Estimating dry matter content for winter wheat using MODIS reflectance data

YONGHONG YI[1], DAWEN YANG[1] & DAOYI CHEN[2]

1 *State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China*
yiyh05@mails.tsinghua.edu.cn

2 *Department of Engineering, University of Liverpool, Liverpool L69 3BX, UK*

Abstract The accumulation process of dry matter content during the crop growing period is a good indicator of crop condition and yield. This paper investigated the application of remote sensing in estimating the dry matter content. Several vegetation indices (VI) calculated from the MODIS 500 m reflectance data were used for regression analysis. A negative correlation was found between all the VIs and the dry matter content after the VIs reached the peak. This correlation during the later growth period may be more valuable for NPP estimation since the dry matter content was then directly related to the crop yield. However, the daily VI time series were found to be very sensitive to atmospheric conditions and angular effects, which may cover the changes on the VI induced by the crop growth to a certain degree. The Savitzky-Golay filter was employed to smooth the daily NDVI time-series as a test. The smoothed NDVI temporal curves were found closer to the 8-day composite ones. The correlation between the dry matter content and NDVI was also improved after smoothing. Both of these demonstrated that this method could partly remove the noise in the MODIS data, specifically the cloud effect, and disclose the changes in NDVI from the crop growth.

Key words MODIS; dry matter; vegetation index; Savitzky-Golay filter

INTRODUCTION

The accumulation process of the dry mass in the crop vegetative and reproductive stage is closely related to the plant net primary productivity (NPP). Remote sensing can provide effective estimation of dry matter content in a rapid and regional way, which facilitates the prediction of regional crop yield. Vegetation indices (VI) are widely applied in the monitoring of crop growth condition and yield prediction because of the simplicity and computational efficiency (Doraiswamy *et al.*, 2004). NDVI, an index combining the absorptive and reflective characteristics of vegetation in the red and near-infrared wavelengths, is sensitive to the vegetation amount (Huete *et al.*, 1999). Changes in NDVI time-series indicate changes in vegetation conditions proportional to the absorption of photosynthetically active radiation (Sellers, 1985). NDVI has been widely applied for the estimation of crop yield and end-of-season dry biomass for its availability. However, NDVI is also sensitive to many other factors, e.g. the background reflectance, the changes in the sun and view angles, and the cloud conditions, which could introduce great disturbance in the NDVI time-series. Additionally, NDVI becomes saturated at dense vegetation cover.

As supplements, several indices were proposed to account for these effects. The soil adjusted vegetation index (SAVI, Huete, 1988) and modified soil adjusted vegetation index (MSAVI, Qi *et al.*, 1994) were developed to consider the contribution of the soil background through aligning the VI isolines with the greenness isolines (usually expressed in terms of LAI). The Enhanced Vegetation Index (EVI, Huete *et al.*, 1999) incorporated both background adjustment and atmospheric resistance concepts for correcting the interactive canopy background and atmospheric influences. Gao (1996) designed a Normalized Difference Water Index (NDWI) using 1240 nm and 860 nm wavelengths, which was proved less perturbed by atmospheric effects than NDVI. The study by Chen *et al.* (2005) proved that NDWI using longer SWIR wavelengths (1640 nm and 2130 nm) showed a lagged saturation in comparison to NDVI for corn and soybeans. In the study of Yi *et al.* (2007), NDWI (1640 nm and 2130 nm) was less sensitive to the atmospheric disturbances and angular effects than NDVI.

Winter wheat is one of the major crops in the northern China plain, so monitoring of its growth and yield has important economic and social impacts. MODIS data shows potential use for retrieval of crop biophysical parameters and for improving accuracy in crop yield assessment, with its merits of moderate spatial resolution (250–500 m) and daily data availability (Doraiswamy *et al.*, 2004). Moreover, MODIS data onboard Terra and Aqua sensors potentially provide twice coverage diurnally, which greatly enhances the ability of monitoring crop condition during the cloudy season.

In this study, four MODIS vegetation indices mentioned above were used to estimate the dry matter content using the linear regression method. The disturbances from the cloud contamination, atmospheric variability and bi-directional effects in VI time-series can greatly affect the monitoring of land cover and terrestrial ecosystems, especially for the daily data (Chen *et al.*, 2004). Therefore, as a test, the Savitzky-Golay filter was employed to reduce these noises in the NDVI time series.

DATA DESCRIPTION

Ground data

Field experiments were conducted around the eddy flux tower (116°3′ 15.3″E, 36°38′ 55.5″N) (see Fig. 1) in the Weishan Irrigation Zone (115°24′–116°30′E, 36°12′–37°00′N). The Weishan Irrigation Zone, with a total area of 36 000 hm^2, is the largest irrigation zone along the downstream of the Yellow River. This area is located in a semi-humid zone with averaged annual precipitation 580 mm. The land cover is mainly constituted of croplands, planted woods, secondary grass and shrubs, and artificial surfaces. The rotation of winter wheat and summer corn is the local primary agriculture practice.

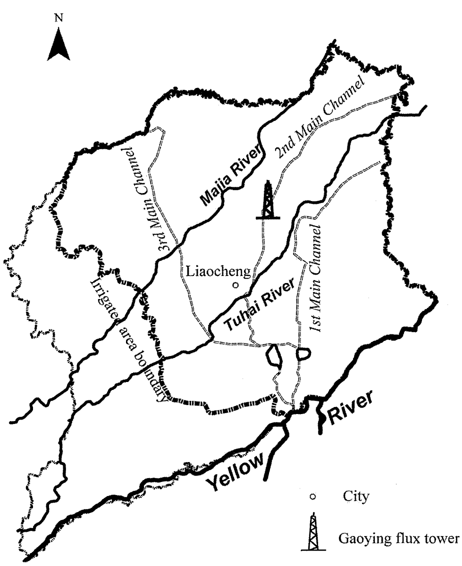

Fig. 1 The Weishan irrigation zone along the downstream of the Yellow River and the experiment site (location of the flux tower).

The ground data were collected in 2006 and 2007. In 2006, the experiment was conducted every other week from 1 April to 3 June. Within a radius of several kilometres around the flux tower, the croplands were almost homogeneously covered by the same wheat crop. Three 1-m^2 blocks were selected randomly around the flux tower, but with nearly uniform row distance. The total number of tillers of wheat in each block was counted, and then the samples taken from the outside of the plot were dried in an oven for 8 h at 70°C to obtain the dry weight. The dry matter of the three samples was averaged and finally the variables in the 1-m^2 area were derived by multiplying the ratio of the tillers in the 1-m^2 area to the sample tillers (in unit of kg m^{-2}). In 2007, the experiment was augmented to six plots located in an area of 1.5 × 1 km^2 around the flux tower. The measurement was conducted weekly from 17 March to 6 June. The same method was taken to measure the dry matter content. The temporal curves of the measured dry matter content in 2006 and 2007 are shown in Fig. 2.

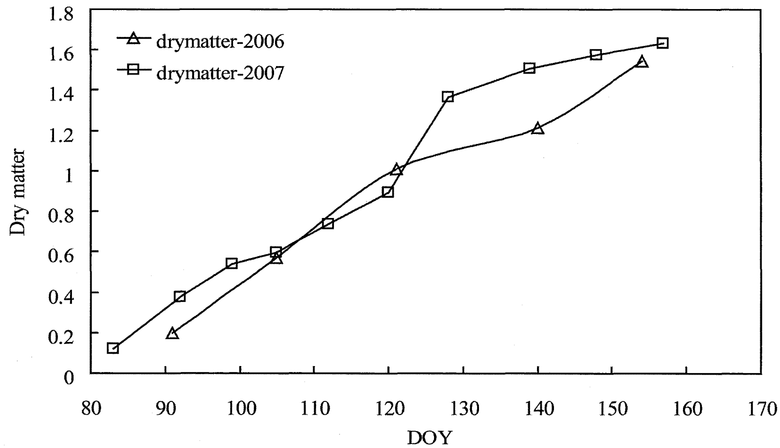

Fig. 2 The temporal curves of dry matter content during wheat growing period in 2006 and 2007.

MODIS data

The MODIS daily and 8-day composite 500 m surface reflectance products were downloaded and used for analysis. The daily products provide an estimation of the surface reflectance for MODIS channels 1 to 7, as it would be measured at the land surface if there were no atmosphere. The 8-day product is a composite over an 8-day period of the daily products. The criteria for qualified pixel screening in the composite are filtering out those pixels with cloudiness, cloud shadows, and low solar zenith angle. When several values pass these quality criteria, the minimum of channel 3 (blue) values are considered, which are mostly contributed by the atmospheric scattering process. The composite process can eliminate the cloud contamination to a certain extent. The data were first re-projected from Sinusoidal Grid (SIN) projection to UTM projection using the MODIS Reprojection Tool (MRT). MODIS Land Data Operational Product Evaluation (LDOPE) software was used to check the cloud status and data quality using the associated quality control and state flags.

METHODOLOGY

Vegetation Index

NDVI has the following form with a normalized transform of the NIR to red reflectance ratio:

$$\mathrm{NDVI}=(\rho_{\mathrm{NIR}}-\rho_{\mathrm{Red}})/(\rho_{\mathrm{NIR}}+\rho_{\mathrm{Red}}) \tag{1}$$

where ρ_{NIR} and ρ_{Red} are reflectances in NIR and red wavelengths, respectively.

SAVI is proposed by Huete (1988) to use vegetation isolines equations to minimize soil background effects. MSAVI is a modified version of SAVI (Qi *et al.*, 1994) and represented by:

$$\mathrm{MSAVI}=\left[2\rho_{\mathrm{NIR}}+1-\sqrt{(2\rho_{\mathrm{NIR}}+1)^2-8(\rho_{\mathrm{NIR}}-\rho_{\mathrm{Red}})}\right]/2 \tag{2}$$

EVI is designed for MODIS, with improved sensitivity into high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influence (Huete *et al.*, 1999).

$$\mathrm{EVI} = 2.5 \cdot (\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}}) / (\rho_{\mathrm{NIR}} + 6 \cdot \rho_{\mathrm{Red}} - 7.5 \cdot \rho_{\mathrm{Blue}} + 1) \quad (3)$$

where ρ_{Blue} is blue band reflectance.

NDWI was proposed by Gao (1996) and proved to be less sensitive to atmospheric effects and shows a lag in saturation compared with NDVI. NDWI is calculated as:

$$\mathrm{NDWI} = (\rho_{\mathrm{NIR}} - \rho_{\mathrm{SWIR}}) / (\rho_{\mathrm{NIR}} + \rho_{\mathrm{SWIR}}) \quad (4)$$

where ρ_{SWIR} is reflectance of SWIR wavelengths (e.g. MODIS bands 6, 7).

The Savitzky-Golay filter

The quality of the daily MODIS data was greatly affected by clouds and aerosol contamination. In this study the image was excluded if the cloud and mixed area in the study region exceeded 30%, or the blue band reflectance was above 0.07, to remove undetected cloud effects. However, even after cloud masking, the sub-pixel cloud contamination could not be totally removed, which induced a great variation in the daily reflectance data. This variation may cover the changes in the reflectance due to the crop growth. In this study, the Savitzky-Golay filter was used to reduce the noise caused by cloud contamination and atmospheric variability in the NDVI time-series.

The Savitzky-Golay filter was a simplified least-squares-fit convolution for smoothing and computing derivatives of a set of consecutive values or a spectrum (Savitzky & Golay, 1964). The convolution can be understood as a weighted moving average filter with weighting given as a polynomial of a certain degree. This filter was designed to consecutive data at a fixed and uniform interval. The general equation of this filter can be given as follows:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+i}}{N} \quad (5)$$

where Y is the original value, Y^* is the resultant value, C_i is the coefficient for the *i*th value of the filter (smoothing window), and N is the number of convoluting integers equal to the smoothing window size (2 m + 1).

This method assumed that the effects of the cloud contamination and atmospheric variability usually depressed the NDVI values and the sudden drops in the NDVI time-series were regarded as noise and were removed. In this study, the cloudy data was first recognized using the above defined cloud mask, and replaced by a linearly interpolated value using the adjacent points. However, if there were over 7 days with cloudy data, the missing data would not be interpolated. Then the Savitzky-Golay filter was used to smooth the NDVI variation and obtain the long-term changing trend. Comparing with the raw data, a new NDVI time-series was generated by replacing the noisy NDVI points with the corresponding points in the smoothed time-series. The new time-series would more closely approach the upper NDVI envelope through an iteration process and best fit the NDVI variations during the full vegetation season. The details of the processing can be found in the study of Chen *et al.* (2004).

RESULTS AND DISCUSSION

Figure 3 presented the daily VI time-series from both Terra and Aqua sensors in 2006 and 2007, respectively, which was only processed using cloud mask. It can be seen that the temporal curves were rather variable, especially for the three indices (NDVI, MSAVI, and EVI). The variations in these indices were found to be closely related to the changes in the blue band reflectance, which was mainly contributed by atmospheric scattering. Especially the NDVI values in both years had

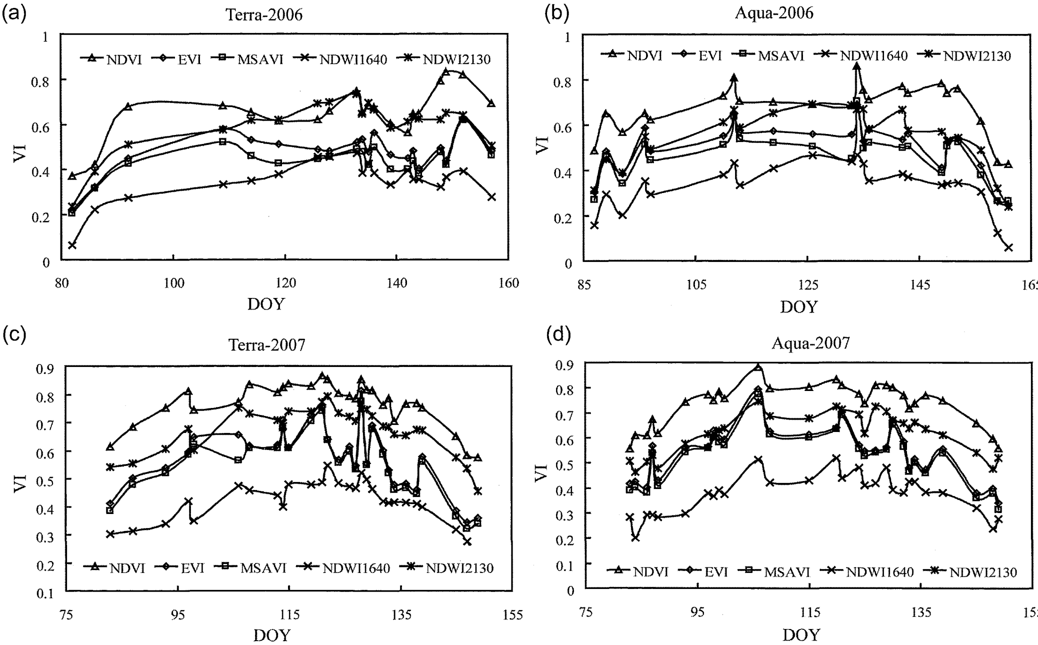

Fig. 3 The temporal curves of daily Terra and Aqua vegetation indices in 2006 and 2007.

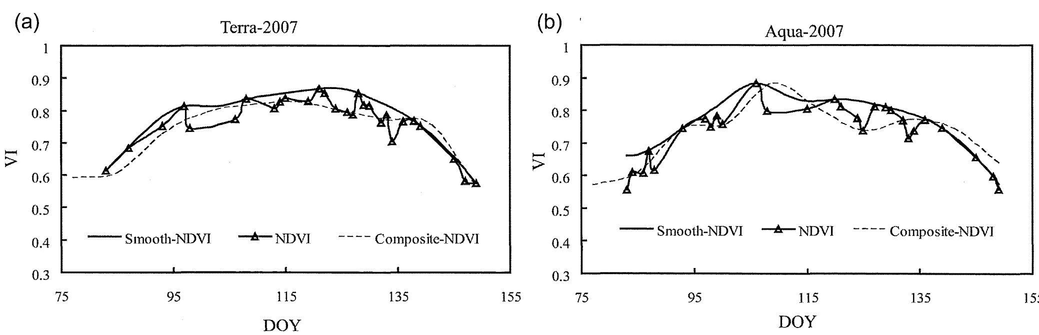

Fig. 4 The comparison of the original, smoothed, and 8-day composite NDVI temporal curves in 2007.

obvious correlations with the blue reflectance. Therefore, it was speculated that these variations were mainly caused by changes in atmospheric conditions. Apart from this effect, the indices were also influenced by the changes in the sun-target-view geometry.

Figure 4 shows the smoothed NDVI time-series in 2007. It was not applied in the 2006 daily NDVI time series because there were too few clear dates from DOY 90 to 110 for both sensors. The filter lost its effect under this condition. The smoothed time series were compared with the 8-day composite ones, in which the influence from the atmospheric variability is partly removed through the composite process. It can be seen from the figure that the smoothed time-series were close to the composite ones, except around DOY 125 for Aqua. However, the valley around this date may be caused by the cloud disturbance and the composite process can not totally remove this effect. Although the smoothed curve was closer to the composite, the smoothed values were generally larger than the composite ones. In the filter method, it was assumed that the cloud always depressed the NDVI values, and the other effects on NDVI were neglected. Therefore, the smoothed NDVI time series were generally matched with the largest values in the time-series.

However, the composite values may choose the values with minimum blue reflectance among the cloud-free dates, not always the date with the largest NDVI value, which could partly remove the angular effects in the reflectance. In consideration of this, the modified Walthall method was applied to the daily NDVI values before filtering to normalize the sun and view angles. The four empirical coefficients in the model were regressed using the modelled reflectance from a 1-D SAIL (Scattering by Arbitrarily Inclined Leaves) model. However, it was found that the BRDF (Bi-directional reflectance) correction did not have much effect on the results. One reason may be that the SAIL model did not provide much reasonable modelling of the canopy reflectance. The other reason may be that the chosen pixels for analysis were not pure pixels, which was limited by the surface heterogeneity in the study area.

Figure 5 shows the scatter plots between several selected indices and dry matter content. For 2006, NDVI and $NDWI_{2130}$ are presented, and for 2007 only NDVI before and after smoothing are shown. In both years, the scatter plots between VI and the dry mass were similar. In the early period, the VI and dry mass showed a positive correlation and after VI arrived at the peak, they showed a negative positive correlation. The different correlation forms may be related to the characteristics of the wheat growth. Before the heading stage (around DOY 120), the growth of the wheat was characterized by the growth in the vegetative components. After the wheat came into the heading stage, the crop grew with notable accumulation of dry mass, but the vegetative parts stopped rapid growth and slowly withered.

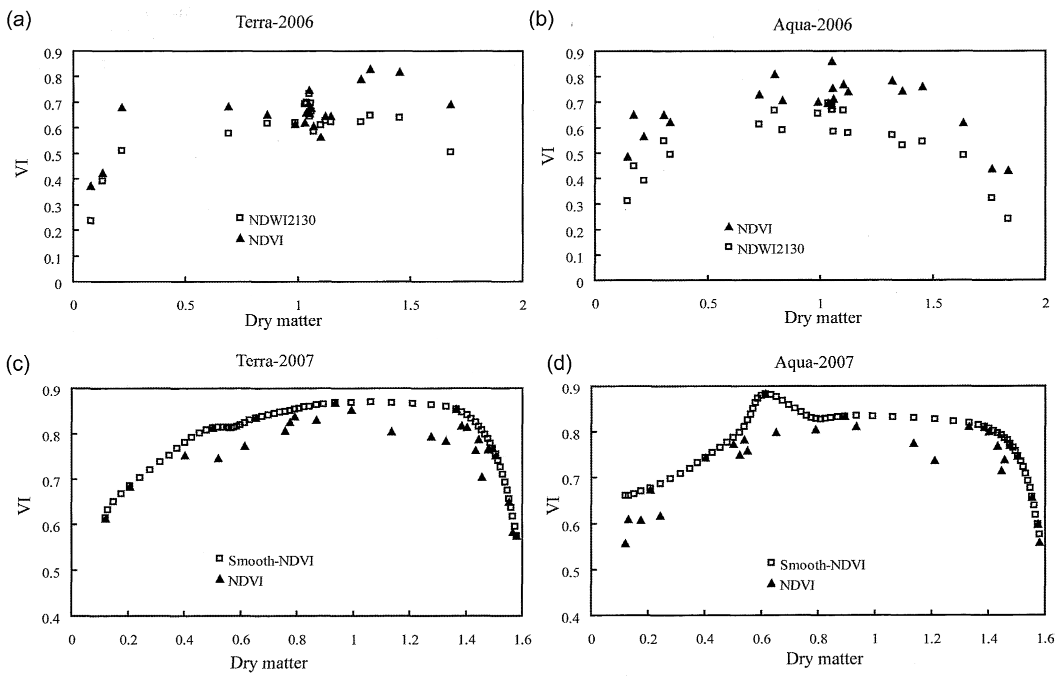

Fig. 5 The relationship between VI and dry matter content in 2006 and 2007, respectively.

Tables 1 and 2 list the determination coefficients (R^2) between all the indices and the dry matter content at the later growth stages in 2006 and 2007, respectively. In 2006, the period was from DOY 119 to 157 for the Terra sensor and from DOY 119 to 161 for Aqua. There were 15 and 14 images during the period for Terra and Aqua, respectively. In 2007, the period was from DOY 115 to 149 for both sensors. There were 19 and 16 images during this period for Terra and Aqua,

Table 1 The summary of R^2 values between dry matter content and daily vegetation indices during the later wheat growth period of 2006.

	NDVI	EVI	MSAVI	$NDWI_{1640}$	$NDWI_{2130}$
Terra	0.29	*0.01*	*0.13*	0.38	**0.41**
Aqua	**0.58**	**0.76**	**0.61**	**0.82**	**0.88**

Bold figures mean that the correlation is significant with $p < 0.01$, and italics mean no obvious correlation with $p > 0.05$. There were 15 samples for Terra from DOY 119 to 157 and 14 samples for Aqua C4 from DOY 119 to 161.

Table 2 The summary of R^2 values between dry matter content and daily vegetation indices during the later wheat growth period of 2007.

	NDVI	NDVI _smooth	EVI	MSAVI	$NDWI_{1640}$	$NDWI_{2130}$
Terra	**0.48**	**0.51**	**0.40**	**0.44**	**0.43**	**0.48**
Aqua	**0.42**	**0.51**	**0.51**	**0.52**	**0.46**	**0.42**

Bold figures mean that the correlation is significant with $p < 0.01$. There were 19 samples for Terra and 16 samples for Aqua both from DOY 115 to 149 in 2007.

respectively. The difference in the chosen dates between the two years was due to the difference in the crop phenology, which can be observed from the dry matter accumulation curve in Fig. 2.

From Table 1, it can be seen that Terra and Aqua VIs had very different performance in the correlation with the dry matter content in 2006. For Terra, only $NDWI_{2130}$ showed significant negative correlation with the dry matter content. EVI and MSAVI showed no correlation with the dry matter content at all. However, for Aqua, significant correlation was found between all the indices and the dry matter content. The two NDWI indices performed best in both sensors. This may be because NDWI, employing longer wavelengths, was less sensitive to the disturbance from the cloud and atmospheric conditions (see Fig. 3).

In 2007, the indices performed similarly for both sensors. A significant correlation was found between all the indices and the dry matter content. The correlation between the smoothed NDVI and the dry matter content did not improve much compared with the original NDVI and the other indices. From Fig. 5, it can be found that the observed dry matter content did not increase much, whilst the VI values kept decreasing after DOY 139. The change in the dry matter content at the last period in 2007 was much different from that in 2006. In 2006, the dry matter content had a sharp increase after DOY 140. If only considering the dates before DOY 139, the correlation between the smoothed NDVI and dry matter content was greatly improved. The R^2 values were increased from 0.51 to 0.85 for Terra and from 0.51 to 0.93 for Aqua. As we knew, the wheat did not experience serious water stress in 2007, and the difference in the dry matter accumulation at the later stages should be investigated. One of the possible error sources may be from the measurements. In 2007, the measurement was averaged from six plots and made once a week, and the error from the multiple measurements was only defined arbitrarily by the growth process of each sample.

CONCLUSIONS

The MODIS derived vegetation indices were used to estimate the dry matter content during the winter wheat growing period in 2006 and 2007.

The daily VI time-series, especially the three indices (NDVI, EVI and MSAVI) employing shorter wavelengths, showed great deviation, which may be caused by the cloud and atmospheric variability and the changes in the sun-target-view geometry. Comparatively, NDWI was more resistant to these disturbances and may potentially be more reliable for monitoring crop growth conditions, especially when the uncertainty in the data set was large, e.g. the Collection 4 Terra data in 2006.

After smoothing, the daily NDVI time-series was closer to the 8-day composite ones, which showed the success of the algorithm in removing the cloud contamination and other disturbances. However, the smoothed NDVI values were typically larger than the composite values. This was because the filter method ignored the angular effects and only considered the depressing effect of the cloud on NDVI values. The sun and view angles in the NDVI time-series should be normalized before the filtering, especially for sensors with a large Field of View (FOV) like MODIS.

The VIs showed a positive correlation with the dry matter content before the VIs arrived at the peak, and after that, a negative correlation was found between them. Except for the Terra indices in 2006, all the indices showed significant correlation with dry matter content during the later growth period, which was characterized by the notable accumulation of the dry mass in the crop. The correlation at the later growth stages may be much more valuable for NPP estimation since the dry matter content is then directly related to the crop yield.

Acknowledgements This research was supported by the National Natural Science Foundation of China (50679029).

REFERENCES

Chen, D., Huang, J. & Jackson, T. J. (2005) Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands. *Remote Sens. Environ.* **98**(2–3), 225–236.

Chen, J., Jonsson, P., Tamura, M., Gu, Z., Matsushita, B. & Eklundn, L. (2004) A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ.* **91**, 332–344.

Doraiswamy, P., Hatfield, J., Jackson, T., Akhmedov, B., Prueger, J. & Stern, A. (2004) Crop condition and yield simulation using Landsat and MODIS. *Remote Sens. Environ.* **92**(4), 548–559.

Gao, B. (1996) NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **58**(3), 257–266.

Huete, A. (1988) A soil adjusted vegetation index (SAVI). *Remote Sens. Environ.* **25**, 295–309.

Huete, A., Justice, C. & van Leeuwen, W. (1999) MODIS vegetation index (MOD13). Algorithm Theoretical Basis Document (ATBD). http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H. & Sorooshian, S. (1994) A modified soil adjusted vegetation index. *Remote Sens. Environ.* **48**(2), 119–126.

Savitzky, A. & Golay, M. J. E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chem.* **36**, 1627–1639.

Sellers, P. J. (1985) Canopy reflectance, photosynthesis and transpiration. *Int. J. Remote Sens.* **6**, 1335–1372.

Yi, Y., Yang, D., Chen, D. & Huang, J. (2007) Retrieving crop physiological parameters and assessing water deficiency using MODIS data during winter wheat growing period. *Canadian J. Remote Sens.* **33**(3), 189–202.