City University of New York (CUNY)
CUNY Academic Works

International Conference on Hydroinformatics

2014

Empirical Regression Model Using Ndvi, Meteorological Factors For Estimation Of Wheat Yield In Yunnan, China

Jing Huang

Qiang Dai

Huimin Wang

Dawei Han

How does access to this work benefit you? Let us know!

More information about this work at: https://academicworks.cuny.edu/cc_conf_hic/5
Discover additional works at: https://academicworks.cuny.edu

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

EMPIRICAL REGRESSION MODEL USING RETRIEVED NDVI, METEOROLOGICAL FACTORS FOR ESTIMATION OF WHEAT YIELD IN YUNNAN, CHINA

JING HUANG (1, 3), QIANG DAI (2), HUIMIN WANG (1, 3), DAWEI HAN (2)
(1): Institute of Management Science, Hohai University, Nanjing, 211100, China
(2): Department of Civil Engineering, University of Bristol, Bristol, BS8 1TR, UK
(3): State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, 211100, China

Crop yield estimation is of great importance to food security. NDVI, as an effective crop monitoring tool, is extensively used in crop yield estimation. However there are few studies conducted in the regions where mixed crops are grown. In this study, a statistical approach for crop area identification is proposed and applied to wheat in Jianshui County in the Nanpan River Basin, Yunnan Province of China. Based on the correlation analysis between MODIS NDVI data and crop yield, the planting areas are identified, as well as the best periods for a reliable estimation. Regression models are presented to predict the crop yield with the retrieved NDVI from the corresponding crop planting-areas. Besides, the crop yield is also strongly influenced by meteorological factors, such as precipitation, temperature and potential evapotranspiration data. Therefore, new regression model by adding those factors is presented and compared with the former one. This study has proposed a simple and convenient method on crop yield estimation using meteorological factors and NDVI data in small regions where crop type is unknown exactly.

INTRODUCTION

Crop production plays a vital role in human society food security and economic development. In the past years, the fluctuation of crop yield caused a great loss in economy, and even lead to food crisis of the whole country in China.

Since the extensively application of remote sensing data on agriculture and crop production [1-4], especially based on the predictive empirical model, it is possible to estimate crop yield efficiently and quantitatively. The Normalized Difference Vegetation Index (NDVI) data could be used to estimate the vegetation health and monitor changes in vegetation. The NDVI temporal profile rises with the growth of crops typically, reaches peak level during the productive stage, and declines around the harvest [5]. The NDVI data derived from NOAA-Advanced Very High Resolution Radiometer (AVHRR) has been used to forecast crop yields worldwide since 1980s. For instance, Mkhabela *et al.* [6] developed regression models to forecast maize yields with average NDVI data in Swaziland; Balaghi *et al.* [7] set up empirical regression models with NDVI data and wheat yields in Morocco. Similar studies have been conducted for various crop types in many other regions [8-10].

Although previous studies found useful statistical relationships between final crop yields and NDVI data around the world, few studies have been conducted in small regions (as small as county level). Moreover, most of the studies were based on specified experimental fields or a region with mono-type and known crops, and no studies have been carried out with unknown crop area exactly. Furthermore, meteorological factors such as precipitation, temperature and potential evapotranspiration data also affect the final crop yield. There is sufficient evidence that the models perform more accurately by adding some other variables such as soil, rainfall, and temperature than that only using NDVI [7, 9, 11]. This study aimed to estimate crop yield in small regions practically using MODIS NDVI data and meteorological factors. In particular, the first efforts are devoted to analyzing the correlation between time-series NDVI data and crop yield. The research is then directed to the identification of the geographical distribution of crops and the best period for a reliable crop yield forecast using NDVI. Next, the regression models based on the NDVI data of best periods and selected planting areas are developed to estimate the crop yield. Finally, through regression analysis, new model is proposed by adding meteorological factors, and then compared with the former ones. The ultimate objective is to put forward a method to forecast crop yield in a small region with available NDVI data and meteorological factors in order to assess regional agricultural risk in the future.

MATRIALS AND METHODS

Study area

In this study, Jianshui county in Nanpan River basin in Yunnan Province, China, with a total area of 3940 square kilometres, is selected as the study area (Figure 1.). Jianshui extends northward from 23°13’N to 24°10’N latitudes and westward from 102°37’E to 103°55’E longitudes. The climate is subtropical monsoon with an average annual temperature of 18-20℃ and an average annual precipitation of 800-1000 mm.

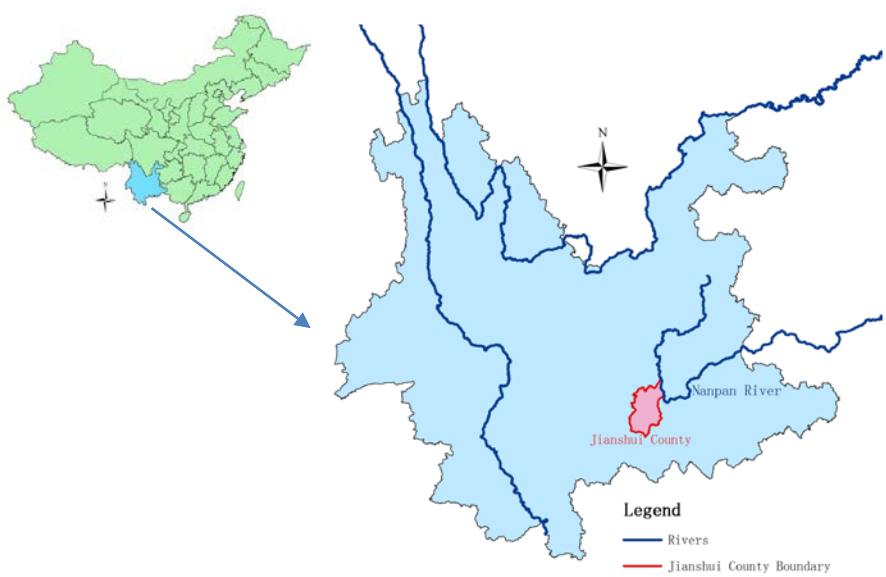

Figure 1. Jianshui County in Nanpan River Basin, Yunnan Province, China

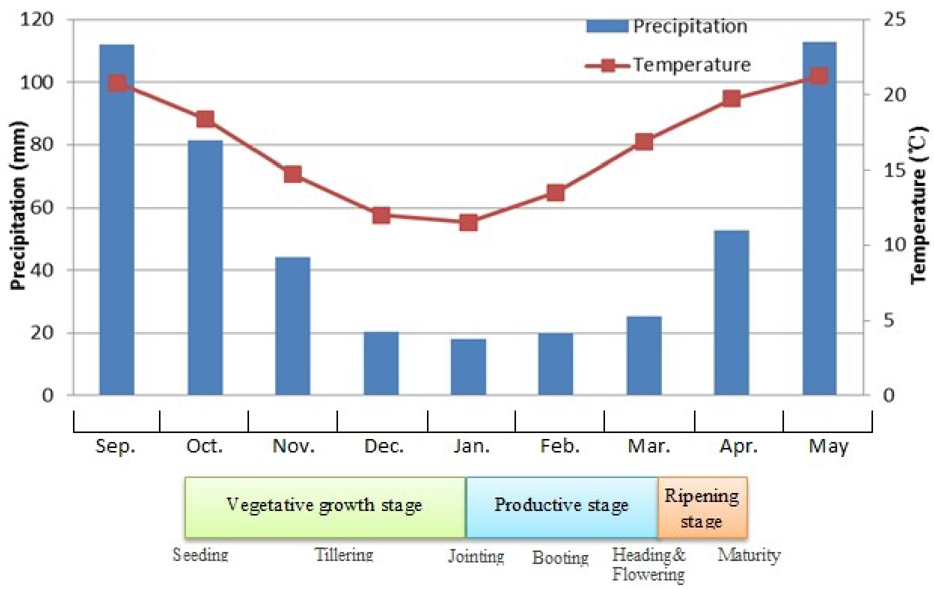

Figure 2. Temperature and precipitation during the wheat growth period in Jianshui County

Data

In Jianshui, winter wheat is the main crop, which grows from middle October to the end of April next year (Figure 2.). Wheat grain production and planted areas for the years 2000-2009 are obtained from Yunnan Statistical Yearbook. Figure 2. shows the wheat growth path in Jianshui, relating to temperature and precipitaion (average for the years 1991-2010).

Processed global Moderate Resolution Imaging Spectroradiometer (MODIS) 16-day NDVI data in 1-km resolution for the years 2000 to 2009 were acquired from Earth Observing System Data and Information System (EOSDIS) of National Aeronautics and Space Administration (NASA). The basic geographic information data of China are downloaded from the National Fundamental Geographic Information System, which is used to extract the existing NDVI images for Jianshui with the GIS (Geographic Information System) tool. Besides, the Chinese land cover map is obtained for the crop land identification from the Cold and Arid Regions Science Data Centre at Lanzhou [12]. The map is used to eliminate the influence of non-agriculture crops on NDVI. Therefore, only those areas of agricultural usage are extracted for further analysis. The meteorological data were obtained from the gridded Climatic Research Unit (CRU) TS (time-series) datasets produced by the Climatic Research Unit (CRU) at the University of East Anglia.[13]. Since they are calculated on high-resolution (0.5*0.5 degree) grids around the world, the subset data of Jianshui are retrieved through latitude and longitude.

Methods

Statistical analysis has been done separately for every crop in each study site. Correlation analysis is performed using the 16-day NDVI values and crop yield data. Particularly, NDVI data of each 16-day period for every pixel (1km*1km) are used to set up a relationship with the yield of winter wheat from year 2000 to 2009.

After the correlation analysis is applied, there are three steps should be implemented based on the matrixes. Firstly, a threshold value of the correlation coefficient (r=0.6) is set. The efforts are then turned to classify the planting areas of winter wheat. Based on the r matrix, the pixel p in which r value over the threshold value in any crop growth period is recognized as the planting area. Such pixel belongs to P. And the corresponding period t when the highest r got belongs to T. That is recognized as the best period for crop yield estimation in every pixel. Lastly, the mean 16-day NDVI data of all extracted planting areas in every period are calculated in each site for correlation analysis as below:

$$NDVI_t(i) = \frac{\sum_{j=1}^{m} NDVI\left(j\right)_t (i)}{m} \quad (1)$$

where m is the number of elements in P, which indicates the number of planting pixels; t is the period of the NDVI data; i is the number of the study year.

The correlation analysis is re-conducted using the mean NDVI during the best period and wheat yield. Regression analyses, both linear and nonlinear models, are performed to predict the crop yield. The mean NDVI data during the best periods are taken as the independent variable and the crop yield as the dependant variable.

In the following step, the ordinary least squares (OLS) regression tecniques are applied. The wheat yields (dependent Y-variable) will be regressed on NDVI, precipitation, temperature and potential evapotranspiration (independent X-variables). The precipitation, mean air temperature and potential evapotranspiration (PET) sum for all possible groups of consecutive months from October to March. For example, t(O) and t(O-M), respectively, indicate the mean temperatures in October and in the period from October to March in the second year. Therefore, the subset of explanatory variables is checked through regression fitting by selecting any group in each meteorological parameter. Next, the selected regression equations were tested in more depth using leave-one-out (LOO) cross-validation.

RESULTS AND DISCUSSION

Table 1 shows the results of correlation coefficient, including correlation coefficient (r) between the mean NDVI and wheat yields, p-Values, the best periods for making reliable yield forecast and the plant-area. The yield of winter wheat is strongly correlated with the NDVI that counted all the plant area of 203 pixels. Since winter wheat grows through the winter, the crop yields are affected not only by the conditions during planting periods in the same year, but also by that in fomer year. Therefore, the NDVI in the last 16-days of the former year are averaged with the mean NDVI in the first month in the same year for winter wheat in Jianshui. In general, the period approximately coincides with the periods of jointing and booting of wheat in this study. The results agreed with the previous studies which had found the high correlations between crop yield and NDVI data during the jointing, booting and filling periods [4, 6, 14-16]. It is probably due to the fact that these three stages are most crucial moment to crop yield, in which any water stress would lead to the reduction of the yield.

In practical situation, there are many types of crops growing together in several small areas. However, the crop yield is only related with the NDVI data of the corresponding crop areas. Therefore, efforts have been focused on the crop identification which ultimately determines the extraction of the NDVI for crop yield forecasting. From Figure 3a, winter wheat only occupies approximately 24% of the crop land. From Figures 3b, the crop distribution could be discriminated more clearly. The green pixels are croplands, which are extracted from the study by Ran *et al.* [17]. The triangles signify the land planting wheat. The croplands are mainly concentrated in the middle of Jianshui which is a fertile river valley area. The crops are rarely grown in the northern part, primarily due to the mountainous terrain. In consideration of the correlation coefficients between the NDVI data and the wheat yield in these pixels, it is interesting to point out that the higher r values are found in the areas of high elevations.

Table 1. Correlation coefficient, p-Values, and the best periods for crop yield forecasting

Crop Type	*r*	*P*	Best Periods	Plant-area
Winter Wheat	0.7238	0.0275	19/12-03/01(23) & 01/01-01/02 (1-2)	203 pixels

In the previous studies, the crop type was fixed for the study site, and the correlation was established between the NDVI throughout the entire area and the crop yield. For example, Ren *et al.* [18] used county level NDVI to relate winter wheat yield; Mkhabela et al. [6] forecasted corn yield with the cultivated NDVI in the four agro-ecological regions of Swaziland. Such NDVI data, however, could not reflect crop yield perfectly because the NDVI data represents the growth condition of all the crops planted in this area. A study by Wang *et al.* [19] found a high correlation in an experimental area in which single crop type is grown. Moreover, Maselli and Rembold [20] described a similar statistical method on crop land identification based on computation per-pixel inter-annual correlations between NDVI and crop yield. However, the study just distinguished the agriculture vegetation from non-agriculture vegetation, and still addressed one generic crop type as agriculture vegetation in whole country. Therefore, the extraction of NDVI for a specific crop would make a significant improvement in crop yield estimation model.

The curve estimation analysis has been conducted, and several regression models which passed the significance test of both the equations and regression coefficients are compared (Table 2.). It seems that S model (y=exp(a+b/x)) fitting winter wheat yield well, which explained 53% of the variability of winter wheat. Mkhabela et al. [4] found a power function to be best-fit for barley, canola, field peas, and spring wheat on the Canadian Prairies. Meanwhile, in a study by Jiang *et al.* [21] found the cubic polynomial regression was the best one from linear, cubic polynomial, and exponential regression models. Holzapfel *et al.* [22] and Begue *et al.* [23] both found linear and exponential equations were suitable for relating the NDVI data to crop yield. Ma *et al.* [24] considered a power function suited the relationship between NDVI and soybean grain yield appropriately.

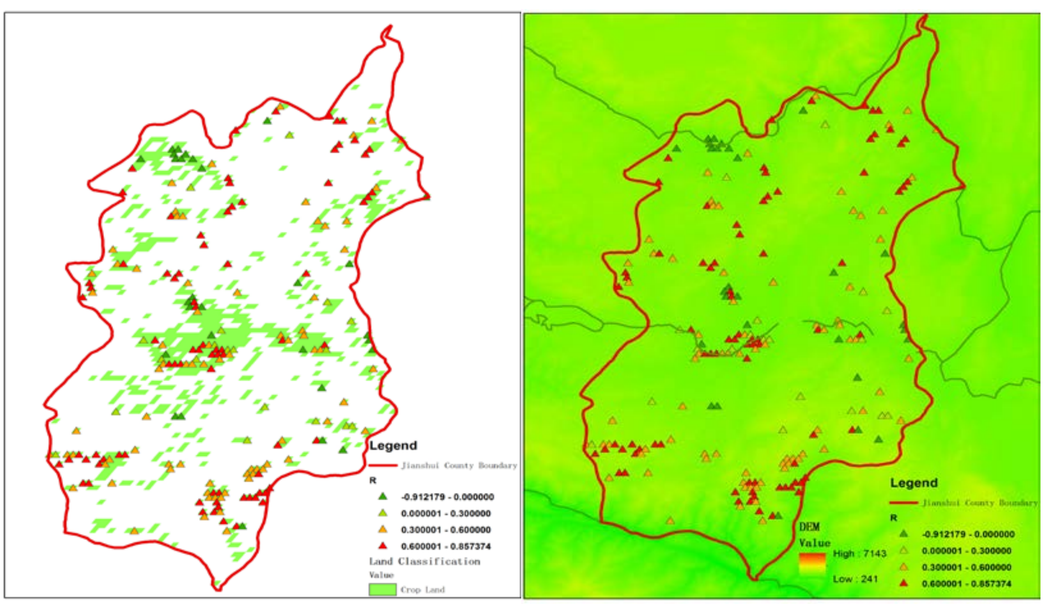

Figure 3. (a) Crop land distribution and (b) the correlation coefficient in every wheat area

Table 3 shows the regression model which retain four significant *X*-predictors. Apparently, NDVI is by far the most important explanatory variable, while some meteorological factors are of great importance. By adding precipitation, temperature and PET variables, the new regression model performs better than any other former ones. This agrees with the research by Balaghi et al. [7], Rasmussen [9], Groten [11]. Moreover, the predicted model with mean precipitation and temperature during the critical period performance better than that with single month value. The probable reason is that the weather situation during several growth stages determines the changes of crop yield.

Although these predictor variables explain the bulk of the variability of winter wheat, there are also some remaining unexplained variances not covered by this model. For instance, diseases, agricultural management, soils, etc.

Table 2. Curve models for winter wheat yield estimation using retrieved NDVI in Jianshui

Model	Equation	R-square	Adjusted R-square	Root Mean Square Error (RMSE)
Linear	$Y = 55.34 + 4162 \times NDVI$	0.5242	0.4562	149
Inverse	$Y = 3912 - \frac{885.7}{NDVI}$	0.5371	0.4709	146.9
Power	$Y = 4168 \times NDVI^{0.9643}$	0.5244	0.4565	148.9
S	$Y = e^{(8.56 - \frac{0.4462}{NDVI})}$	0.5314	0.4645	147.8
Growth	$Y = e^{6.631 + 2.069 \times NDVI}$	0.5162	0.4471	150.2
Exponential	$Y = 785.1 \times e^{2.069 \times NDVI}$	0.5162	0.4471	150.2

Table 3. Regression models for winter wheat yield estimation using retrieved NDVI, precipitation, temperature, PET in Jianshui

NDVI, Precipitation, Temperature, PET	R^2	R_p^2
$Y = 491.8 + 1772.7 \times \text{NDVI} + 6.2 \times \text{p(M)} - 193.1 \times \text{t(N)} + 50.2 \times \text{PET(O)}$	0.9853	0.9163
$Y = -2576.4 + 9734.3 \times \text{NDVI} - 9.4 \times \text{p(O-F)} + 115.8 \times \text{t(D-M)} - 42.4 \times \text{PET(J)}$	0.9909	0.9222

Note: The period of precipitation, temperature and PET are indicated by combining start and end month, with single characters (O, N, D, J, F, M) for the months October-March. For instance: p(O-F) = mean precipitation over period from October to February. R_p^2 is the dertermined coefficient for LOO cross validation.

CONCLUSION

This study presented a simple and easy implement scheme to estimate crop yield in mixed crop planting regions using MODIS NDVI data and meteorological data based on geospatial and regression analysis. With the assumption that the strong correlation between NDVI data and crop yield indicates a high probability of the crop type, the identification method of the crop planting areas is presented through the correlation analysis in each pixel. The extraction of NDVI data on a specific crop is conducted for the establishment of crop estimation models. Such models accounted for 44% to 47% of the variance in winter wheat yield. Besides, the best periods for yield estimation model are presented, which approximately coincide with the jointing, booting and filling stages of the crops. Futhermore, some meteorological factors (precipitation, temperature and potential evapotranspiration) during certain periods are selected by using regression tecniques. The new regression model explained most of the winter wheat yield variability compared with any former ones only using NDVI. In conclusion, the method in this study to estimate the crop yield with the mean NDVI data extracted from the pixels of the crop grown and meteorological data is practical and reasonable and it is also a simple method to study the crop yield in a small region where several crops are planted together.

In this study, the crop identification is conducted firstly in order to improve the correlation between NDVI data and crop yield, which is different to the studies conducted by other researchers in which the crops were mono-type and known by default. In addition, the significant meteorological factors are introduced in new regression model which shows great improvement compared with former models. If a long record of data is available in the future, the models could to be validated and updated. Although the models proposed in this study show some promising results in crop yield estimation, a more generalised model structure with additional inputs from various site physical characteristics may be needed for different crops in different locations. If a wide range of applications of this proposed method could be carried out in various locations, more factors affecting the correlations between crop yield and NDVI, such as soil moisture, solar radiation, etc., could be analysed as model input variables. In brief, with such a generalised model, the crop yield could be estimated in the regions without the historical crop yield records.

ACKNOWLEDGEMENTS

This work is supported by the National Social Science Foundation of China (Grant No.12&ZD214), the Research Fund for the Doctoral Program of Higher Education of China (Grant No.20120094110018) and Research Innovation Program for College Graduates of Jiangsu Province (Grant No.CXLX12_0259).

REFERENCES

[1] Doraiswamy P. C., Moulin S., Cook P. W., and Stern A., "Crop yield assessment from remote sensing," *Photogrammetric Engineering and Remote Sensing*, vol. 69, (2003), pp 665-674.

[2] Chaudhari K. N., Tripathy R. and Patel N. K., "Spatial wheat yield prediction using crop simulation model, GIS, remote sensing and ground observed data," *Journal of Agrometeorology*, vol. 12, (2010), pp 174-180.

[3] Boken V. K. and Shaykewich C. F., "Improving an operational wheat yield model using phenological phase-based Normalized Difference Vegetation Index," *International Journal of Remote Sensing*, vol. 23, (2002), pp 4155-4168.

[4] Mkhabela M. S., Bullock P., Raj S., Wang S., and Yang Y., "Crop yield forecasting on the Canadian Prairies using MODIS NDVI data," *Agricultural and Forest Meteorology*, vol. 151, (2011), pp 385-393.

[5] Soudani K., Hmimina G., Delpierre N., Pontailler J. Y., Aubinet M., Bonal D., Caquet B., de Grandcourt A., Burban B., Flechard C., Guyon D., Granier A., Gross P., Heinesh B., Longdoz B., Loustau D., Moureaux C., Ourcival J. M., Rambal S., Saint Andre L., and Dufrene E., "Ground-based Network of NDVI measurements for tracking temporal dynamics of canopy structure and vegetation phenology in different biomes," *Remote Sensing of Environment*, vol. 123, (2012), pp 234-245.

[6] Mkhabela M. S., Mkhabelab M. S. and Mashininic N. N., "Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR," *Agricultural and Forest Meteorology*, vol. 129, (2005), pp 1-9.

[7] Balaghi R., Tychon B., Eerens H., and Jlibene M., "Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco," *International Journal of Applied Earth Observation and Geoinformation,* vol. 10, (2008), pp 438-452.

[8] Benedetti R. and Rossini P., "On the Use of NDVI Profiles as a Tool for Agricultural Statistics: The Case Study of Wheat Yield Estimate and Forest in Emilia Romagma," *Remote Sensing of Environment,* vol. 45, (1993), pp 311-326.

[9] Rasmussen M. S., "Operational yield forecast using AVHRR NDVI data: Reduction of environmental and inter-annual variability," *International Journal of Remote Sensing,* vol. 18, (1997), pp 1059-1077.

[10] Lewis J. E., Rowland J. and Nadeau A., "Estimating maize production in Kenya using NDVI: some statistical considerations," *International Journal of Remote Sensing,* vol. 19, (1998), pp 2609-2617.

[11] Groten S. M. E., "NDVI—crop monitoring and early yield assessment of Burkina Faso," *International Journal of Remote Sensing,* vol. 14, (1993), pp 1495-1515.

[12] Ran Y. and Li X. Multi-source Integrated Chinese Land Cover Map [Online]. Available: http://dx.doi.org/10.3972/westdc.010.2013.db

[13] University of East Anglia Climatic Research Unit (CRU). [Jones P. and Harris I.]. CRU Time Series (TS) high resolution gridded datasets, [Internet]. NCAS British Atmospheric Data Centre, 2008, 2014.01. Available from http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_0c08abfc-f2d5-11e2-a948-00163e251233

[14] Labus M. P., Nielsen G. A., Lawrence R. L., Engel R., and Long D. S., "Wheat yield estimates using multi-temporal NDVI satellite imagery," *International Journal of Remote Sensing,* vol. 23, (2002), pp 4169-4180.

[15] Wang J., Rich P. M., Price K. P., and Kettle W. D., "Relations between NDVI, Grassland Production, and Crop Yield in the Central Great Plains," *Geocarto International,* vol. 20, (2005), pp 5-11.

[16] Salazar L., Kogan F. and Roytman L., "Use of remote sensing data for estimation of winter wheat yield in the United States," *International Journal of Remote Sensing,* vol. 28, (2007), pp 3795-3811.

[17] Ran Y., Li X., Lu L., and Li Z., "Large-scale land cover mapping with the integration of multi-source information based on the Dempster–Shafer theory," *International Journal of Geographical Information Science,* vol. 26, (2012), pp 169-191.

[18] Ren J. Q., Chen Z. X., Zhou Q. B., and Tang H. J., "Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China," *International Journal of Applied Earth Observation and Geoinformation,* vol. 10, (2008), pp 403-413.

[19] Wang L., Bai Y. L., Lu Y. l., Wang H., and Yang L. P., "NDVI Analysis and Yield Estimation in Winter Wheat Based on Green-Seeker," *Acta Agronomica Sinica,* vol. 38, (2012), pp 747-753.

[20] Maselli F. and Rembold F., "Analysis of GAC NDVI Data for Cropland Identification and Yield Forecasting in Mediterranean African Countries," *Photogrammetric Engineering and Remote Sensing,* vol. 67, (2001), pp 593-602.

[21] Jiang D., Wang N. B., Yang X. H., and Wang J. H., "Study on the Interaction between NDVI Profile and the Growing Status of Crops," *Chinese Geographical Science,* vol. 13, (2003), pp 62-65.

[22] Holzapfel C. B., Lafond G. P., Brandt S. A., Bullock P. R., Irvine R. B., Morrison M. J., May W. E., and James D. C., "Estimating canola (Brassica napus L.) yield potential using an active optical sensor," *Canadian Journal of Plant Science,* vol. 89, (2009), pp 1149-1160.

[23] Begue A., Lebourgeois V., Bappel E., Todoroff P., Pellegrino A., Baillarin F., and Siegmund B., "Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI," *International Journal of Remote Sensing,* vol. 31, (2010), pp 5391-5407.

[24] Ma B. L., Dwyer L. M., Costa C., Cober E. R., and Morrison M. J., "Early prediction of soybean yield from canopy reflectance measurements," *Agronomy Journal,* vol. 93, (2001), pp 1227-1234.