AN ALGORITHM FOR THE DIRECT ESTIMATION OF NDVI WITH LANDSAT TOA REFLECTANCE

Qingwei Wu[1] and Tao He[1]
[1]School of Remote Sensing and Information Engineering, Wuhan University, No.129 Luoyu Rd. Hongshan District, Wuhan, Hubei 430079, China
Email: wuqingwei@whu.edu.cn; taohers@whu.edu.cn

KEY WORDS: OLI, Vegetation index, BRDF, Segmented model

ABSTRACT: Vegetation index can effectively describe the coverage, distribution and growth of vegetation. As a simple and sensitive vegetation index, normalized difference vegetation index (NDVI) is widely used in crop yield estimation, surface classification and vegetation extraction. However, due to the influence of atmospheric conditions and other factors, the atmospheric correction results in some areas are uncertain and even difficult to obtain accurate atmospheric variables (e.g., aerosol optical depth), resulting in the inability to estimate NDVI from surface reflectance. This paper introduces an algorithm to directly estimate NDVI with Landsat top of atmosphere (TOA) reflectance. The algorithm builds a segmented regression model of TOA reflectance and surface NDVI, obtains the model coefficients under different geometric conditions, and generates a look-up table to store regression coefficients. Through this algorithm, the instantaneous NDVI can be obtained directly from the TOA reflectance, which not only provides a way to obtain the surface NDVI for the areas where the atmospheric correction is difficult or the atmospheric correction results are not accurate, but also can be used to produce high-resolution surface NDVI products with spatiotemporal continuity. In order to verify the accuracy and reliability of our algorithm, we use the Landsat 8 TOA reflectance data of Hunan area in China to calculate the surface NDVI, and use another surface NDVI products obtained by atmospheric correction and band operation to verify the accuracy of the estimation results, count the relevant accuracy indicators, and make a comparative analysis. The statistical results show that the accuracy of NDVI estimated by this method is good, and the overall results are reliable.

1. INTRODUCTION

Vegetation, as a widely distributed and ecologically significant land cover type, has great influence on climate, environment and hydrology, and has always been the focus of global ecological research. Vegetation index is composed of several bands of multispectral remote sensing data through linear, nonlinear combination or spatial transformation. It is one of the important parameters to describe the characteristics of surface vegetation, and can reflect the growth characteristics, growth status, coverage, distribution characteristics, biomass and so on (Guo, 2003). More than 40 vegetation indices are currently in use and are widely used in crop yield estimation, surface classification, drought monitoring, vegetation extraction, climate change analysis and so on (Liang et al., 2013). Among them, NDVI proposed by Rouse has relatively simple calculation and concentrated numerical range. Through the nonlinear combination between red band and near-infrared band, the difference between reflectance is strengthened. It can not only eliminate the influence of radiometric calibration, atmosphere and observation angle of some sensors, but also reflect the growth of vegetation effectively and have high sensitivity to the change of vegetation. Therefore, it has become the most widely used vegetation index (Rouse, 1973).

Due to the difference of atmospheric conditions, the change of observation angle, the difference of radiometric calibration of sensors and the difference of spectral response function setting, the NDVI between images can not be directly compared and analyzed. In order to monitor the vegetation change dynamically, it is necessary to use the vegetation index products of long time series as the research parameters to empirically measure the change of vegetation. At present, the common long-time series vegetation index products mainly include GVI, GIMMS and PAL products of AVHRR data, MOD13 and MYD13 products of MODIS data, VGT-S10 and VGT-D10 products of VEGETATION data (Zhang et al., 2020). Among these products, the highest temporal resolution is only 8-days, and the highest spatial resolution is only 250 meters. There are few vegetation index products with medium or high spatial resolution, and it is difficult to eliminate the influence of atmospheric noise and cloud pollution. In the analysis and research of vegetation remote sensing, the reliable and large-scale vegetation index products with spatial-temporal continuity are urgently needed. Therefore, it is necessary to study the direct estimation method of NDVI. In this paper, an algorithm for the direct estimation of NDVI with Landsat TOA reflectance is presented. By establishing a segmented regression model and generating a look-up table, the NDVI with high temporal resolution and good continuity can be obtained by simple calculation, which not only provides data support for producing NDVI time series products, but also lays a foundation for subsequent related research (Li et al., 1995) .

2. DATA AND ALGORITHM

2.1 Data and Preprocessing

The MODIS BRDF database is established based on the MODIS BRDF product (MCD43A1). A total of 2399 MODIS BRDF data samples were collected from MODIS BRDF products. The samples included a variety of land cover types, including vegetation, soil, water, snow and so on. The 2399 samples collected were selected randomly, but in order to ensure the reliability and accuracy of the data, only samples with good quality of all bands in MODIS BRDF products were selected. Each sample represents a set of BRDF parameters corresponding to one pixel point. Because of the difference between MODIS and Landsat 8 spectral responses, MODIS BRDF parameters can not be directly used for Landsat reflectance calculation. The MODIS BRDF parameters of each band need to be multiplied by the band conversion coefficients to obtain the Landsat BRDF parameters (He et al., 2015).

TOA reflectance of different sensors, different bands, different aerosol optical depth and different angles was simulated by the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) model. The 6S model is written in Fortran programming language and is used to simulate the atmospheric radiative transfer in the solar reflection band, it takes into account the entire radiative transfer of solar radiation energy from the top of the atmosphere to the surface and the reflected radiation energy from the surface to the sensor through the atmosphere (Liang et al., 2013). By inputting the corresponding parameters and using the 6S model to calculate, the required TOA reflectance data were obtained.

2.2 Spectral Bidirectional Reflectance Distribution Function

Spectral bidirectional reflectance distribution function (BRDF) is a function that can completely describe the directional reflection characteristics of a surface. It is defined as the ratio of spectral radiance reflected from the surface of an object in one direction to spectral radiance incident on the surface of an object in another direction (Liang, 2009). The surface BRDF statistical model can be roughly divided into three types: physical model, empirical model and semi-empirical model. The Ross-Li kernel-driven model used in our algorithm belongs to the semi-empirical model, which has the advantages of both empirical model and physical model. Through the approximation and simplification of the physical model, the semi-empirical model not only has some physical meaning, but also reduces its complexity and makes the calculation more simple (Yang et al., 2002) .

The kernel-driven model is one of the most widely used semi-empirical models. It combines and simplifies the radiative transfer model of horizontal uniform canopy and the canopy geometric optical model. It is generally composed of isotropic kernel, volume scattering kernel and geometric optical kernel (Li, 1989). Its general expression is as follows:

$$R(\theta_s, \theta_v, \varphi; \lambda) = f_{iso}(\lambda)k_{iso} + f_{geo}(\lambda)k_{geo}(\theta_s, \theta_v, \varphi) + f_{vol}(\lambda)k_{vol}(\theta_s, \theta_v, \varphi) \quad (1)$$

Among them, the k_{iso} is an isotropic kernel function, the k_{geo} is the geometric optical kernel function, the k_{vol} is the volume scattering kernel function, the f_{iso}, f_{geo} and f_{vol} are the coefficients of each item, the λ is the wavelength, the θ_s, θ_v and φ are solar zenith angle (SZA), viewing zenith angle (VZA) and relative azimuth angle (RAA).

In this study, the value of k_{iso} is constant 1, and the k_{geo} used the RossThick kernel which was used to describe the dense vegetation canopy (Roujean et al., 1992; Ross et al., 1981). The k_{vol} used the LiSparseR kernel which was proposed by Lucht et al. (Lucht et al., 1998; Lucht et al., 2000). It is obtained by rewriting the LiSparse kernel proposed by Wanner according to the reciprocity principle (Wanner et al., 1995), and it is a simplified form of geometric optics model, which is used to describe the sparse canopy distributed on the ground background of Lambert scattering.

2.3 Normalized Difference Vegetation Index

The construction of vegetation index is based on the reflection spectrum characteristics of vegetation to incident radiation. The vegetation index is an important parameter that can effectively, directly and simply measure the surface vegetation (Guo, 2003). NDVI makes use of the characteristics that vegetation reflects more near-infrared light and absorbs more red light, combines the near-infrared band and red band in the reflection or radiation of vegetation canopy in a proportional form, and increases the contrast between the red band and the near-infrared band. Generally, that the value of NDVI is low means that the contrast between the red band and the near infrared band is small and the corresponding area is an area with no or little vegetation. On the contrary, it is the vegetation area. The value of NDVI is limited to (-1, 1). The value is centralized and easy to use. It can suppress the influence of sensor calibration,

atmospheric change, terrain, solar angle and soil. The contrast between the reflection of soil and vegetation is enhanced by nonlinear stretching, which can enhance the low value range and inhibit the high value range, and enhance the sensitivity to vegetation, the formula is as follows:

$$\mathrm{NDVI} = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \tag{2}$$

Among them, ρ_{NIR} and ρ_{RED} are the reflectance of near infrared band and red band respectively.

2.4 NDVI Estimation Algorithm and Application

The flow chart of the research algorithm used is shown in Figure 1.

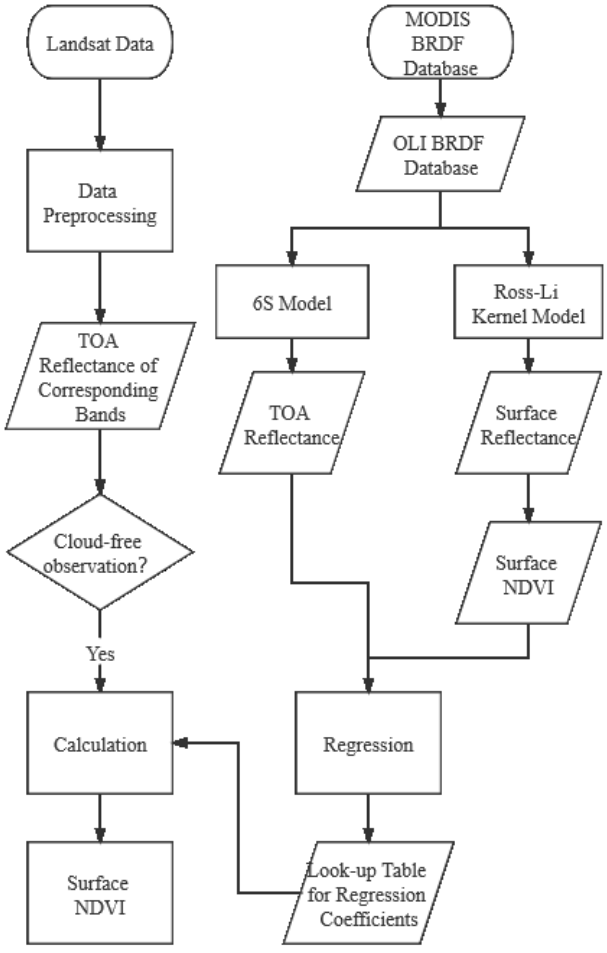


Figure 1. Flow chart of NDVI direct estimation method

For the instantaneous NDVI, set different angles of SZA, VZA and RAA, use the converted Landsat BRDF database, substitute the angle values and BRDF parameters into Ross-Li kernel-driven model, and calculate the surface reflectance under ideal conditions, and obtain the surface NDVI through band operation.
The surface NDVI is regressed with the TOA reflectance simulated by 6S model to establish a segmented regression model, where a_i is the regression coefficient of the model, ρ_i is the value of TOA reflectance or the reflectance combination of different bands used in the model, n is the number of ρ_i. The model is segmented according to the range of TOA NDVI value. Through regression and fitting, the segmented model coefficients under different SZA, VZA and RAA are obtained, and the corresponding look-up table is established. The model is formulated as follows:

$$\mathrm{NDVI} = a_0 + \sum_{i=1}^{n} a_i \cdot \rho_i \tag{3}$$

The fitting method uses the least square method and adds boundary constraints. The parameters are solved by the following formula:

$$\begin{cases} \operatorname{argmin} \frac{1}{2} \left\| \mathrm{NDVI} - (a_0 + \sum_{i=1}^{n} a_i \cdot \rho_i) \right\|_2^2 \\ -1 \le a_i \le 1 \end{cases} \tag{4}$$

Using the model coefficients in the look-up table, the instantaneous NDVI can be estimated from the Landsat TOA reflectance data by simple calculation. In the application, it is necessary to remove the cloud from the image firstly, then get the angles, time and position information from the image, find the corresponding model coefficients, and finally bring the TOA reflectance values of the corresponding band into the model to estimate the corresponding NDVI values.

3. EXPERIMENT

3.1 Study Area and Time

The study area is located in Hunan, China, with latitude from 24 ° 38’ N to 30 ° 08’ N, longitude from 108 ° 47’ E to 114 ° 15’ E, covering an area of about 211,800 square kilometers. The location of the study area is shown in Figure 2.

The study area covers a wide range of terrain, including plains, basins, hills, mountains, rivers, lakes and so on, with abundant vegetation cover and forest coverage of 59.9%. It is a typical sub-tropical monsoon climate with cloudy and rainy weather throughout the year, and the remote sensing images with good data quality and cloud-free coverage are relatively rare. Figure 2 is NDVI image of Hunan Province in the third quarter of 2019. It can be seen from Figure 2 that Hunan Province has dense and widely distributed vegetation and high vegetation coverage.

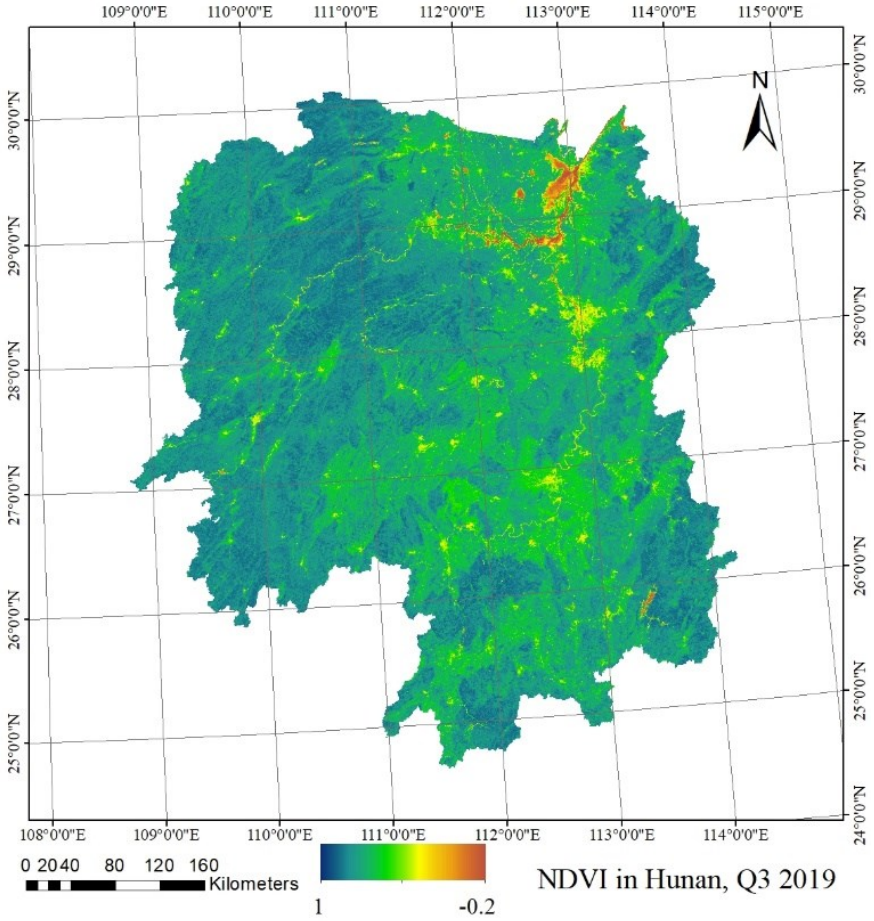


Figure 2. Location of study area

3.2 Landsat 8 Data and Processing

Landsat 8 uses the Worldwide Reference System 2 (WRS-2) to classify and number images, and each image corresponds to a specific set of numbers. Hunan Province involves 21 Landsat 8 OLI images in space. Based on the experimental needs, 117 Landsat 8 OLI images in summer of 2019 and 2020 were selected. The selected TOA reflectance data were substituted into the model to calculate the surface NDVI. In order to verify the accuracy and reliability of the model, the atmospheric correction of the TOA reflectance data and the band operation were used to obtain the NDVI. The NDVI directly estimated by the model was compared with the NDVI calculated by atmospheric correction and band operation, and the relevant results were counted. In order to get rid of the invalid information in the cloudy area, Fmask 4.0 was used to mask the NDVI images, and only the non-cloud pixel points of the images were preserved. The comparison before and after the mask is shown in Figure 3. In the comparison experiment, the values of cloud and NDVI less than -0.2 were removed.

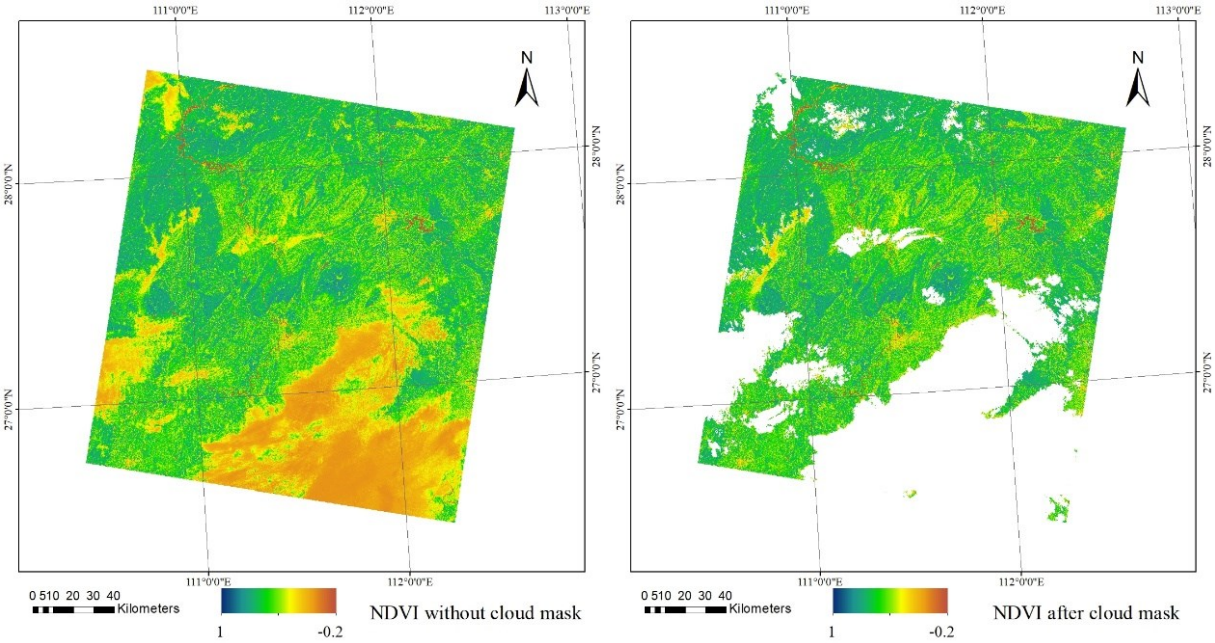


Figure 3. Comparison before and after cloud mask

3.3 Land Surface NDVI Estimation Results and Validation

To validate and analyze the NDVI products obtained by using the NDVI direct estimation algorithm based on Landsat TOA reflectance proposed in this paper, we used Landsat 8 TOA reflectance data in Hunan Province to estimate NDVI products by our model, and used another surface NDVI products obtained by atmospheric correction and band operation to verify the accuracy of the estimation results, counted the relevant accuracy indicators, and made a comparative analysis. The statistical results of the determination coefficient of these two images (R^2) of all the experimental data are shown in Table 1. The verification results are shown in Figure 4. Some verification results in the summer of 2019 and 2020 were selected for display. Visually compare the NDVI images with higher accuracy and lower accuracy in all results, and the results are shown in Figure 5.

Table 1. Statistical results of R^2 of experimental data

R^2	<0.7	0.7-0.8	0.8-0.85	0.85-0.9	0.9-1
Quantity	13	13	13	29	49
Proportion	11.11%	11.11%	11.11%	24.79%	41.88%

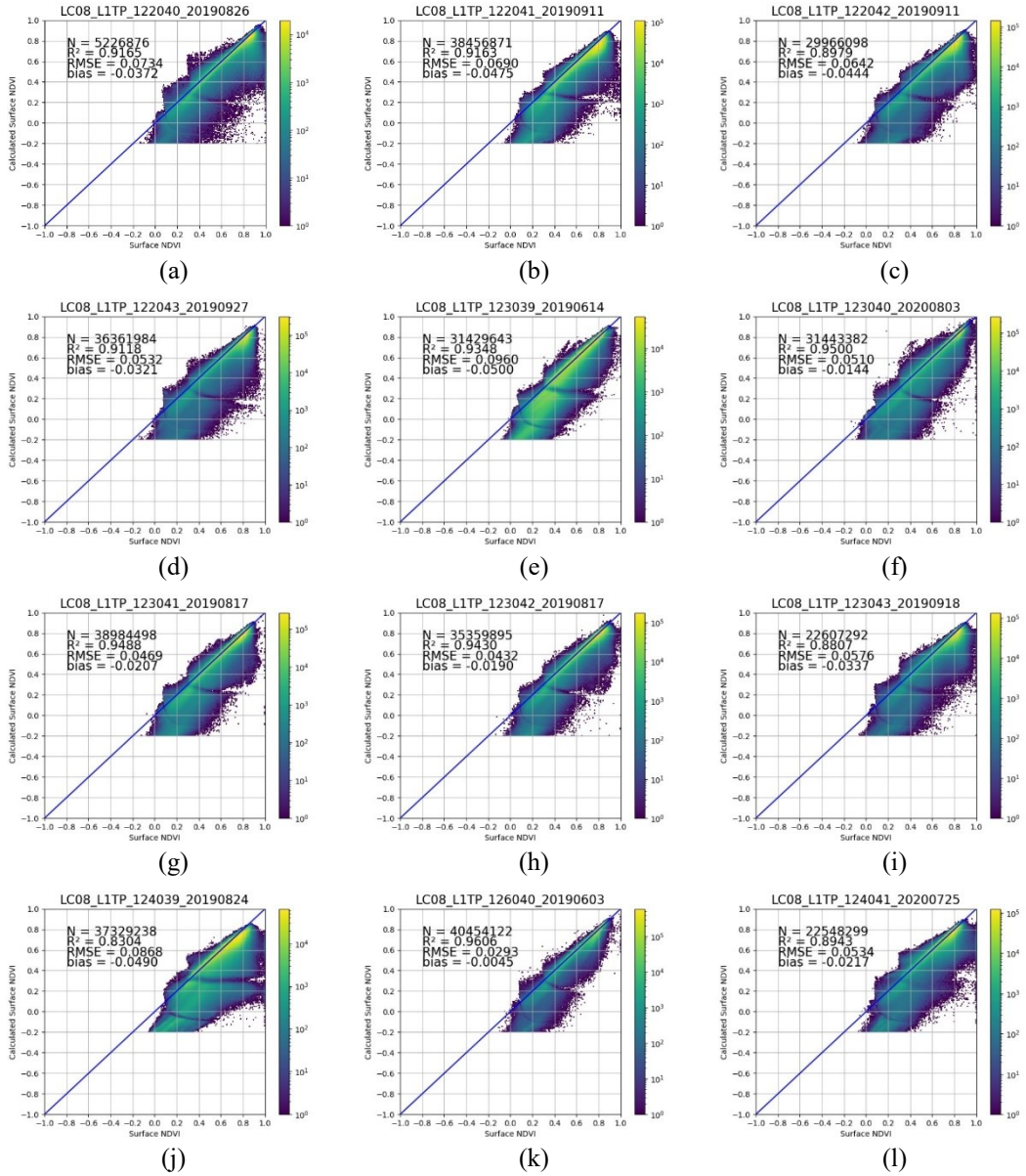
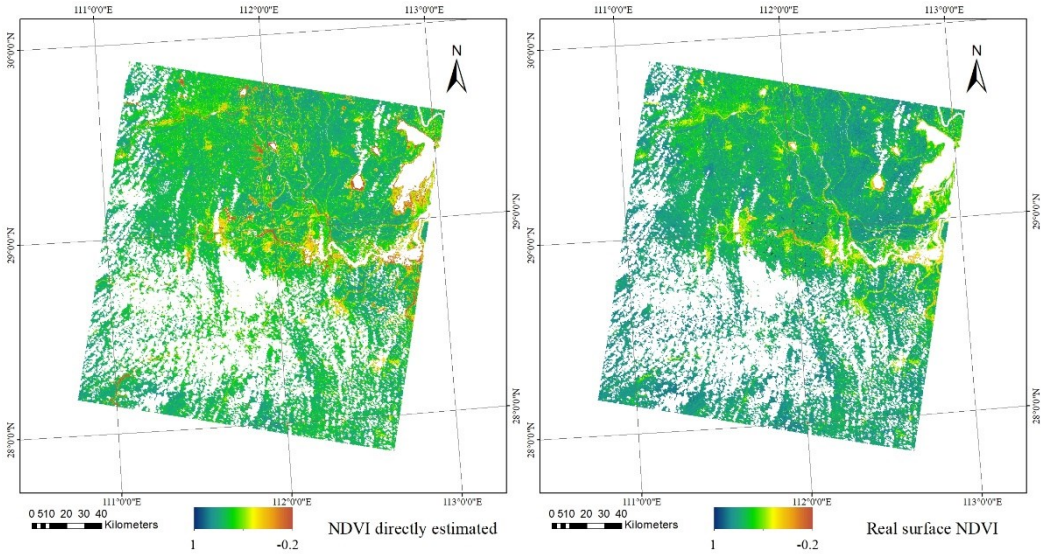


Figure 4. Comparison results of NDVI

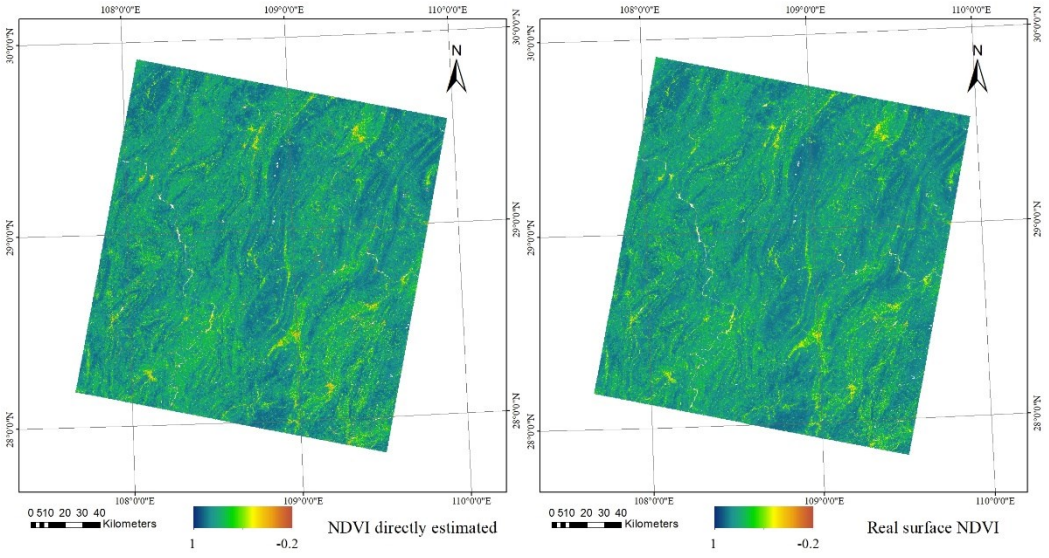
In this experiment, the NDVI obtained by atmospheric correction and band operation is regarded as the real surface NDVI, which is written as real surface NDVI directly. In Figure 4, the horizontal coordinates represent the real NDVI values, and the vertical coordinates represent the NDVI values estimated by this algorithm. The R^2, root mean square error (RMSE), bias, and the number of pixel points (N) involved in the statistics are given. Through the scatter density charts and three evaluation indicators in Figures 4, we can see that the estimation results of the model basically meet the accuracy requirements. The results show that the R^2 values of nearly 80% images are greater than 0.8, more than half of the values are above 0.9, and the bias values of all images are less than 0.1. In general, the estimated NDVI results have high accuracy, but there are still some deficiencies. As can be seen from Figure 4(e) and Figure 4(j), since the model is a segmented model, there is an obvious boundary at the segmentation threshold. In addition, on the whole, the estimation result of the model is lower than the real value. Table 1 shows the statistical results of R^2 of all the experimental data, including images with high cloud coverage or poor data quality, so the statistical results are slightly lower than those in Figures 4, but still have ideal precision and reliability, the R^2 values of 41.88% images are more than 0.9, and the R^2 values of more than half of the images are more than 0.85. Therefore, it can be considered that this algorithm overcomes the influence of data quality and estimates the reliable results.

Figure 5(a) is a group of images with low accuracy. It is located in the north of Hunan, with low vegetation coverage, more water bodies and large area of cloud pollution. It can be seen from the figure that the overall estimated NDVI value is low, and the pixel points with obvious error are mainly concentrated in low vegetation coverage areas, such as water body, bare soil and other areas, as well as areas with obvious mountain shadow. Figure 5(b) shows a group

of images with high accuracy of estimation results. The corresponding area has high vegetation coverage, almost cloud-free, relatively flat terrain. The pixel points different from the real value are still concentrated in the areas with low NDVI value.



(a)



(b)

Figure 5. Estimated NDVI image and real surface NDVI image

Based on the comparison of the scatter density charts and three statistical indicators, and the visual comparison of the two groups of images with high and low accuracy. It can be concluded that the NDVI direct estimation algorithm based on Landsat TOA reflectance proposed in this paper has high accuracy and can meet the subsequent application requirements. The values of the three statistical indicators verify the reliability of this method. However, in areas with low vegetation coverage and large topographic relief, the accuracy of the estimation results will be reduced to some extent. In addition, in areas with high cloud coverage, even if the clouds are removed, the estimation results of NDVI may be affected to some extent.

4. DISCUSSION AND CONCLUSION

NDVI is one of the most commonly used indices in vegetation analysis. It is an important parameter in the subsequence analysis of remote sensing images and has been widely used in agricultural remote sensing. It has important guiding significance in vegetation monitoring, phenology analysis, information extraction, crop yield estimation, drought monitoring, forest disturbance monitoring and so on. At present, the main NDVI products of long time series often have low temporal and spatial resolution, and they are easily affected by the factors such as angle, atmospheric condition, soil background, topography and so on.

The algorithm of directly estimating NDVI based on Landsat TOA reflectance proposed in this paper constructs a segmented regression model between the TOA reflectance data and surface NDVI, and generates a look-up table. Through simple calculation, the surface NDVI can be directly estimated from the TOA reflectance data, and there is no need for atmospheric correction. It provides a convenient and feasible way to estimate NDVI for the areas where the atmospheric correction is difficult or the atmospheric correction results are not accurate.

The main research contents and conclusions of this paper include the following three points:

(1) An algorithm that can directly estimate NDVI from TOA reflectance is proposed, and a look-up table is generated. Through the look-up table, the NDVI can be directly estimated, which is simpler and more convenient than the traditional method of calculating NDVI. The NDVI whose accuracy meets the requirements of subsequent processing can be obtained without atmospheric correction. While simplifying the image processing steps, the reliability of the results is guaranteed.

(2) Landsat 8 OLI images of Hunan Province in China in the summer of 2019 and 2020 were selected to verify the accuracy and reliability of this method. The difference between the NDVI directly estimated and the NDVI obtained by atmospheric correction and band operation was compared through experiments. Three evaluation indicators (R^2, RMSE, and bias) were calculated and the results were statistically analyzed. The R^2 values of more than half of the images are greater than 0.85, and the RMSE and bias values of more than 90% of the images are less than 0.1. The statistical results show that the NDVI estimated by this algorithm performs well in accuracy, and the overall result is relatively reliable.

(3) Combined with the statistical results, some images with slightly lower accuracy were compared and analyzed, and some problems and deficiencies were summarized. Because it is a segmented model, there is obvious discontinuity in the connection part of the segmented threshold. In addition, the overall value of NDVI directly estimated is lower. In areas with low vegetation coverage, large topographic fluctuation, high cloud coverage or poor data quality, the accuracy of estimation results will be reduced to some extent.

With the development and progress of quantitative remote sensing, the research on vegetation becomes more and more intensive. Relevant research in many fields urgently needs large-scale vegetation index products with high spatial and temporal resolution and reliable accuracy. In order to meet the production and application requirements of subsequent NDVI products with spatio-temporal continuity, the next work will further improve and perfect the algorithm in this paper, such as solving the continuity problem of segmented model, improving the product accuracy in some regions, and generating high spatio-temporal resolution sequential NDVI products combined with temporal filtering.

REFERENCES

Guo, N., 2003. Vegetation index and its advances. Arid Meteorology, 21 (4), pp. 71-75.
He, T., Liang, S., Wang, D., Cao, Y. & Min, F., 2018. Evaluating land surface albedo estimation from Landsat MSS, TM, ETM+, and OLI data based on the unified direct estimation approach. Remote Sensing of Environment, 204, pp. 181-96.
He, T., Liang, S., Wang, D., Chen, X., Song, D. & Jiang, B., 2015. Land Surface Albedo Estimation from Chinese HJ Satellite Data Based on the Direct Estimation Approach. Remote Sensing, 7 (5), pp. 5495-5510.
Liang, S., 2009. Quantitative remote sensing. Science Press, Beijing, pp. 20-24.
Liang, S., Li, X. & Wang, J., 2013. Quantitative remote sensing: concept and algorithm. Science Press, Beijing, pp. 142-224.
Li, X., 1989. Characteristics of bidirectional reflectance and directional spectrum of ground objects. Remote Sensing of Environment, 004 (001), pp. 67-72.
Li, X. & Wang, J., 1995. Vegetation optical remote sensing model and vegetation structure parameterization. Science Press, Beijing, pp.70-78.
Lucht, W. & Wolfgang, 1998. Expected retrieval accuracies of bidirectional reflectance and albedo from EOS-MODIS and MISR angular sampling. Journal of Geophysical Research Atmospheres, 103 (D8), pp. 8763-8778.
Lucht, W. & Schaaf, C., 2000. An algorithm for the retrieval of albedo from space using semi-empirical BRDF models. IEEE Transactions on Geoscience and Remote Sensing, 38 (2), pp. 977-998.
Ross, Iukhan & Ross, J., 1981. The Radiation Regime and Architecture of Plant Stands. Dr. W. Junk Publisher, The Hague, pp.122-123.
Roujean, J., Leroy, L. & Deschanps, P., 1992. A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. Journal of Geophysical Research Atmospheres, 972 (D18), pp. 20455-20468.
Rouse, J., 1973. Monitoring Vegetation Systems in the Great Plains with Erts. Paper presented at the Third ERTS

Symposium, Washington, DC.
Wanner, W., Li, X. & Strahler, A., 1995. On the derivation of kernels for kernel-driven models of bidirectional reflectance. Journal of Geophysical Research Atmospheres, 100 (D10), pp. 21077-21089.
Yang, H., Li, X. & Gao, F., 2002. An algorithm for retrieving surface albedo using the new geometric optical kernel driven BRDF model. Journal of Remote Sensing, 006 (004), pp. 246-251.
Zhang, L., Zhong, T., Liu, H., Zhu, M., Wang, N. & Tong, Q., 2020. China land 1:1 million vegetation index UNVI multidimensional data set (2017). Journal of Remote Sensing, V.24 (11), pp. 4-9.