The Fifth International Conference on Agro-Geoinformatics 2016
18-20 July, Tianjin,China

Regression based Corn Yield Assessment using MODIS Based Daily NDVI in Iowa State

Ranjay Shrestha*[1], Liping Di*[1], Eugene G. Yu[1], Lingjun Kang[1], Lin Li[1], Md. Shahinoor Rahman[1], Meixia Deng[1], Zhengwei Yang[2]

[1]Center for Spatial Information Science and Systems (CSISS)
George Mason University
Fairfax, VA, USA
{ldi, gyu, mdeng}@gmu.edu, { rshresth, mrahma25, llin2, lkang3}@masonlive.gmu.edu

[2]United States Department of Agriculture, National Agricultural Statistics Service (USDA, NASS)
Washington D.C.
Zhengwei.Yang@nass.usda.gov

***Abstract*—Traditional method of visiting field and surveying farmers to estimate crop yield has been considered inefficient and impractical especially in cases when fields are not easily accessible. Remote sensing techniques, therefore, has been utilize to overcome these obstacles with good success. Normalize Difference Vegetation Index (NDVI) based models are considered to be most effective and utilized technique in crop yield assessment and can provide up to field level assessment. This research utilize MODIS based 250m daily NDVI to estimate corn yield in 4 Agricultural Statistics Districts (ASD) in Iowa state. Corn is considered the primary crop in the US accounting for 90% of the total feed grains, hence utilized to be study in this research. Linear regression model was derived between NDVI curve and corn yield using all counties within the 4 ASD between years 2000 to 2014. The regression model showed statistically significant relation between NDVI curve and corn yield with coefficient of deterministic (R-square) over 0.80 in all 4 ASD. Similarly validating the model using new 2015 yield, the average predictability error was between 5 to 7 percent.**

Keywords—MODIS NDVI; Liner Regression; Corn Yield; Remote Sensing

I. INTRODUCTION

Field visit and surveying local farmers has been one of the traditional methods to estimate crop yield [1]–[3]. These methods are only practical within a small study area and when fields are easily accessible, however could be an extremely time consuming and difficult process otherwise. Remote sensing techniques has been implied by many scientists to overcome these obstacles with good success [1]–[5]. Satellite images and their products can be utilized in crop yield assessment efficiently because of their: wide area coverage, accurate geocoding, frequent revisit, rapid data distribution, relatively low data cost, and strong crop/land discrimination [3].

One of the methods to derive remote sensing products for crop yield assessment is band ratioing. Band ratioing is a process of calculating ratio of the spectral responses between different bands within a single image. As infrared band of the spectrum are much more sensitive to the green plants compare to visible portion of the spectrum, ratio between these bands can be used to delineate healthy crops from its adjacent surroundings [3]–[6]. Vegetation Indices (VI) are the primary example of spectral band ratioing to derive indicators of vegetation condition. These products are primarily derived from global covering high temporal resolution images such as Moderate Resolution Imaging Spectroradiometer (MODIS) and Advance Very High Resolution Radiometer (AVHRR). VI based crop loss assessment methods are much more robust compare to the image classification based methods [7], [8]. By transforming the daily products into weekly/bi-weekly composite products, the VI are able to reduce data abnormalities such as cloud cover effect. Once the abnormalities are removed, very small or even negative value will indicate water/land surface whereas vegetation cover will show positive values (0.3-0.8) [9]. Normalized Difference Vegetation Index (NDVI) considered to be the most utilized vegetation index in crop yield (especially corn) and NDVI studies [2], [10]–[12]. Hence this study will use NDVI product for corn yield assessment. However unlike in previous researches, this study will utilize MODIS based 250 m daily NDVI product to provide up to field level assessment. Furthermore, by using the daily NDVI, detection of the growing season timeline will be more accurate compare to the weekly/bi-weekly NDVI products.

Corn will be the primary crop used for this study. Based on USDA's report, corn account for more than 90% of the total feed grain in the US and rank first in the world with 40% of the global corn production [13]. It is also consider multipurpose crop as besides being used in diverse food products, it is also used in industrial alcohol and fuel ethanol. Hence this research

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:48:15 UTC from IEEE Xplore. Restrictions apply.

will focus on corn to understand how well it can be assess using NDVI product.

So the aim of the research will be to utilize daily MODIS NDVI product to detect its relation with corn yield using linear regression and examine whether these NDVI indicators can actually be utilize to assess the corn yield. For the case study, 4 Agricultural Statistics Districts (ASD) in Iowa State was selected. The next chapter will go over these study areas as well as the datasets used in the study. Chapter 3 will highlight the method carried out in the study followed by Result/Discussion and Conclusion in chapter 4 and chapter 5 respectively.

II. Study Area/Data

A. Study Area

For the study area, 4 ASD and all counties it contains within the Iowa State was selected. The ASD code for these area are: 1920, 1940, 1970, and 1990. Each ASD contains between 10-14 counties. Fig 1 below shows the study area.

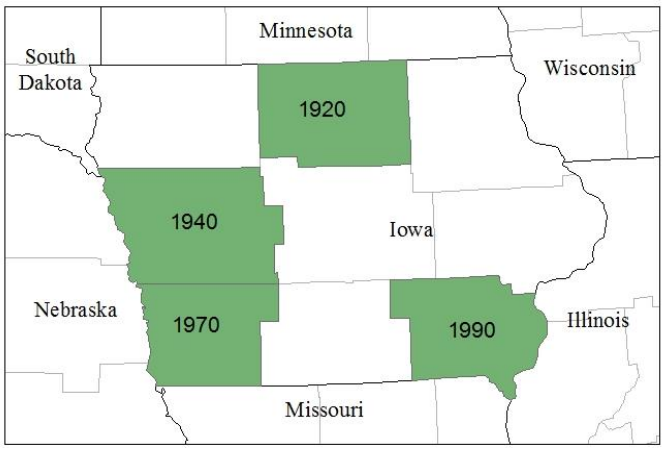


Fig. 1. Study Area- ASD 1920, 1940, 1970, and 1990- IA State

B. Data

The primary dataset used for the modeling of the study was 250m MODIS daily NDVI from year 2000 to 2014. For the crop type identification, 50 m LandSat Crop Data Layer (CDL) product for year 2000 to 2014 was used. Both the dataset along with necessary geopolitical boundaries were obtained through the online web portal designed and maintained by Center for Spatial Information Science and System (CSISS), George Mason University (GMU). Additionally, county level corn yield data for the 4 ASD in IA State for year 2000-2014 were obtained through USDA, NASS quick stat data repository.

III. Method

Below are the method carried out in the research to process/utilize the daily NDVI and Landsat CDL data and use it with corn yield data to build the regression model for yield assessment.

A. Corn Pure Pixel Selection

The Landsat CDL based corn pixels and MODIS based NDVI pixels are not in the same spatial scale. With Landsat 50m and MODIS 250m resolution, the former dataset was scaled as ratio to the latter dataset. Using the county boundaries, both CDL and NDVI datasets were masked out for all counties within the study area of the 4 ASD in IA State. Depending on the total area of the corn CDL pixel contained within a NDVI pixel, the fraction percentage of corn was calculated and assigned to the NDVI pixel. All NDVI pixels which has 99% or more corn cover area within a single county was selected as “Pure Corn Pixels”.

B. Normal NDVI

Normal NDVI is a single cure representative of annual yield for corn within a single county. Once the pure corn pixels are identified based on previous section, next step is to derive normal NDVI curve for that county. Utilizing daily MODIS NDVI layer, first all the NDVI value for all identified “Pure Corn Pixels” are obtained. Next, the normal NDVI curve is achieved by taking median value for all “Pure Corn Pixel” NDVI across the same day. For example, all NDVI value of day 1 for all “Pure Corn Pixel” were sorted and the median value was taken to represent the day 1 NDVI for the Normal NDVI curve. Similar approach was performed for all daily values for all 14 years (2000 – 2014).

C. NDVI Noise Reduction

Before the Daily NDVI values can be utilize for further analysis, the noise in the NDVI time-series needed to be removed. These noises occurs due to various reasons such as cloud cover, shadows etc. Hence it is extremely important to remove these imparities form the time-series to achieve more accurate results. Fig 2 shows the Daily NDVI curve with noise in Fremont County, IA for year 2014.

To remove the noise from the NDVI time-series, two level filtering: Best Index Slope Extraction (BISE) with Savitzky-Golay (SavGol) filter was applied. BISE is a threshold based method which uses a moving window approach to filter the noise in the time series. In the first level filtering, BISE with moving window of 40 days sliding period was used as threshold. For the second level of filtering SavGol filtering was utilize to smooth the time-series. SavGol utilizes the polynomial function to smooth the time-series and in this research, SavGol with second degree polynomial function was used. Additionally, with SavGol smoothing, corn growing season (Start day and End day) was also dynamically defined. Fig 3 shows the cleaned Normal NDVI after applying two level of noise reduction filtering with BISE and SavGol in Fremont County, IA, 2014.

This study was supported by a grant from NASA Applied Science Program *(Grant # NNX14AP91G, PI: Prof. Liping Di*)

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:48:15 UTC from IEEE Xplore. Restrictions apply.

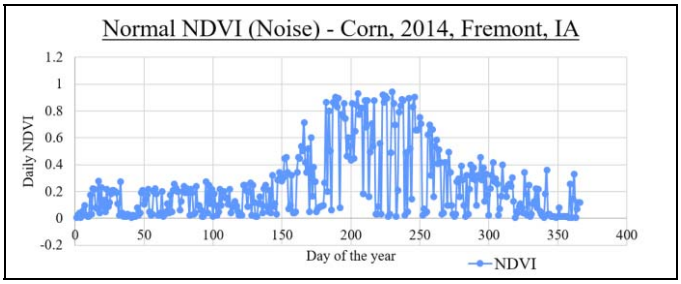


Fig. 2. Normal NDVI showing noise impact, Corn, 2014, Freemont, IA

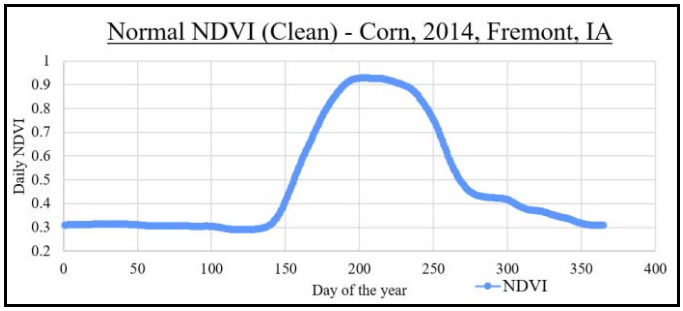


Fig. 3. Normal NDVI without noise, Corn, 2014, Freemont, IA

D. *Area Under the Curve*

Before running the regression model between the normal NDVI curve and the annual yield, the entire year NDVI curve needed to quantify into single value. For this purpose to represent the annual yield with the NDVI curve, Area Under the Curve (AUC) within the growing season was calculated. As NDVI represents the greenness (health) of the vegetation, the total area of the curve within the growing season of the normal NDVI curve assume to represent annual yield for that county. Additionally, the annual yield was also normalize into bushel per acre (BU/Acre). These two variables, AUC and yield ratio were used to develop the regression model. Fig 4 shows the derivation of area under the curve for the normal NDVI curve.

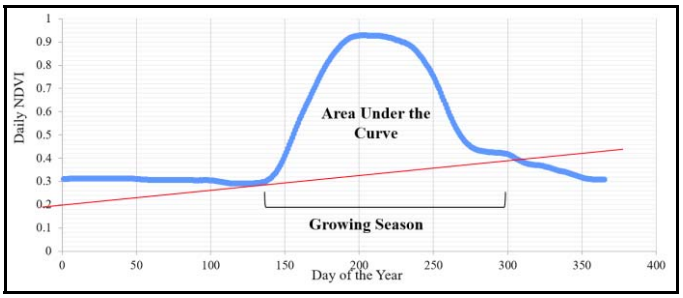


Fig. 4. Deriving Area Under the Curve

E. *Regression*

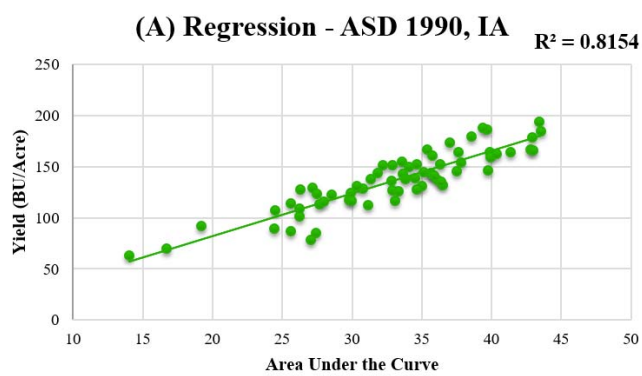
Regression model to assess corn yield based on NDVI was derive using AUC as independent variable and annual county level yield ration as depend variable. This research assumes all counties within a single ASD are homogenous, hence derives one model for each ASD utilizing all counties within the ASD for years 2000 to 2014. For the selected ASD (1920, 1940, 1970, and 1990) as case study for this research, the total counties ranged 10 to 14 per ASD, so the total observation to build each ASD level regression model was between 140 to 194.

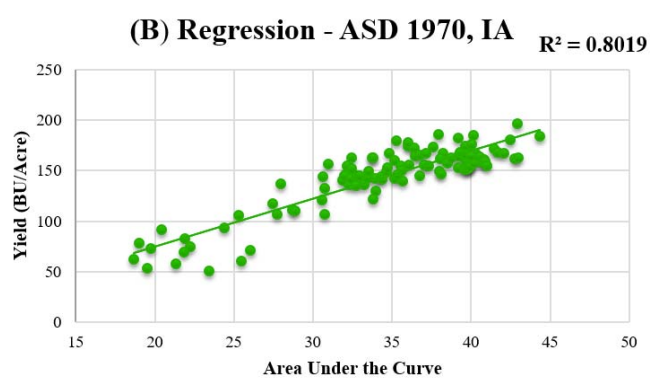
IV. Result/Discussion

The result achieved through the study and the validation of the outcomes are highlighted in the following section.

A. *Regression Result*

For all 4 ASD, The linear regression exhibit statistically significant relation between the AUC and county level annual corn yield ratio. All four regression models had the coefficient of deterministic (R-squared) over 0.80. This indicates that 80% of the time the independent variable is able to define the dependent variable. Fig 5 (A-D) represents the linear regression relationship between the yield and AUC for all 4 ASD.





Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:48:15 UTC from IEEE Xplore. Restrictions apply.

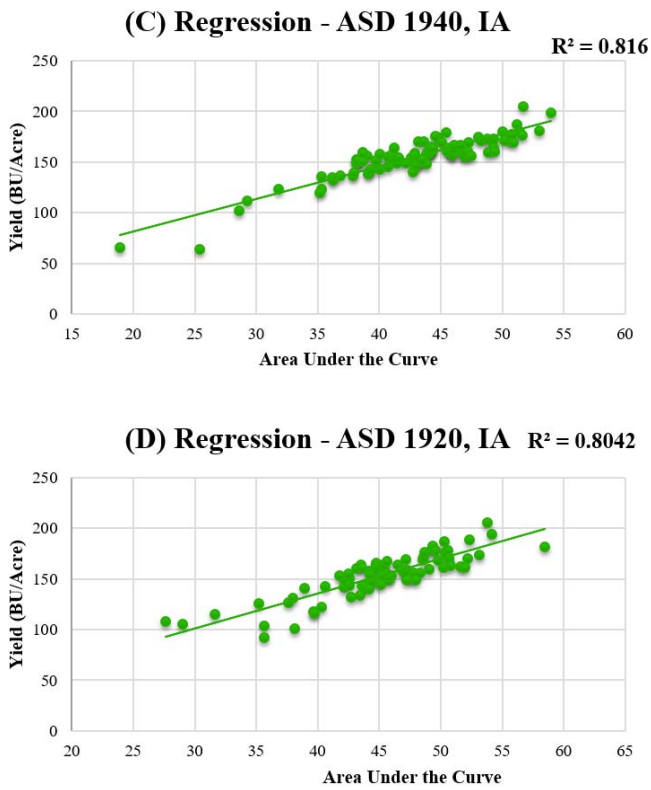


Fig. 5. ASD level regression model – (A) ASD 1990, (B) ASD 1970, (C) ASD 1940, (D) ASD 1920.

B. *Validation*

The validation of the regression model was done by using 2015 data. As the regression model was derived from the corn NDVI and corn yield data from 2000 – 2014 data, using 2015 corn NDVI and yield data, the validation result will be unbiased. Following the similar approach highlighted in method section, for each ASD in the study area, normal corn NDVI was derived for 2015. After smoothing the NDVI time-series, AUC was calculated for each ASD. These AUC were then used in the regression model as independent variable to calculate the corn yield for all the counties within the study area ASD for 2015. The results are shown in Table 1 (A-D).

The validation outcome showed promising results. For all 4 ASD, the average percentage difference between yield observed and yield calculated was between 5 to 7 %, with ASD 1990 had the best result with 5%. These results indicates that the derived regression model's performance in assessing corn yield with NDVI is fairly accurate. Although in cases the percentage difference between observed and calculated yield was as high as 28%, however the majority of the difference was within 10%.

Table 1: Regression Validation Result

Table 1 (A) : 2015, ASD 1940, IA		
Yield Observed	Yield Calculated	Percentage Diff
189.8	180.2	5%
194.7	178.2	8%
195.4	190.5	3%
199.1	183.7	8%
189	162.6	14%
179.2	160.2	11%
189.5	174.8	8%
203.1	192.7	5%
169	174.7	-3%
204.8	192.8	6%
195.3	189.4	3%
202.3	181.5	10%
Average Percentage Difference		**6%**

Table 1(B) : 2015, ASD 1920, IA		
Yield Observed	Yield Calculated	Percentage Diff
193.3	175.2	9%
196.9	200.6	-2%
196.5	177.6	10%
200.2	200.8	0%
201.2	186.8	7%
194.8	171.2	12%
199	161.3	19%
202.1	188.5	7%
203.6	183.2	10%
202.1	187.5	7%
198.4	192.7	3%
Average Percentage Difference		**7%**

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:48:15 UTC from IEEE Xplore. Restrictions apply.

Table 1 (C): 2015, ASD 1990, IA		
Yield Observed	Yield Calculated	Percentage Diff
142.4	142.6	0%
176	147.4	16%
178.4	152.0	15%
168.2	151.6	10%
181	179.4	1%
145.1	154.3	-6%
181.7	168.3	7%
190.3	186.5	2%
158.3	114.2	28%
159.3	154.4	3%
195	194.3	0%
Average Percentage Difference		**7%**

Table 1 (D): 2015, ASD 1970, IA		
Yield Observed	Yield Calculated	Percentage Diff
176.5	170.8	3%
169.6	159.7	6%
193.7	190.5	2%
179.9	165.4	8%
185.2	177.0	4%
165.8	161.0	3%
204	185.7	9%
Average Percentage Difference		**5%**

V. Conclusion

Based on the study and observed result, the NDVI showed statistically significant relation with the corn yield in the 4 ASD in IA State. With the ASD level regression model between NDVI AUC and the corn yield, the mode showed high coefficient of deterministic (R- squared) over 0.80. For the further validation, new 2015 corn yield data were obtained and used in the model. Based on the validation outcome, the average percentage difference among all ASD regression model was between 5 to 7 %. This suggest that the regression model provides good predictability results. That said, the modeling was only performed within single state and for a single crop (corn), hence testing broader spatial level of modeling as well as diverse crop groups in future might provide better understanding in the topic.

Acknowledgment

This study was supported by a grant from NASA Applied Science Program (Grant # NNX14AP91G, PI: Prof. Liping Di).

References

[1] E. Pantaleoni, B. A. Engel, and C. J. Johannsen, "Identifying agricultural flood damage using Landsat imagery," *Precis. Agric.*, vol. 8, no. 1–2, pp. 27–36, Apr. 2007.

[2] N. A. Quarmby, M. Milnes, T. L. Hindle, and N. Silleos, "The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction," *Int. J. Remote Sens.*, vol. 14, no. 2, pp. 199–210, Jan. 1993.

[3] L. C. Smith, "Satellite remote sensing of river inundation area, stage, and discharge: A review," *Hydrol. Process.*, vol. 11, no. 10, pp. 1427–1439, 1997.

[4] P. S. Frazier and K. J. Page, "Water body detection and delineation with Landsat TM data," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 12, pp. 1461–1468, 2000.

[5] R. Shrestha and Liping Di, "Land/water detection and delineation with Landsat data using Matlab/ENVI," *Agro-Geoinformatics Agro-Geoinformatics 2013 Second Int. Conf. On*, pp. 211–214, Aug. 2013.

[6] Donglian Sun, Yunyue Yu, and M. D. Goldberg, "Deriving Water Fraction and Flood Maps From MODIS Images Using a Decision Tree Approach," *Sel. Top. Appl. Earth Obs. Remote Sens. IEEE J. Of*, vol. 4, no. 4, pp. 814–825, Dec. 2011.

[7] A. K. Prasad, L. Chai, R. P. Singh, and M. Kafatos, "Crop yield estimation model for Iowa using remote sensing and surface parameters," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 8, no. 1, pp. 26–33, Jan. 2006.

[8] F.-O. Tapia-Silva, S. Itzerott, S. Foerster, B. Kuhlmann, and H. Kreibich, "Estimation of flood losses to agricultural crops using remote sensing," *Recent Adv. Mapp. Model. Flood Process. Lowl. Areas*, vol. 36, no. 7–8, pp. 253–265, 2011.

[9] C. Domenikiotis, A. Loukas, and N. R. Dalezios, "The use of NOAA/AVHRR satellite data for monitoring and assessment of forest fires and floods," *Nat. Hazards Earth Syst. Sci.*, vol. 3, no. 1/2, pp. 115–128, 2003.

[10] Lingjun Kang, Liping Di, Yuanzheng Shao, E. Yu, Bei Zhang, and R. Shrestha, "Study of the NDVI-precipitation correlation stratified by crop type and soil permeability," *Agro-Geoinformatics Agro-Geoinformatics 2013 Second Int. Conf. On*, pp. 194–199, Aug. 2013.

[11] M. S. Rasmussen, "Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR.," *Int. J. Remote Sens.*, vol. 13, no. 18, pp. 3431–3442, Dec. 1992.

[12] Genong Yu, Liping Di, Bei Zhang, Yuanzheng Shao, R. Shrestha, and Lingjun Kang, "Remote-sensing-based flood damage estimation using crop condition profiles," *Agro-Geoinformatics Agro-Geoinformatics 2013 Second Int. Conf. On*, pp. 205–210, Aug. 2013.

[13] USDA ERS, "USDA Economic Research Service - Corn," 2016. [Online]. Available: http://www.ers.usda.gov/topics/crops/corn.aspx. [Accessed: 2-June-2016].

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:48:15 UTC from IEEE Xplore. Restrictions apply.