Province in China. The results showed that the determination coefficient between the estimated and the measured LAI was 0.791, and the RMSE was 0.61. The results suggest that this algorithm can be widely applied to determine high-resolution leaf area indexes using data from the China HJ-1 satellite even if the information from single-angle observations are insufficient for quantitative application.

Index Terms—Bayesian method, HJ-1 CCD, leaf area index, MODIS.

I. INTRODUCTION

THE leaf area index (LAI) is one of the most important structural parameters in land ecosystems, and it influences the substance and energy exchange between vegetation and the atmosphere. The LAI is also a key input factor for hydrological, ecological, and climate models [1], [2]. Satellite remote sensing is an effective way to obtain regional LAI data. Currently, most systems that can operationally determine the LAI from remote sensing data are based in North America and Europe, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI, the VEGETATION CYCLOPES global products, and the Advanced Very-High-Resolution Radiometer (AVHRR) LAI products [3]–[5]. These data sets provide LAI products with a coarse resolution (approximately 1 km). Higher-resolution (approximately 30 m) products may provide important input data for land and resource monitoring and ecological and environmental evaluations. In the past decades, many studies have attempted to produce higher-resolution LAI products [6]–[8].

The first land environment satellite (HJ-1) launched by China in 2009 has a revisit cycle of 4 days and a 4-band CCD camera (Blue: 0.43–0.52 μm, Green: 0.52–0.60 μm, Red: 0.63–0.69 μm, and Near-infrared: 0.76–0.90 μm) with a 30-m spatial resolution [9]–[11]. Li et al. [12] compared the spectral signature of HJ1-CCD and Landsat5 TM sensors and found that the coefficient of determination approached unity considering the imaging and spectral response characteristics. Currently, concern about estimating the vegetation parameters from HJ-1 CCD data is increasing [13]. However, due to the weather conditions (e.g., cloud cover), most of HJ-1 CCD data are hard to interpret and use, which will lead to time-series discontinuities and low precision levels. Thus, in this study, additional MODIS time-series information was incorporated into the estimation of the LAI to compensate for the insufficiencies in the HJ-1 CCD data.

The time-series MODIS LAI can provide vegetation growth information, which can effectively constrain the temporal trajectories of the estimated LAI from the HJ-1 CCD data. This hypothesis has been widely used in remote sensing data assimilation to enhance the temporal continuity of the estimated results [14]. We implemented this under the framework of dynamic Bayesian networks (DBN) because it addresses uncertainties using a Bayesian probability [15], [16]. Our previous studies demonstrated the efficiency of DBN in estimating time-series LAI from multi-source datasets [17], [18]. The purpose of the present study was to study the use of the dynamic information of land surface parameters extracted from coarse-resolution historical data to generate quantitative high-resolution LAI from China HJ-1 CCD data.
II. DATASETS

A. The Study Area

The field experiment was conducted in Guantao County in the Hebei Province (115.13°E, 36.52°N) of China. The 6-m resolution imagery from Google Earth of the study area is shown in Fig. 1(a). The experimental observation team designed the field observation scheme corresponding to the pixel scale of the HJ-1 CCD. The LAI values were measured using the LAI-2000 instrument on June 22, July 24, August 14, and September 16, 2010. The distribution of the LAI sampling points approximately forms a cross, as illustrated by the red cross shown in Fig. 1(a). All of the sampling spots surround an automatic meteorological station (Fig. 1(b)). The spatial extent of study area is approximately 3°3 km\(^2\) and is shown as a blue square in Fig. 1(a). During this time period, the land cover in the research area was dominated by corn, as shown in Fig. 1(c) and Fig. 1(d), which show photographs of the crop growth state in the spring and summer, respectively. In addition, the geological coordinates of some of the crossroads were also measured for geometric precision corrections of the HJ-1 satellite data.

B. MODIS LAI Product

The MCD15A2 product, which was constructed from 8 days of observation data with a spatial resolution of 1 km, was adopted as the data source for constructing the background information. The temporal distribution is from the Julian day 161 to Julian day 273 in 2010. The time-series information of the MODIS LAI data reflects the dynamic changes of the LAI. Using a dynamic process model driven by the MODIS LAI data, the background information was introduced into the estimation of the time-series LAI by providing prior knowledge of the vegetation LAI dynamics during the growth season, which may improve the accuracy and temporal continuity of the retrieved results.

C. HJ-1 CCD Data

We chose images with low cloud cover (<10% overall) for the LAI retrieval. There are four HJ-1 CCD images that satisfy this requirement during the period from June 22, 2010 to September 16, 2010, when the field LAI measurements were available. The satellite data were available on June 28, 2010, July 6, 2010, July 20, 2010, and August 16, 2010. Currently, the available HJ-1 CCD data are level-2 products that have not been precisely geometrically corrected. Before the HJ-1 CCD data were applied to the LAI retrieval, the data were re-processed using precise geometric and atmospheric corrections and transformed into surface reflectance values. We used the GPS coordinates of the crossroads with typical geological symbols in Google Earth as auxiliary targets for the geometric correction of the HJ-1 CCD images. The correction precision (RMSE) of the four images was equal to 0.2122, 0.2176, 0.0862, and 0.3417 pixels, respectively. Then, a FLAASH algorithm was used for the atmospheric correction to obtain the corresponding surface reflectance values.
III. ALGORITHM AND MODEL

A. Time-Series Bayesian Model

Fig. 3 shows schematic diagrams of the time-series LAI retrieval using dynamic Bayesian networks extended to the time slice T. The round nodes represent random variables. The gray nodes represent the observed reflectance, and the non-filled nodes represent the target variable LAI. Ref\textsubscript{T} is the reflectance at time slice T, and LAI\textsubscript{T} is the LAI at time slice T. The directional arcs connecting the nodes in Fig. 3 represent the dependence or influence relationship between the variables. The arcs pointing to the right represent the state transition relationship between two adjacent time slices of the LAI (the green dashed frame in Fig. 3). The transition relationship is represented quantitatively as a dynamic process. The arcs pointing upward represent the mapping relationship between the target variable LAI and the observation variable Ref (indicated by the red curved frame). The mapping relationship is represented as a canopy reflectance model.

Assumed that the remote sensing observation of the time series has already been obtained: \( Ref_1, Ref_2, Ref_3, \ldots, Ref_{T-1}, Ref_T \). Then, the posterior probability of LAI\textsubscript{T} at the Tth time slice is \( P(LAI_T|Ref_1, Ref_2, Ref_3, \ldots, Ref_{T-1}, Ref_T) \), which can be expressed as \( P(LAI_T|Ref_{1:T-1}, Ref_T) \). According to the Bayesian principle (detailed information are referred in our published work [17]).

Equation (1), shown at the bottom of the page, is the basic algorithm for LAI retrieval using the DBN method. This equation integrates the dynamic information of the LAI, the canopy reflectance model, and the remote sensing observations to compute the posterior probability distribution of LAI. The core of the DBN method consists of the following four steps: 1) obtaining the posterior probability distribution of LAI at the previous moment, which is described as \( P(LAI_{T-1}|Ref_{1:T-1}) \), 2) calculating the state transition probability \( P(LAI_T|LAI_{T-1}) \) from the dynamic process model, 3) calculating the likelihood probability \( P(Ref_T|LAI_T) \) of the observational data using the look-up table generated by the canopy reflectance model, and 4) calculating the posterior probability of the LAI at the Tth time slice based on (1). The LAI time series is obtained through the sequential iteration of the four steps. The estimated LAI can be either the mean value or the maximum probability value of the distribution. In this study, we adopted the mean value as the retrieval result.

B. The Canopy Reflectance Model and Look-Up Table

The SAILH (Scattering by Arbitrarily Inclined Leaves with Hot spot effect) model proposed by Kuusk is an extension of the SAIL model [19], [20]. The canopy structure is assumed to be a horizontally well-distributed turbid medium, which is more suitable for the canopy of crops. The bidirectional reflectance function of the canopy was considered to be a function involving the sun, observational geometry, canopy structure parameter (LAI and leaf inclination angle distribution function), and leaf/soil spectral features. Because it takes into consideration the influence of leaf size and shadow, it can describe the hot-spot effects of the canopy. In most studies, the LAI was retrieved using an inverting SAILH model that required the input of parameter values, except for the free LAI parameter [19], [21]. In most inversion problem, a certain optimization algorithm is used for the iteration to minimize the error between the simulated canopy reflectance and the observed data [22], [23]. An efficient way to

\[
P(LAI_T|Ref_{1:T}) = \frac{P(Ref_T|LAI_T) \times \sum_{LAI_{T-1}} P(LAI_T|LAI_{T-1}) \times P(LAI_{T-1}|Ref_{1:T-1})}{\sum_{LAI_T} P(Ref_T|LAI_T) \times P(LAI_T|Ref_{1:T-1})}
\]
C. LAI Dynamic Process Model

The MODIS LAI products were chosen and re-analyzed for dynamic process fitting and for calculating the state transition probability. Under ideal conditions, one can assume that the temporal trajectories of the MODIS LAI data coincide with that of vegetation growth. In this case, a simple equation can be used to represent the dynamic vegetation processes.

However, many studies have shown that MODIS LAI data are likely to be overestimated or underestimated in some regions, which leads to temporal and spatial discontinuities. The influence of these factors should be reduced or eliminated before using the MODIS LAI data in the dynamic process model used to conduct the LAI estimation. In this study, the MODIS time-series LAI were filtered to reduce or eliminate any possible temporal and spatial discontinuities and thus generate temporally and spatially continuous MODIS LAI data. The filtered MODIS LAI data were used to simulate the dynamic process of vegetation growth using the following equation:

\[ I_{t+1} = I_t \times \frac{l_{t+1}}{l_t} \]  \hspace{1cm} (2)

where \( I_{t+1} \) is the LAI at time \( t \) and \( l_t \) is the filtered MODIS LAI at time \( t \). \( I_t \) is the filtered MODIS LAI (or \( l_t \)). It should be noted that, for the first time slice, there was no predicted LAI from the process model. Thus, we used the MODIS LAI directly as the initial driving data of (2) and then (1).

However, the spatial resolution difference between the MODIS LAI data and the HJ-1 satellite observations is very significant: one MODIS pixel contains approximately 33-by-33 HJ-1 CCD pixels. Therefore, from the perspective of probability and statistics, we considered the MODIS LAI value as the overall average information of the 33-by-33 HJ-1 CCD pixels. This average information can represent the prior distribution of the individual pixel distribution. Every pixel of the HJ-1 CCD can be independently sampled multiple times to obtain the overall average value. The pixels that are sampled independently share the same background information, and their spatial heterogeneity can be represented as the difference between the pixels in the HJ-1 CCD data, i.e., the HJ-1 CCD reflectance was used to update the background a priori probability distribution of the LAI.

IV. RESULTS AND VALIDATION

Based on the HJ-1 CCD data and the process model fitted with the MODIS LAI, the time-series LAI were retrieved using the
Fig. 4. Spatial and temporal pattern of the estimated LAI from the HJ-1 CCD and the MODIS data for the year 2010 in Guantao County. The spatial extent consists of 112 rows and 121 lines. The LAI for June 28, 2010, July 6, 2010, July 20, 2010, and August 16, 2010 were estimated from the HJ-1 CCD data and the MODIS data, whereas the LAI for July 12, July 28, August 5, and August 21 were estimated from the MODIS data alone through the dynamic process model.

The spatial and temporal patterns of the estimated LAI in one growth season are shown in Fig. 4. The estimated LAI for the day of June 28 is based on prior information provided by the MODIS LAI that was updated with the reflectance data from the HJ-1 satellite. Based on this, the dynamic process model was used for further estimation of the LAI distribution with a 30-m pixel resolution. When HJ-1 CCD data are available, the posterior probability distribution of the LAI was updated with the HJ-1 CCD data. As mentioned in Section II-C, the HJ-1 satellite observations were available for four days: June 28, 2010, July 6, 2010, July 20, 2010, and August 16, 2010. Thus, the estimated LAI for these days includes the contribution of both the dynamic process model and the HJ-1 CCD data. In contrast, for July 12, July 28, August 5, and August 21, 2010, the LAI were retrieved from the dynamic process model alone. When the observed data are severely insufficient, a dependence on the model prediction alone could cause a deviation in the model estimated value from the true value. This deviation could only be corrected when the observed data were again available. Fig. 4 suggests that, when no observed data are available, the LAI results based on the process model are inferior to the retrieved results based on the combination of observations and model predictions. In this case, the estimated LAI values are more affected by the MODIS LAI data. As a result, the LAI of some pixels exhibit the same value as their surrounding pixels within an area. However, with the incorporation of additional HJ-1 CCD data, the texture information became richer (Fig. 4). To evaluate the LAI retrieval results, the field measurement from June 22, 2010, July 24, 2010, and August 14, 2010 were used to validate the LAI values of the corresponding pixels on 2010-06-28, 2010-07-20, and 2010-08-16 (Fig. 4), respectively. The results are shown in Fig. 5. The determination coefficient ($R^2$) between the measured value and the estimated value was 0.791, and the RMSE was 0.61. This result showed a high consistency between the retrieved value and the measured value. However, it can also be found that the retrieved LAI of some pixels exhibited large deviations from the field-measured values partially because the measured LAI and the retrieved LAI were not temporally matched. For example, the retrieved LAI from 2010-07-20 and the observed LAI from July 24, 2010 had a 4-day lag. During periods of rapid vegetation growth, a four-day lag could cause a significant deviation. In addition, errors in the HJ-1 CCD surface reflectance data could also cause a large deviation in some pixels. When the observed reflectance had errors in a specific spectral band, the weight of a priori LAI increased compared to the reflectance data, and the retrieved result would deviate toward the prior LAI. In such cases, the accuracy of the retrieved results of the corresponding pixels would be compromised.

V. CONCLUSIONS

To overcome difficulties in the retrieval of high-resolution quantitative products from remotely sensed data obtained by the Chinese environmental satellite HJ-1, this paper proposed a method that utilizes prior knowledge extracted from low- and medium-resolution historical data products. By expanding the traditional Bayesian retrieval model, dynamic changes in the land surface LAI were introduced to form a new method for the Bayesian retrieval of time-series LAI. In the proposed method, the dynamic process model, the canopy reflectance model, and
the remotely sensed data are combined. The algorithm validation was conducted using LAI data based on HJ-1 CCD data and field measurements. The validation results show that the LAI retrieved using the time-series Bayesian algorithm developed in this paper were consistent with measured values. The retrieval algorithm fully utilized the high spatial resolution of the HJ-1 CCD data and the time-series information of the MODIS LAI data. This allowed the extraction of land surface parameter dynamic information from low-resolution data, which assisted in the LAI retrieval. This method has the potential to generate high-resolution LAI products with the assistance of low-resolution remote sensing data.

In the present work, the proposed algorithm was only validated using crops. For other types of land cover, the look-up table must be generated using the corresponding canopy reflectance model. Therefore, before popularizing this method for the operational generation of high-resolution LAI products, existing land surface classification maps are needed, and the LAI retrieval should be conducted based on different look-up tables for different types of land cover. In addition, additional validation work for remote sensing products must be conducted. Field measurements must be acquired for different land types to validate the quality of the LAI retrieval. The present paper only used approximately 150 measurement spots for the validation of the retrieval results. In future work, additional data from different land types should be included in the LAI retrieval process, and validations should be conducted in different regions.

ACKNOWLEDGMENT

The authors would like to thank Prof. L. Shaomin and Prof. W. Jindi, as well as the graduate students C. Ping, Y. Xuejun, L. Yan, G. Libiao, and Z. Kai in the School of Geography, Beijing Normal University. These individuals provided the ground-measured LAI data and the geometry control points for the experiment in Guantao County, Hebei Province, China.

REFERENCES


Yonghua Qu received the M.Sc. degree in geography from China University of Mining Technology, Jiangsu, China, in 2000, and the Ph.D. degree in remote sensing from the Beijing Normal University, Beijing, China, in 2005. From 2007 until 2012, he held a position as Assistant Professor in the School of Geography at Beijing Normal University. Since March 2012, he has been working as Project Leader in Remote Sensing Inversion Using Dynamic Bayesian Network which is supported by National Natural Science Foundation of China.

Yuzhen Zhang received the M.Sc. degree from Beijing Normal University, Beijing, China, in 2011. She is currently pursuing the Ph.D. degree at Beijing Normal University. Her research interests focus on forest biomass mapping from lidar, optical and radar data, assessment of impacts of forest disturbances on forest biomass dynamics, and radiative forcing of forest disturbances.

Huazhu Xue received the Ph.D. in Cartography and Geographic Information Systems in the area of Quantitative Remote Sensing from the School of Geography, Beijing Normal University, Beijing, China, in 2012. His research interests include vegetation parameters inversion, satellite image processing and GIS applications. From 2012, He held a position as Assistant Professor in the School of Surveying and Land Information Engineering, Henan Polytechnic University, Henan Jiaozuo.