Cotton Area and Yield Estimation at Zhanhua County of China Using HJ-1 EVI Time Series

Qing-Sheng LIU, Xin-Yang LI, Gao-Huan LIU, Chong HUANG and Yu-Shan GUO

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, LREIS, 11A Datun Road, Chaoyang, Beijing, P.R. of China

Abstract. Cotton is a significant cash crop of China. Timely and accurate cotton area and yield estimation is useful for management decisions related to the cotton procurement and sales. This study is a first research on cotton area and yield estimation based on remote sensing at Zhanhua County which is one of the high-quality cotton production demonstration bases of China. After normalization of Enhanced Vegetation Index (EVI) time series derived from Huanjin 1 A/B satellite (HJ-1 A/B), decision tree classifier was used to identify the cotton, and then K-Means classifier was applied to estimate cotton yield. The results indicated an overall accuracy of 95% for the cotton area estimation and 91% for the cotton yield classification. With further validation, it suggests that this method can be used to timely achieve the cotton area and growth information of this region.

1 Introduction

Cotton is a significant cash crop of China which is mainly planted in Xinjiang, Henan, Shandong, Hebei Province and so on [1]. Timely and accurate cotton area and yield estimation is useful for management decisions related to the cotton procurement, sales, import and export program of a country [2, 3]. Remote sensing imagery can offer a repeated unbiased view of large areas, and has thus been widely used to estimate crop yields. Lots of studies have shown that cotton area and yield can be efficiently estimated by remote sensing [2-13]. National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) data with the higher temporal resolution are usually used to assess early cotton production information and predict the cotton yields at national and regional levels [6, 7]. Landsat Thematic Mapper (Landsat TM), Indian Remote Sensing (IRS), China-Brazil Earth Resource Satellite (CBERS) and Huanjin (HJ-1 A/B) satellite with the higher spatial resolution are often used to estimate cotton yields at regional and field levels [2, 3, 8, 11-13]. Site-specific lint yields were estimated using field soil and multispectral images data [4, 5, 9, 10].

In both cotton area and yield estimation, the use of remote sensing multispectral data from the visible and near infrared range of the electromagnetic spectrum generally involves the use of the vegetation indices. By far, the most commonly used vegetation indices include the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI). The NDVI is chlorophyll sensitive, and the EVI is more responsive to

canopy type, plant physiognomy and canopy architecture. Compared with NDVI, EVI was found to be more linearly correlated with green leaf area index in crop field, less prone to saturation in the regions with the high vegetation fraction, and minimally sensitive to residual aerosol contamination from extensive fires [12, 14].

HJ-1 A/B satellites were launched at Taiyuan city of China on September 6, 2008. HJ-1 A/B were equipped with wide-coverage CCD scanner with the same design principle and observed in parallel to complete scanning and imaging for earth with swath width of 700 km, ground pixel resolution of 30 m with 4 spectrum bands and the revisit period of 2 days after two CCD cameras networking [15]. These high spatial and temporal resolution characteristics of HJ-1 A/B are very useful for crop acreage estimation and yield prediction, so it has been used to estimate winter wheat, rice, sugarcane and cotton area and yields in China [8, 16-19].

However, in China, most of the studies on cotton area and yield estimation were carried out in Xinjiang Province and site-specific applications. Zhanhua County of Shandong Province is one of the high-quality cotton production demonstration bases of China which recently was confirmed as one demonstration base for Cotton Integrated Pest Management by the United Nations and one important production base by Better Cotton Initiative [20, 21]. The aim of this research is to extract the spatially distributed cotton fields and estimate the cotton yields at Zhanhua County using HJ-1 A/B EVI time series.

2 Materials and Methods

2.1 Study Area

Zhanhua County (Fig. 1) lies in the north of Binzhou City of Shandong Province of China from 117° 43' 45" to 118° 21' 52" East longitude and 37° 33' 28" to 38° 9' 59" North latitude and covers a total geographical area of 221,500 ha. The district has the semi-arid East Asia monsoon climate in warm-temperate zone with four distinct seasons. The mean annual temperature is 12.5°C and the average annual precipitation is 543 mm. The main soil type in the region is calcaric fluvisols. Agriculture plays an important role in the economy of the Zhanhua County. Zhanhua County is well known for winter Jujube in China, which covered about 3,300 ha. The most import crop of Zhanhua County is cotton, which covered about 140,000 ha in 2012, while winter wheat, corn and other crops have a markedly smaller area share. In general, the cotton growing season in Zhanhua County extends from mid-April to the end of October.

Figure 1. Zhanhua County-5 May, 2014 EVI image of HJ 1B.

2.2 Data

2.1.1 Ground Survey Data

In this study, for acreage estimation, ground data were collected at the 36 sampling sites in May, 2014, and for yield estimation, the data were collected from questionnaires on 19 farming households in November, 2014 (Fig. 2) and the cotton acreage and yield statistical data were collected on the county level.

2.1.2 HJ-1 A/B Data

In order to classify and identify the cotton acreage, imagery data comprised 24 acquisition dates from 21 March 2014 to 13 November 2014 which was provided in Table 1 below.

The raw digital numbers (DNs) of single date satellite data were converted to at-satellite reflectance using the

parameters from the header file of each image which was described by Satellite Environment Center, Ministry of Environmental Protection²² and the default values for HJ-1 A/B CCD which was described by China Centre for Resources Satellite Data and Application²³. Then, a relative radiometric normalization method using spectrally pseudo-invariant features (PIFs) was used to reduce atmospheric and other unexpected variations in the date of data acquired²⁴. EVI was derived to generate a profile curve of the crop (Fig. 3) using the formula:

$$EVI = G \times \frac{(\rho_{nir} - \rho_{red})}{(\rho_{nir} + C_1 \rho_{red} - C_2 \rho_{blue} + L)} \quad (1)$$

where ρ_{nir} , ρ_{red} and ρ_{blue} are relative radiometric normalization at-satellite reflectance in near-infrared, red and blue bands respectively; $G=2.5$ is a gain factor, $C_1=6$, $C_2=7.5$ are the coefficients of the aerosol resistance term and $L=1$ is the soil adjustment factor.

Table 1. Dates of HJ-1 A/B Used For Cotton Area And Yield Estimation

Satellite Sensor	Path/Row	Date
HJ-1 A CCD2	454/68	21 March 2014
HJ-1 B CCD2	455/68	23 March 2014
HJ-1 B CCD1	453/71	4 April 2014
HJ-1 A CCD1	453/68	6 April 2014
HJ-1 B CCD1	454/68	8 April 2014
HJ-1 A CCD1	455/68	14 April 2014
HJ-1B CCD2	455/68	23 April 2014
HJ-1A CCD2	455/68	29 April 2014
HJ-1B CCD2	456/68	1 May 2014
HJ-1A CCD2	456/68	3 May 2014
HJ-1B CCD2	457/68	5 May 2014
HJ-1B CCD2	454/71	28 May 2014
HJ-1B CCD2	455/70	5 June 2014
HJ-1B CCD2	455/68	14 July 2014
HJ-1A CCD2	456/68	20 July 2014
HJ-1B CCD2	457/68	26 July 2014
HJ-1A CCD2	453/68	20 Aug. 2014
HJ-1B CCD1	1/68	4 Sept. 2014
HJ-1A CCD1	455/68	3 Oct. 2014
HJ-1A CCD1	456/70	7 Oct. 2014
HJ-1A CCD1	457/70	15 Oct. 2014
HJ-1B CCD2	454/68	28 Oct. 2014
HJ-1B CCD2	455/68	9 Nov. 2014
HJ-1B CCD2	456/68	13 Nov. 2014

Figure 2. Land use map in 2010 and the location of ground survey data in 2014.

Figure 3. HJ-1 A/B EVI time series curve of winter wheat, winter jujube and cotton in 2014.

2.1.3 Acreage Classification

The area under cotton at Zhanhua County for the year 2014 was estimated using 24 EVI time series data which almost cover the cotton and other crops/vegetations development stage. County boundary overlaying and decision tree classifier was used for cotton crop identification and acreage estimation. For acreage estimation, the ground survey sites for cotton, winter wheat, winter jujube and other crops/vegetation were overlaid on the RGB false color image (R, G and B stand for EVI image acquired on 29 April, 14 July and 29 October 2014 respectively). Then, decision tree was created to extract cotton crop using machine learning techniques with a training subset (Fig. 4). The decision tree classifier performs multistage classifications by using a series of binary decisions to place pixels into classes. Each decision divides the pixels in a set of images into two classes based on an expression. You can divide each new class into two more classes based on another expression. In the initial decision tree, background, non-vegetation, winter wheat, winter jujube, cotton and other crops/vegetations were able to be separated using more than one band. The band or band combinations used to

separate these classes were determined while performing the classification by visually checking their classification performance on the sites which the training samples were taken. Therefore, the final decision tree was determined while classifying the images. For example, winter wheat was observed to be easily separable using the maximum $EVI > 0.5$ of the images acquired from 6 April to 5 May and the $EVI < 0.35$ of the images acquired on 5 June and 14 July. Winter Jujube was easily separable using the maximum of EVI of all images less than 0.5 and $EVI > 0.35$ of the 28 May image. Cotton can be identified using the maximum $EVI < 0.4$ of the images acquired from 6 April to 5 May, $EVI < 0.35$ of the 28 May image and $EVI > 0.35$ of the 14 July image. After the decision tree classification, farm land mask derived from an existing land use map in 2010 was used to reduce some classification errors resulted from spectral similarity between cotton and natural vegetations. The total classification accuracy was calculated by confusion matrix and amounted to 90% with an accuracy of 95% regarding the classification of cotton compared to the county statistical data on cotton planting acreage.

2.1.4 Yield Estimation

The spectral profile is the temporal variation of spectral vegetation index of the cotton during the growth period. The area under a profile is related to the total photosynthesis over that period. Thus, it is related to above ground biomass and cotton yield. Through analysis on the profile curves of the different cotton fields (Fig. 5), it proved that there is a strong relationship between the profile and yield. K-Means unsupervised classifier was used to classify the cotton fields into low, medium and high yield lands (Fig. 6). K-Means unsupervised classification calculates initial class means evenly distributed in the data space then iteratively clusters the pixels into the nearest class using a minimum distance technique. Each iteration recalculates class means and reclassifies pixels with respect to the new means. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria. This process continues until the number of pixels in each class changes by less than the selected pixel change threshold or the maximum number of iterations is reached. The total classification accuracy was evaluated by the data collected from questionnaires on 19 farming households and the county statistical data, and amounted to 91%.

3 Results and Discussions

3.1 HJ-1 A/B EVI Time Series Images

Table 1 shows the images data used in this research were from two satellites (HJ-1 A and HJ-1 B) with four CCD cameras (HJ-1 A CCD1, CCD2 and HJ-1 B CCD1 and CCD2). Therefore, it is easy to acquire the images of crop growing season and hence improve the cotton acreage and yield estimation accuracy. But it brings about some

problems. Although a relative radiometric normalization was done, EVI time series data from the different sensors could be influenced and then cause cotton acreage and yield estimation errors. In addition, HJ-1 A/B equipped with wide-coverage CCD scanner could bring about the geometric distortion which is difficult to be corrected, it could cause the registration errors between the images and hence increasing the cotton acreage and yield estimation errors.

Figure 4. Cotton areas in Zhanhua County in 2014.

Figure 5. The profile curves of EVI time series of low, medium and high yield cotton fields.

3.2 Cotton Classification

An area 38,393.1 ha of cotton was estimated at Zhanhua County in 2014. Compared with 140,000 ha in 2012, cotton acreage has declined by 72.6 percent. Through analysis on the relative materials, it could be found that this was due to low income from cotton planting. It also indicated that cotton acreage estimation of this paper was reliable. Figure 4 shows the cotton area distribution at Zhanhua County in 2014. Some cotton fields were found around the winter jujube land. The majority of cotton fields were found in the more salinized northern and northeastern areas near the Bohai Sea.

Because the growing season of cotton often overlays those of spring corn, watermelon, sorghum and other crops/natural vegetations such as reed and endive, it led to cotton classification errors. Through an existing farmland map mask, this type of cotton classification errors could be partly reduced. Of course, the more cotton classification accuracy needs the more spatial and temporal resolution remote sensing data and field survey data.

Figure 6. Low, medium and high yield cotton fields in Zhanhua County in 2014.

3.3 Spatial Cotton Yield Estimation

The results from the present research indicated that on the basis of the cotton acreage map, K-Means unsupervised classifier could be used to classify the cotton fields into low, medium and high yield lands using HJ-1 A/B EVI time series data. The high yield cotton fields covered about 8.5% of the total cotton area and sparsely distributed in the southern and southwestern areas, and the low yield cotton fields covered about 7.5% of the total cotton area and sparsely distributed in the northern and northeastern areas, and the medium yield cotton fields covered about 84.0% of the total cotton area.

Because the cotton yield data collected from questionnaires on 19 farming households were rough and the cotton yield statistical data were at county level, the quantitative yield estimation of raw cotton could not be done. In the future, the quantitative yield data of cotton field will be collected including more field-based spatially distributed information such as cotton types, agricultural practices that influence yields.

4 Conclusions

In this paper, HJ-1 A/B EVI time series data was applied to estimate cotton area and yield estimation based on at Zhanhua County. The results showed that decision tree classifier could be used to identify the cotton acreage and

K-Means unsupervised classifier could be used to classify the cotton fields into low, medium and high yield lands using HJ-1 A/B EVI time series data, and there is a strong relationship between the profile and yield of cotton. Thus, with further validation, it suggests that this method can be used to timely achieve the cotton area and yield information at county scale.

Acknowledgment

This research work was jointly financially supported by the National Science-Technology Support Plan Projects (Project No.2013BAD05B03), the Strategic Priority Research Program of CAS (Project No.XDA05050601), International Cooperation in Science and Technology Special Project (Project No.2012DFG22050, 2013DFA91700) and National Natural Science Foundation of China (Project No.41271407, 41023010).

References

1. X. B. Pan, Y. Y. Liu, *Proc. of the 2005 annual conference of the Chinese Meteorological Society*, 3424-3432 (2005). (In Chinese)
2. S. S. Ray, S. S. Pokharna, Ajai, *International Journal of Remote Sensing*, 15, 1085-1090 (1994).
3. S. S. Ray, S. S. Pokharna, Ajai, *International Journal of Remote Sensing*, 20, 2693-2702 (1999).
4. C. Richard, *Photogrammetric Engineering & Remote Sensing*, 66, 1219-1227 (2000).
5. K. R. Reddy, D. L. Zhao, V. G. Kakani, J. J. Read, K. Sailaja, *Proc. SPIE 5153*, doi:10.1117/12.515226 (2004).
6. C. Domenikiotis, M. Spiliotopoulos, E. Tsiros, N. R. Dalezios, *International Journal of Remote Sensing*, 25, 5373-5388 (2004)
7. B. Little, M. Schucking, B. Gartrell, B. Chen, K. Ross, R. Mckellip, *Proc. of IEEE International Conference on Data Mining Workshops*, 426-435 (2008)
8. M. Li, G. X. Zhao, Y. W. Qin, *Proc. of IEEE International Conference on Computer Distributed Control and Intelligent Environmental Monitoring*, 816-823 (2011)
9. J. Iqbal, J. J. Read, F. D. Whisler, *Turkish Journal of Field Crops*, 18, 158-165 (2013)
10. D. L. Zhao, K. R. Reddy, V. G. Kakani, J.J. Read, S. Koti, *Europ. J. Agronomy*, 26, 335-344 (2007)
11. G. R. Ruecker, Z. Shi, M. Mueller, C. Conrad, N. Ibragimov, J. P. A. Lamers, C. Martius, G. Strunz, S. W. Dech, *Proc. of ISPRS XXXVI-8/W48 workshop*, 123-128 (2006)
12. F. Ahmad, K. Shafique, S. R. Ahmad, S. U. Rehman, R. M. Ali Khan, A. Raoof, *Global Journal of Human Social Science (B)*, 13, 21-42 (2013)
13. H. J. Liu, L. H. Meng, X. L. Zhang, U. Susan, D. H. Ning, S. Y. Sun, *Transactions of the Chinese Society of Agricultural Engineering*, 31, 215-220 (2015)
14. Z. Y. Jiang, A. R. Huete, K. Didan, T. Miura, *Remote Sensing of Environment*, 112, 3833-3845 (2008)
15. CBERS, <http://cresda.com/EN/satellite/7117.shtml> (3 November 2015)
16. J. S. Zhang, K. J. Shen, Y. Z. Pan, L. L. Li, D. Hou, *Scientia Agricultura Sinica*, 43, 3306-3315 (2010). (In Chinese with English Abstract)
17. X. C. Wei, X.S. Wang, H. Liu, J. Xu, *Journal of Geo-Information Science*, 14, 382-388 (2012). (In Chinese with English Abstract)
18. M. H. Ding, H. Li, Y. H. Yang, H. Q. Zhang, J.F. Mo, L. He, W. H. Mo, J. H. Wang, *Chinese Journal of Agrometeorology*, 33, 265-270 (2012). (In Chinese with English Abstract)
19. Q. Wang, K. R. Wang, S. K. Li, C. H. Xiao, J. G. Dai, L. F. Fang, *Cotton Science*, 24, 503-510 (2012). (In Chinese with English Abstract)
20. Z. Q. Lu, Q. C. Guo, http://binzhou.dzwww.com/bzhxw/201410/t20141010_11149988.htm (10 October 2014). (In Chinese)
21. X. Li, http://www.bzcm.net/news/2015-06/03/content_1772332.htm (3 June 2015). (In Chinese)
22. Satellite Environment Center, Ministry of Environmental Protection of China, <http://www.secmep.cn/secPortal/portal/column/itemDetails.faces?itemid=297e855148f8b9360148f8ed0f4d0008> (10 October 2014). (In Chinese)
23. CBERS, <http://www.cresda.com/n16/n1115/n1522/n2134/9677.html> (29 December 2009). (In Chinese)
24. Y. L. Du, P. M. Teillet, J. Cihlar, *Remote Sensing of Environment*, 82, 123-134 (2002)