Monitoring the Plant Density of Cotton with Remotely Sensed Data

Junhua Bai [a, b, c], Jing Li [b], Shaokun Li [a*]

[a] Institute of Crop Sciences, Chinese Academy of Agriculture Sciences, Beijing 100081, P. R. China
[b] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University., Beijing 100101, P. R. China
[c] College of Agricultural Sciences of Shihezi University, Shihezi Xinjiang, 832003, P. R. China
∗ Corresponding author at: Institute of Crop Sciences, Chinese Academy of Agriculture Sciences, Beijing 100081, P. R. China. Tel.: +86 10 82108891. E-mail: shaokun0004@sina.com

Abstract PDC (Plant Density of Cotton) was an essential parameter for estimating the cotton yield and developing the zone-management measurements. This paper proposed a new method to retrieve PDC from the satellite remote sensing data. The thirteen fields of Xinjiang Production and Construction Corps (XPCC) (total 630 hm^2) were selected as the study area, where the sowing date, emergence date, and PDC were investigated. Based on the investigation data the linear models to estimate PDC are established using *EVI* and *DEVI* respectively. The results indicated that the difference of seedling size caused by the emergence time decreased the estimation accuracy of PDC. To improve the estimation accuracy the partition functions were established in terms of sowing date. *DEVI* is capable of reducing the influence of soil background significantly and it can bring the monitoring time forward from June 9th to May 24th in this research. The results indicated that the optimal time monitoring PDC would be from squaring to full-flowering of cotton growing period. A demonstration to monitor *PDC* was taken on June 9th in the 148th farm of XPCC. It can be concluded that the emergence time and the non-cotton background were the main factors affecting the monitoring accuracy of *PDC*, and the partition function with the emergence time could improve the estimation accuracy, and *DEVI* could make the monitoring time forward, and the optimal monitoring time was from the squaring stage to the full-flowering stage. This research provides an efficient, rapid and intact way to monitor PDC, and it is significant for operational application at a regional scale.

Key words Cotton · Plant Density · Remote sensing

1 Introduction

PDC (Plant Density of Cotton) is one of key factors affecting cotton yield, and is the most important index for analyzing cotton growth and taking the management measurements. The cotton yield in Xinjiang accounts for about 30% in China and 8% in the international cotton trade. PDC is a key item in all measurements for the cotton farmland management (*W. F. Zhang et al., 2004; X. L. Hou et al., 2006*). The low PDC weakens the applications of cultivation technology for increasing cotton yield.

Agriculture is one of important applications of satellite remote sensing technology, and the numerous researches have been done on the area estimation, yield estimation and growth monitoring (*O. Wendroth et al., 2003; P. C. Doraiswamy et al., 2004; X. M. Xiao et al., 2006*). In cotton monitoring, the method of spectrum matching has been used to extract the cotton pixel in the regional scale(*J. Li et al.,*

2005), and the image texture of geo-statistical has been introduced into monitoring cotton growth(*C. J. Liao, et al., 2006*), and the hyper-spectrum can be used to inverse the biomass and leaf area index quantitatively (*D. L. Zhang et al., 2007; J. H. Bai, et al., 2007*), V. Alchanatis (*2005*) used the hyper-spectrum image to distinguish the weeds from cotton field, and the action of water stress is compared with the root disease by infrared sensor (*N. R. Falkenberg et al., 2007*). And although PDC plays an important role in the farmland practices, few researches was done on PDC monitoring by satellite data.

Normal Difference Vegetation Index (*NDVI*) is one of the most popular parameters used to evaluate the crop growth condition (*J. W. Rouse, 1974; F. Baret, G. Guyot, 1989, 1991*). And some researches show that *NDVI* is susceptible to the influence of the non-vegetation background, and it saturates when the vegetation cover increases to the some extent (*Q. Wang et al., 2005; A. E. K. Douaoui, 2006*). Other indices were constructed to make up for the inadequacy, and the Enhance Vegetation Index (*EVI*) is one of the most popular in them.

At present, the field investigation is almost the only way to obtain *PDC*. The disadvantage of this way is that it can only get the point data, and consumes much time and energy. In this paper, we analyzed the multi-temporal satellite images coupling the field investigation data from the emergence to full-flowering stage of cotton to (1) construct the linear model to estimate PDC, (2) analyze the main factors affecting the monitoring accuracy of *PDC* and perfect the model, (3) obtain the optimal period to monitor *PDC*, (4) demonstrate the PDC monitoring method in the regional scale.

2. Methods

2.1 Study area

The study area is located in the148[th] farm of XPCC, beside the Manas river valley ((45°00′N 86°11′E; 44°46′N 86°23′E). The climate type is the typical continental-arid climate, and the sunshine rate is about 80% during the cotton growing period. The cotton planting proportion accounts for over 90% in the farm.

Generally it seeds from the early April to the middle April, and Squaring at the early of June, flowers at early July. The row space is usually 55cm+20cm+55cm and plants space is usually 9.6cm. At the middle of July, the vegetation coverage fraction rises to over 80%, and the top-ending measurement was taken to control the unlimited growing character of cotton. In *2007*, the sowing date is mostly from April 13th to 19th.
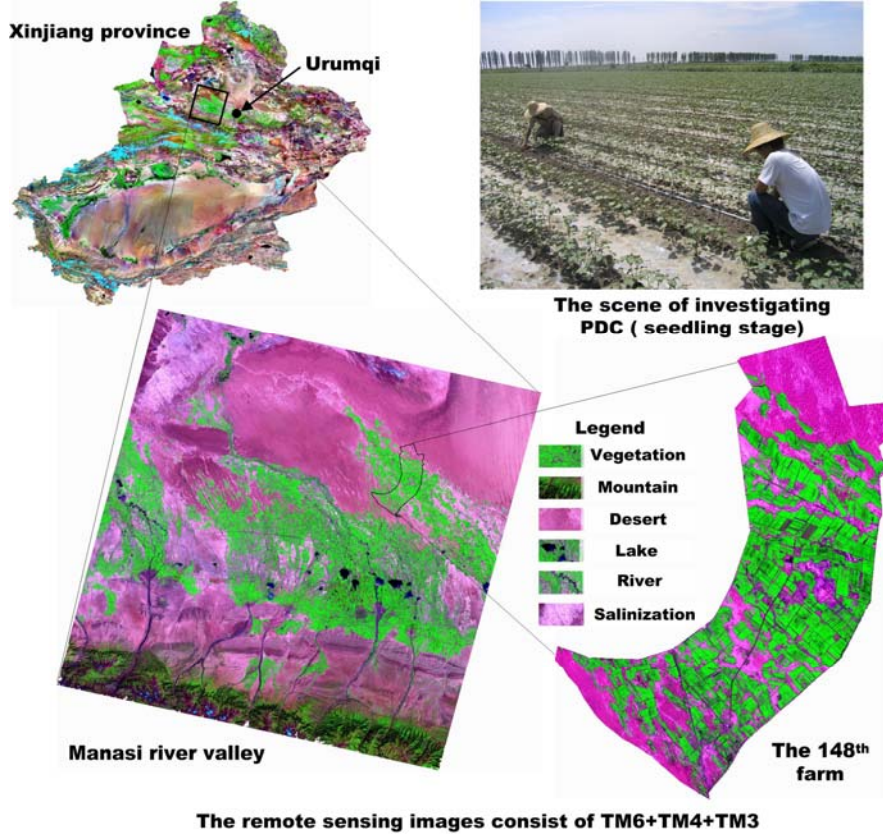
The remote sensing images consist of TM6+TM4+TM3

Fig.1 Location of the experiment

2.2 Investigation and computation for PDC

PDC was investigated from June 5th to 15th, in 2007 and 2008. In order to make the sample data representative, we chose the investigation points far from the field patch boundary to avoid the boundary effect on the crop growing. Three sampling units were investigated in every investigation points. The sampling points were positioned by *GPS*. The area of sampling units is approximately 26.7m[2]. We visited the landowner of each sampling fields to record the seeding and emergence date. The sowing time was divided into three partitions including the early, middle, and late of the sowing date

52 sample points are obtained through investigating thirteen fields (630 hm[2]), including 40 sample points from the middle sowing date, 4 from the early sowing date, and 8 from the late sowing date. The data consisting of three sowing date are employed evenly into the establishing and testing model. PDC could be calculated as:

$$N_u = \frac{n_1 + n_2 + n_3}{3} \times 375 \qquad (1)$$

Where N_u was PDC per hectare($\times 10^5$ plant $\cdot$ hm^{-2}), n_i (i = 1, 2, 3) was the number of plants in the sampling unit.

2.3 Remote sensing images and pre-processing

LANDSAT-5/TM images in April 22th, May 24th, June 9th, June 25th, and July 11th, 2007 were used. The geo-location was done by *ENVI4.2* with the geo-referenced panchromatic waveband image from *LANDSAT-7* as the basic image. The atmospheric correction was done using the AgRSIS software. The atmospheric radiation transferring model uses 6S model and the aerosol optical thickness was inversed by the algorithm of dense dark vegetation. The pixel size was 28.5m×28.5m.

2.4 Vegetation index

$$EVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + C_1 \times \rho_{red} - C_2 \times \rho_{blue} + L} \times G \tag{2}$$

Where ρ_{nir}, ρ_{red}, ρ_{bule} was reflectance in near-infrared (*NIR*), red and blue waveband respectively. C_1 and C_2 were the correction coefficients of atmospheric resistance in red and blue; L is the canopy background brightness correction factor and G is the gain factor. The coefficients adopted in the EVI algorithm are, L=1.0, C_1=6.0, C_2=7.5 and G (gain factor) =2.5 (Huete et al., 1994; Huete et al., 1997).

To further explore the suitable remotely sensed parameters, *DEVI* is the difference of *EVI* between the growth stage and the bared soil stage:

$$DEVI = EVI_i - EVI_e \tag{3}$$

2.5 The evaluation of estimation accuracy

The Relative Extremenum Predication Error (*REPE*) is used to evaluate the estimation accuracy of PDC, REPE is calculated as

$$REPE = \frac{RMSE}{N_U \max - N_U \min} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (N_u p - N_u t)^2}{n-1}}. \tag{5}$$

$N_{u\ max}$, $N_{u\ min}$ is the maximum and minimum value in the testing models, respectively; $N_u p$, $N_u t$ is the estimation and measured value, respectively.

3. Results

3.1 PDC estimation with EVI

PDC estimation based on the data from three emergence stages. As shown in Fig.2, from May 24th to July 11th, EVI rose with the cotton growing. The average EVI for each growing stage is 0.11, 0.27, 0.53 and 0.82. On May 24th, the range of EVI distribution is narrow and it shows insignificant linear relationship between EVI and PDC. However, *EVI* distribute more widely on June 9th and June 25th, and it shows the obvious linear relationship. EVI on June 25th is higher than on June 9th, the correlation degree isn't kept increasing tendency as time goes on, EVI difference on July 11th is less than on June 25th and the linear isn't significant. From the above analysis, we could deduce the emergence time caused the difference of seedling size, which affects the linear between PDC and EVI. Furthermore, the mixed pixels with non-cotton information decrease the relation on May 24th and July 11th, and the relations aren't nearly interrupted on July 9th and July 25th.

The correlation coefficients between PDC and EVI haven't reached significant level on May 24th and July 11th and reached significant level on June 9th and June 25th, and correspondingly R^2 is 0.23 on June 9 and 0.30 on June 25. Therefore, the estimation models on PDC with EVI as below:

$$\text{June } 9^{th}: N_u = 2.61EVI + 1.22 \qquad R^2 = 0.23^{**} \tag{6}$$

$$\text{June } 25^{th}: N_u = 2.41EVI + 0.65 \quad R^2 = 0.30^{**} \tag{7}$$
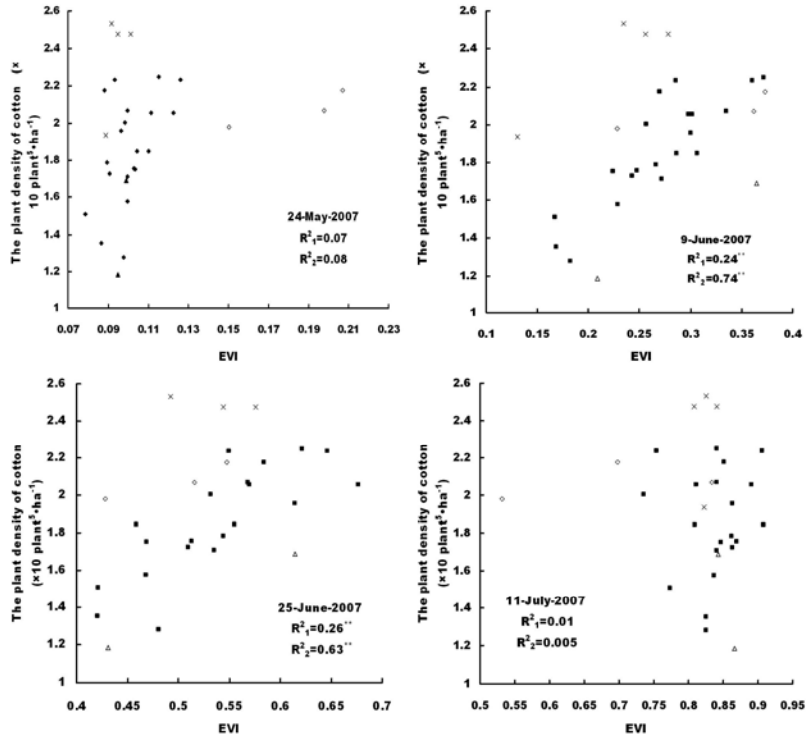
R^2_1, R^2_2 showed the coefficient of determination from three sowing dates and middle sowing date.
△ the late sowing; × the early sowing; ■ the middle sowing; ◇ at the edge of farmland, and the same as below.

Fig.2 Relationship between PDC and EVI

PDC estimation based on the data from middle emergence stage As shown in Fig.2, coefficient of the linear equation between PDC and EVI rise to 0.75 for June 9th and 0.63 for June 25th when the data from middle emergence stage is only used, but unfortunately, the determination coefficient hasn't still reach the significant level on May 24th and July 11th. The estimation models on PDC with EVI as below:

$$\text{June } 9^{th}: \quad N_u = 4.364EVI + 0.686 \qquad R^2 = 0.75^{**} \qquad (8)$$

$$\text{June } 25^{th}: \quad N_u = 3.185EVI + 0.148 \qquad R^2 = 0.63^{**} \qquad (9)$$

3.2 PDC estimation with DEVI

PDC estimation based on the data from three emergence stages EVI on May 24th, June 9th, June 25th and the July 11th, respectively minus EVI on April 22th get DEVI on the corresponding time. It is shown in Fig.3, the determination coefficients of the linear equation between PDC and DEVI are 0.34 reaching the great significant level on May 24th, and on the contrary the linear disappears when EVI estimated PDC on May 24th. In addition, the determination coefficients rise from 0.23 to 0.27 on June

9 and from 0.3 to 0.34 on June 25th. And the linear still disappears on July 11th. The estimation models on PDC with DEVI as below:

$$\text{May } 24^{th}: N_u = 12.13DEVI + 1.13 \qquad R^2 = 0.34^{**} \tag{10}$$

$$\text{June } 9^{th}: N_u = 2.82DEVI + 1.31 \qquad R^2 = 0.27^{**} \tag{11}$$

$$\text{June } 25^{th}: N_u = 2.65DEVI + 0.64 \qquad R^2 = 0.34^{**} \tag{12}$$

PDC estimation based on the data from middle emergence stage From Fig.3, the linear correlations between PDC and DEVI from middle emergence stage are optimized on May 24th, June 9th and June 25th, and the determination coefficients reach the extremely significant level and rise to 0.46, 0.76 and 0.67 distinctly. Unfortunately, the coefficient on July 11th doesn't reach the significant level. The estimation models on PDC with DEVI as below:

$$\text{May } 24^{th}: N_u = 10.95 DEVI + 1.147 \qquad R^2 = 0.46^{**} \tag{13}$$

$$\text{June } 9^{th}: N_u = 4.114 DEVI + 0.980 \qquad R^2 = 0.76^{**} \tag{14}$$

$$\text{June } 25^{th}: N_u = 3.174 DEVI + 0.329 \qquad R^2 = 0.67^{**} \tag{15}$$
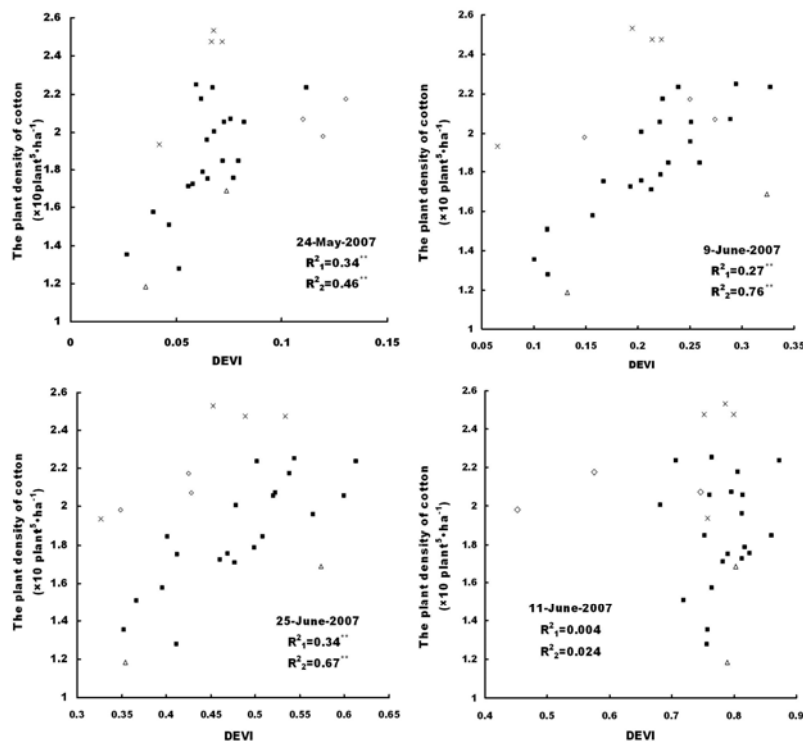

Fig.3 Relationship between PDC and DEVI

3.3 The validation of estimated models and its application

The validation of estimation models As Table 1 shows, the veracity of regression models from (6) to (14) is tested, all determination coefficients between the estimated and measured value are validated, but it is noticeable that the determination coefficients of *III* and *IV* from *DEVI* on May 24th also pass through the significant testing. The RMSE on PDC is 0.301×10^5 plant $\cdot$ hm^{-2} and REPE is 27.1%, which indicates that *DEVI* would bring the monitoring time forward , on the contrary, when EVI is used to estimate PDC on June 9th and June 25th, RMSE is 0.280 and 0.253×10^5 plant $\cdot$ hm^{-2}, respectively, and REPE is 25.7% and 23.2%, respectively. However, the estimation accuracy of models is greatly improved on May 24th, June 9th and June 25th when the data of middle emergence stage is used to estimate PDC. Among them, RMSE and REPE of *I* on June 9th is the lowest value.

Table 1 Testing about the different models of estimating PDC

Time	The estimating methods	The regress equations between measured and estimated values	RMSE	REPE
May 24	III	$y = 0.184x+1.507$ $(R^2 = 0.181^*)$	0.301	27.1
	IV	$y = 0.454x+0.939$ $(R^2 = 0.418^{**})$	0.248	22.7
June 9	I	$y = 0.252x+1.393$ $(R^2 = 0.293^{**})$	0.280	25.7
	II	$y = 0.711x+0.446$ $(R^2 = 0.656^{**})$	0.205	18.7
	III	$y = 0.282x+1.342$ $(R^2 = 0.274^{**})$	0.282	25.4
	IV	$y = 0.727x+0.408$ $(R^2 = 0.660^{**})$	0.207	18.7
June 25	I	$y = 0.322x+1.279$ $(R^2 = 0.466^{**})$	0.253	23.2
	II	$y = 0.559x+0.718$ $(R^2 = 0.664^{**})$	0.208	19.1
	III	$y = 0.307x+1.317$ $(R^2 = 0.431^{**})$	0.259	23.3
	IV	$y = 0.599x+0.636$ $(R^2 = 0.647^{**})$	0.213	19.2

Noting: *I* shows the testing result based on *EVI* from the early, middling and late emergence stage; *II* shows the testing result based on EVI form the middling emergence stage; *III* shows the testing result based on DEVI from the early, middling and late emergence stage; *IV* shows the testing result based on DEVI from the middling emergence stage.

The demonstration for monitoring PDC with LANDSAT-5 It is concluded from analysis above that DEVI can make the PDC monitoring time bring forward to the late May, and the accuracy on June 25th is higher than on June 9th and May 24th. However, I and III on June 9th (Table 1) have the double properties with the earlier time and the higher veracity, and I has the advantage of only using the single-time image, which will enhance the feasibility of estimating PDC. Therefore, I on June 9th can be used as the optimal plan to estimate PDC. Monitoring PDC in the 148th farm by I, the results from Fig.4 indicate that PDC has the strongly spatial correlation and the distributed continuity. The farmland area for cotton in the 148th farm in 2007 is approximate 15811.7 hectare, and 17.1% for $\leq 1.20 \times 10^5$ plant $\cdot$ hm^{-2}, 10.2% for $1.20\sim1.35\times10^5$ plant $\cdot$ hm^{-2}, 13.1% for $1.35\sim1.50\times10^5$ plant $\cdot$ hm^{-2}, 13.6% for $1.50\sim1.65\times10^5$ plant $\cdot$ hm^{-2}, 12.6% for $1.65\sim1.80\times10^5$ plant $\cdot$ hm^{-2}, 10.9% for $1.80\sim1.95\times10^5$ plant $\cdot$ hm^{-2}, 8.7% for $1.95\sim2.10\times10^5$ plant $\cdot$ hm^{-2}, 6.0% for $2.10\sim2.25\times10^5$ plant $\cdot$ hm^{-2}, 3.8% for $2.25\sim2.40\times10$ plant $\cdot$ hm^{-2}, 4.0% for $\geq 40\times10^5$ plant $\cdot$ hm^{-2}. Farmlands with the different PDC can be distinguished (Fig.5).
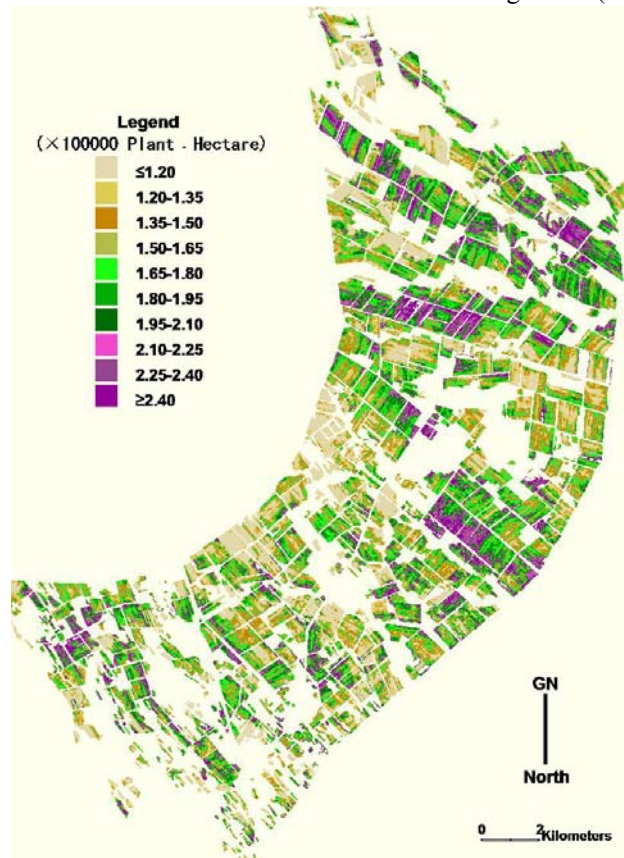

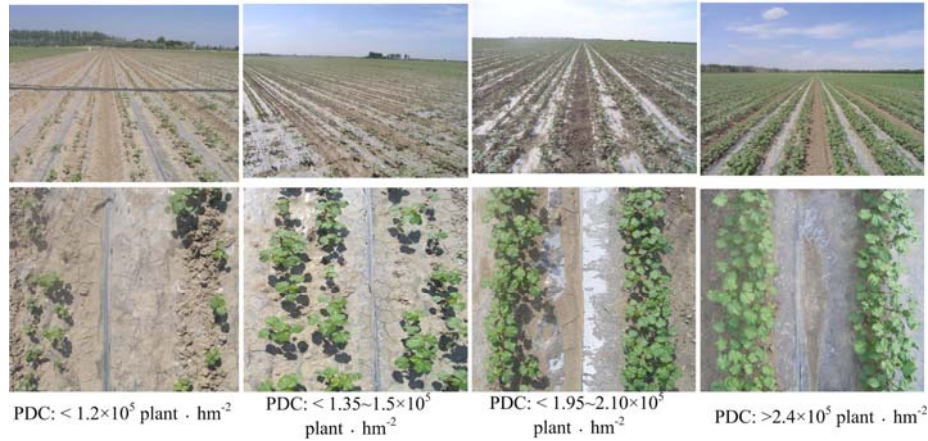Fig.4 Monitoring PDC with remotely sensed data from *LANDSAT-5* in the 148th farm of Xinjiang, china

PDC: < 1.2×10^5 plant · hm^{-2} PDC: < 1.35~1.5×10^5 plant · hm^{-2} PDC: < 1.95~2.10×10^5 plant · hm^{-2} PDC: >2.4×10^5 plant · hm^{-2}

Fig.5 Scene from cotton farmland with the PDC difference

4. Discussion

The above analysis shows that two situations may affect the monitoring accuracy. Firstly, when the coverage fraction of bared soil was high, the reflectance signal of vegetation was weak, the soil properties dominated the difference of pixel reflectance, and it was difficult to retrieve PDC accurately. Secondly, the cotton coverage increased from zero to almost 100 percent when the growth stage developed from the seedling stage to full-flowering period. Under the condition of the high cotton coverage, the contribution of soil to the pixel reflectance decreased, and the saturation of vegetation index happened.

As shown in the results on May 24th EVI is not sensitive to the PDC difference significantly. DEVI could remove the noise of the soil background and highlight the vegetation information. We construct the relationship between PDC and DEVI to monitor PDC, which brings up the PDC monitoring time. Compared with on May 24th, the cotton coverage on June 9th and June 25th was increased remarkably, the proportion of cotton signal rose gradually, and the PDC could be monitored through the green index of cotton canopy population.

The analysis also indicated that it was impossible to estimate PDC effectively in any way when the cotton canopy population in the full-blooming stage had the highest land cover, and especially the cotton coverage rose to over 80%, and the canopy population was fuzzy between pixels because of the saturation phenomenon. Moreover, it was difficult to monitor PDC by remote sensing because of another reason. The canopy population was controlled by PDC before the flowering stage, and yet after the cotton grown into the full-flowering stage, the compensation effect of cotton plants appeared by a series of management measures of top-ending, fertilizer-water and chemical controlling, and so the population with the low PDC might develop into the big one, that meant the population with the low PDC had the higher leaf area index and biomass than the population with the high PDC. As a result, the cotton canopy population was simultaneously controlled by the plant size and PDC.

Therefore, PDC couldn't monitored successfully after the full-flowering stage (at July 11th), which was decided by the defect of remotely sensed technology and the characteristic of cotton growth.

The cotton farmers couldn't seed at the same time because of the limitations of human and mechanism, which could cause the difference of cotton plant size in the early stage of cotton growth, the result added the difficulty of estimating PDC with the high accuracy, the information of sowing date should be known to increase the monitoring accuracy.

5. Conclusions

This paper proposed a method to monitor PDC by satellite remote sensing images. The results showed that satellite images can be used to monitor PDC at the regional scale although the sowing date and the spectrum saturation have an influence on the accuracy of PDC monitoring. The analysis indicates that the optimal PDC monitoring period is late May by DEVI. The optimal growing stage with the highest estimation accuracy to monitor PDC is from the squaring stage to the early-flowering stage, especially the late-squaring stage could be satisfied with the estimation accuracy and time required by the farmland management measurements simultaneously. The method to monitor PDC by satellite images could be a potential way to investigate PDC at the regional scale.

Acknowledgments

References

1. Alchanatis V., Ridel L., Hetzroni A., & Yaroslavsky L. (2005). Weed detection in multi-spectral images of cotton fields. Computers and Electronics in Agriculture, 47(3), 243–260.
2. Baret F., & Guyot G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sensing of Environment, 46(2~3), 161-173.
3. Bai J. H., Li S. K., Wang K. R., Zhang X. J., Xiao C. H., Sui X. Y., 2007. The Response of Canopy Reflectance Spectrum for the Cotton LAI and LAI Inversion. Scientia Agricultura Sinica, 40(1), 63-69.
4. Bai J. H., Li S. K., Wang K. R., Sui X. Y., Chen B., & Wang F. Y. (2007). Estimating Aboveground Fresh Biomass of Different Cotton Canopy Types with Homogeneity Models Based on Hyper Spectrum Parameters. Agricultural Sciences in China, 6(4), 437-445.

5. Ben-Dor E., Levin N., Singer A., Karnieli A., Braun O., & Kidron G. J. (2006). Quantitative mapping of the soil rubification process on sand dunes using an airborne hyperspectral sensor. Geoderma, 131(2~3), 1-21.

6. Boegh E. H., Broge S. N., Hasager C. B., Jensen N. O., Schelde K., & Thomsen A. (2002). Airborne multi-spectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. Remote Sensing of Environment, 81(2~3), 179-193.

7. Doraiswamy P. C., Hatfield J. L., Jackson T. J., Akhmedov B., Prueger J., & Stern A. (2004). Crop condition and yield simulations using Landsat and MODIS. Remote Sensing of Environment, 92(4), 548–559.

8. Douaoui A. E. K., Nicolas H., & Walter C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. Geoderma, 134(1~2), 217– 230.

9. Falkenberg N. R., Piccinni G., Cothren J. T., Leskovar D. I., & Rush C. M. (2007). Remote sensing of biotic and abiotic stress for irrigation management of cotton. Agricultural water management, 87(1), 23-31.

10. Hou X. L., Zhang D., Wang X. J., Li P., & Sheng J. D. (2006). Effect of different nitrogen fertilization on yield and nitrogen using of super-high density cotton system. Cotton Science, 18(5), 273~278. (In Chinese)

11. Huete, A., Justice, C., & Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. Remote Sensing of Environment, 49 (3), 224– 234.

12. Huete, A. R., Liu, H. Q., Batchily, K., & Van Leeuwen, W. J. D. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing of Environment, 59 (3), 440–451.

13. Huang J. F., & Wang X. Z. (1999). Study on cotton phenology and climate in Xinjiang. Journal of Arid Land Resources and Environment, 13(2), 90-95.

14. Li J., Liu Q. H., Liu Q., Liu L. F., Bai J. H., & Li S. K. (2005). Method research of pixel discrimination about CBERS-02 satellite image based on spectral knowledge. Science in China Series E, 35(supplement), 141-155.

15. Liao C. J., Wang C. Y., Li H., & Yang P. R. (2006). Cotton growth monitoring during chemical control stage using geo-statistical image texture: a cases study of Shihezi.Transactions of the CSAE, 22(8), 135-139.

16. Rouse J. W., Haas R. W., Schell J. A., Deering D. W., & Harlan J. C. (1974). Monitoring the vernal advancement and retrogadation (Greenwave effect) of natural vegetation. NASA/GSFC Type III Final Report, Greenbelt, MD, USA.

17. Wang Q., Adiku S., Tenhunen J., & Granier A. (2005). On the relationship of NDVI with leaf area index in a deciduous forest site. Remote Sensing of Environment, 94(2), 244–255.

18. Wendroth O., Reutera H., & Kersebaum K. C. (2003). Predicting yield of barley across a landscape: a state-space modeling approach. Journal of Hydrology, 272(1~4), 250–263.

19. Xiao X. M., Boles S., Frolking S., Li C. S., Babu J. Y., Salas W., & Moore B. (2006). Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sensing of Environment, 100(1), 95–113.

20. Zhang W. F., Wang Z. L., Yu S. L., Li S. K., Fang J., & Tong W. S. (2004). Effects of plant density on canopy photosynthesis, canopy structure and yield formation of high-yield cotton. Acta Phytoecologica Sinica, 28(2), 164~171.

21. Zhao D. L., Reddy K. R., Kakani V. G., Read J. J., & Koti S. (2007). Canopy reflectance in cotton for growth assessment and lint yield prediction. Europ. J. Agronomy, 26(3), 335–344.