International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 5 Issue 8 ǁ August. 2017 ǁ PP. 51-60

Yield Forecasting to Sustain the Agricultural Transportation UnderStochastic Environment

Nimish Dharmadhikari[1*],Kambiz Farahmand[2],EunSu Lee[3], Kimberly Vachal[4]David Ripplinger[5]

1Indian Nations Council of Governments, 2 West 2nd Street, Suite 800, Tulsa, OK 74105
2Dept. of Industrial and Manufacturing Engineering, North Dakota State University, NDSU Dept. 2485, PO Box 6050, Fargo, North Dakota 58108-6050
3Dept. of Management, New Jersey City University, 2039 Kennedy Blvd. Jersey City, NJ 07305
4Program Director, Upper Great Plains Transportation Institute, North Dakota State University, NDSU Dept. 2880, P.O. Box 6050, Fargo, ND 58108-6050
5Dept. of Agribusiness and Applied Economics, NDSU Dept. 7610, PO Box 6050, Fargo ND 58108-6050
**Corresponding author: Nimish Dharmadhikari[1*]*

ABSTRACT: Agricultural transportation is a major part of the United States' transportation systems. This system follows a complex multimodal network consisting of highway, railway, and waterways which are mostly based on the yield of the agricultural commodities and their market values. The yield of agricultural commodities is dependent on stochastic environment such as weather conditions, rainfall, soil type and natural disasters. Different techniques such as leaf growth index, Normalized Difference Vegetation Index (NDVI), and regression analysis are used to forecast the yield for the end of harvest season. The yield forecasting techniques are used to predict the agricultural transportation needs and improve the cost minimization. This study provides a model for yield forecasting using NDVI data, Geographical Information System (GIS), and statistical analysis. A case study is presented to demonstrate this model with a novel tool for collecting NDVI data.
Keywords:Yield Forecasting; Geographical Information System; NDVI

I. INTRODUCTION

Agricultural transportation is a major part of the United States' freight transportation and overall transportation systems. Agricultural commodities such as Cereal Grains represent one of the top ten commodities by weight originating in United States as shown in Table 1.[1]. It involves several stages of the transportation activities namely 1) from farm to storage, 2) from storage to production facility, and 3) from production facility to market. Farmers are involved in the first step 'from farm to storage.' Logistical costs for this step are accrued by farmers. This step is incorporated of different modes of transportation like roads, railways, or waterways. The logistical costs are dependent on the yield of the crop, shipping distance, and the harvest time of the year. Though distance and the time of the year are fairly similar each year, yield of the crop can change drastically year to year. The yield of agricultural commodities is dependent on stochastic environments such as weather conditions, precipitation, soil type, and natural disasters. This consideration of stochastic environment is important for the freight transportation modeling as trip origins are seasonal and yield of the commodity is varying. Transportation models are forecasting models which consist of a series of mathematical equations that are used to represent how people travel [2]. The freight transportation models substitute people with different commodities and forecast how freight travels and how it affects the transportation network. Thus, it is important for farmers and freight models to have a yield forecasting model to predict logistical needs for the agricultural transportation.

Table 1. Top 10 Commodities in U.S. for year 2013

#	Commodity	Weight (millions of Tons)
1	Gravel	2,427
2	Cereal grains	1,665
3	Non-metallic mineral products	1,514
4	Waste/scrap	1,441
5	Natural Gas, coke, asphalt	1,403
6	Coal	1,263
7	Gasoline	1,029
8	Crude Petroleum	839
9	Fuel Oils	757
10	Natural Sands	620

Yield can be defined as the harvested amount of agricultural product. Yield is measured in different unit systems. Senay and Verdin[3] use ton/hectors (t/ha) unit to measure the yield in Ethiopia. They provide yield as ratio between production in tones and planted area in hectors. Agricultural reports in United States use bushels as a unit of agricultural yield. Murphy[4] provides the weights per bushels of different crops on University of Missouri Extension website. According to the data, wheat weighs 60 pounds per bushel and corn weighs 56 pounds per bushel. Agricultural transportation becomes a stochastic process as it is dependent on the yield of the commodity. The yield of the commodity is not constant. It is dependent on weather conditions, rainfall, soil type, and other factors Figure 1.

Highlights

- A yield forecasting model is developed.
- Raster data conversion using Geographical Information System (GIS) is used.
- Statistical analysis is followed the GIS process.

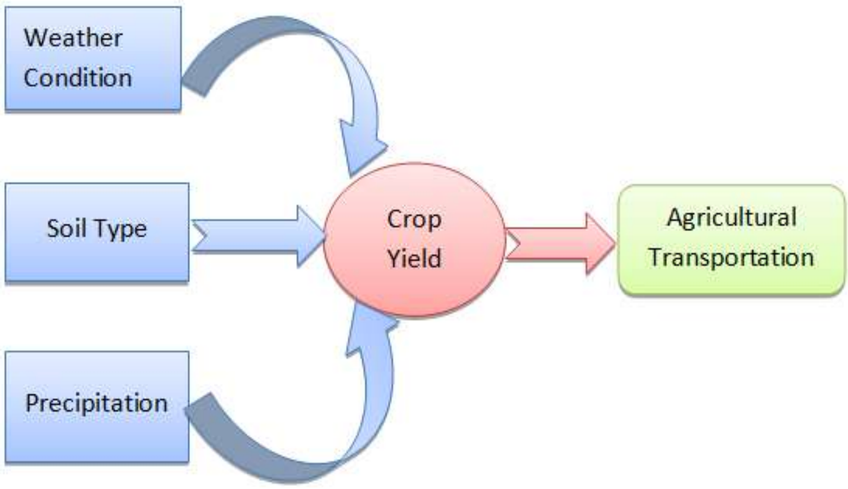


Figure 1. Agricultural Transportation Process.

This paper presents a study of yield forecasting with the help of satellite imaging and geographic information system (GIS). This research can be useful for farmers, transportation professionals, and logistical operators. Models and algorithm developed in this study can be replicated for different crops and harvest times. The paper is organized as follows: Section 2 reviews current and past literature about relevant yield forecasting techniques, Section 3 demonstrates the method used in the study and proposes the variables and data required for the analysis, Section 4 presents the data analysis part with results, and Section 5 discusses insights gained from this study and future research avenues.

II. LITERATURE REVIEW

In the research and production of the agricultural products it is important to study different efforts of yield forecasting. Yield forecasting is used to estimate the yield in one season or one year. The results of the yield forecasting are useful for transportation and economic decision making and providing policy changes. Yield forecasting is carried out with the help of different techniques. This research concentrates on the techniques using Geographical Information System (GIS) and other mathematical algorithms to determine and forecast agricultural yield. Following section reviews different efforts of yield forecasting in the literature.

Donatelli et al. [5] developed the agricultural production and externalities simulator (APES). This simulator consists of different modules starting from soil, water, and crop with diseases and chemicals. A new combined model for biomass growth and crop yield forecasting from low cost measurements has been developed and calibrated for the conditions in Pakistan. Advance very high resolution radiometer, coarse resolution - pixel size 1.1 km [6]. The spatial structure of the relationship between rainfall and groundnut yield has been explored using empirical orthogonal function (EOF) analysis. They finally use district level scale which is highly correlated (R^2= 0.96). Their experiment studied different spatial scales to use in the forecasting[7]Cantelaube and Terres[8] present a model with two parts: 1) Crop growth simulation and2) Regression analysis. Data used is the weather data from DEMETER climate models. They observed use of climatic forecast in crop yield modeling provides better predictions. Net primary productivity (NPP) was estimated using PAR and NDVI. They use NPP equation by Goward and Huemmrich (1992). Results of the study show that it is possible to monitor crop growth and assess grain yield on a large scale through the integration of satellite imagery, field data, and growth modeling. [9]

Louhichi et al. [10]developed the Farm System Simulator (FSSIM), which helps to present a bio-economic modeling to provide insight to complex agriculture structure of Europe. Senay and Verdin [3] connect the yield reduction in Ethiopia with the water balance using a GIS based model. Then, they use the water requirement satisfaction index (WRSI) and data acquired by GIS to find interrelationship between yield of maize, sorghum and teff. They suggest this WRSI based estimates can be used as a warning system for different crops.

Liu[11] in his research about crops prepares a GIS tool for modeling the relation between crops and water. He connected a GIS part with Environmental Policy Integrated Climate (EPIC) model to simulate dynamics of soil–crop–atmosphere-management system. He also enlists different available models for the food production. He categorizes them in Physical model, Economic model, Physical-economic model, Time series model, Regression analysis model, and integrated model. He uses an integrated model. This model adds the crop growth model and soil evaporation model with the help of GIS to simulate the plant growth. In the case study this GEPIC (GIS and EPIC) model estimated the crop water productivity (CWP) for three major cereal crops. The CWP is ratio of crop yield to crop evapotranspiration which is an important indicator to measure relation between crop yield and water consumption.

Liu et al.[12]continuing their research with EPIC model for agricultural and environmental studies provide a new model for the regions where daily weather data is not available. They develop a MODAWEC (MOnthly to DAilyWEather Converter) model to generate the daily precipitation and maximum and minimum temperature. They use monthly precipitation, maximum and minimum temperature, and number of wet days to generate daily values. They generate this reliable data to simulate the crop yield and crop weather use, their main objectives. With the help of case study they show that the quality of generated daily weather data is good enough to use in simulation of crop yield.

Chavas et al. [13] try to study the effect of the climate change on the productivity of major crops in China. They determine domain wide trends and also detect the regions which are vulnerable to future climate change and change in productivity is 10% increase or decrease. They use simulation modeling to assess the region wide changes in long term. They test the scenarios of the greenhouse gas emission and CO_2 farming. In this study, they combine regional climate model output (RegCM3) over the domain, a global soil database (or WISE), a county-level cropland database, and Chinese farm management data with the EPIC agro-ecosystem simulation model. They conclude that if the CO_2 enrichment is used, it will increase the productivity. In absence of this enrichment climate change will affect negatively on the productivity.

Mo et al.[14]developed a process based crop growth model. This model predicts regional crop growth, water consumption and water use efficiency (WUE). They use remote sensing data from North China Plain (NCP) which includes GIS, land use maps, digital elevation model (DEM) and soil texture with leaf area index (LAI). They receive their data from National Oceanic and Atmospheric Administration (NOAA) – Advanced Very High Resolution Radiometer (AVHRR) data. They simulate the winter wheat and summer maize yields in 1992 and 1993 and compare them with county level data. They state that to improve accuracy and reliability of crop yield prediction higher resolution satellite images such as MODIS should be cheaply available, preferably free on-line access, globally.

Ewert et al. [15] state the Integrated Assessment and Modeling (IAM) for the agricultural production can provide impacts on policy changes. They try to provide a flexible IA model which will allow solution of range of issues and not focus on just one problem. They propose a framework SEAMLESS-IF which integrates relationship and processes across disciplines and scales and combines quantitative analysis.

III. METHODOLOGY

From the literature review, the study found that a variety of ways areproposed for the yield forecasting out of which we choose yield forecasting using vegetation index method for this research. These vegetation indices are obtained from the satellite images. Raw or unprocessed satellite images have different issues such as cloud cover which make them difficult to use in the analysis. National Agricultural Statistics Service (NASS) has created a Vegetation Condition Explorer (VegScape) to provide simplified vegetation data as a raster image by simplifying satellite images [16]. Different vegetation indices available at VegScape are as follows: 1) Normalized Difference Vegetation Index(NDVI) Products, 2) Vegetation Condition Index (VCI), 3) Ratio of current NDVI to previous year for the same periods (RVCI), 4) Ratio of current NDVI to the median of previous years since 2000 for the same periods (RMVCI), and 5) Deviation of the vegetation to "normal" vegetation or multiple-year mean (MVCI). This research uses NDVI as the primary vegetation index for the yield forecasting model. NDVI is calculated from different satellite images with the equation 1. It is calculated with the help of visible and near-infrared light reflected by vegetation. Healthy vegetation absorbs most of visible light and reflects large near-infrared light thus NDVI is one of the most useful indices for yield forecasting [17].

$$NDVI = \frac{NIR-Red}{NIR+Red} \qquad (1)$$

Where,
NIR = near-infrared spectral reflectance measurement
Red = red spectral reflectance measurement

The methodology follows the algorithm shown in Figure 2. It begins with the VegScape. The raster datasets for intended area are downloaded. These datasets are available in daily, weekly, or biweekly timelines. Based on the computing power available and the accuracy of the analysis required one of the dataset can be selected. The raster dataset is then converted in to the polygons with each polygon having individual values for NDVI. The polygons coinciding with the yield data from desired farms are selected based on location. This gives data set of desired NDVI values for a time interval based on farm locations. This analysis is performed with the help of geographical information system (GIS). The localized NDVI values are used to perform the regression analysis to generate the forecasting model. The validation and verification tests are performed on this model. Once the tests are done the final model is presented. This is an ongoing process and there is a different model for each season.

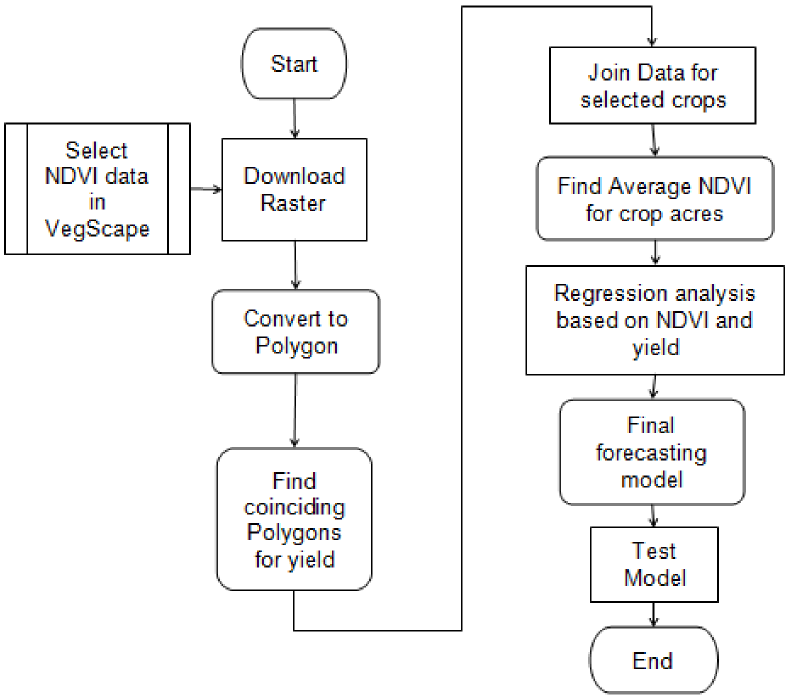


Figure 2. Algorithm for Yield forecasting method using NDVI.

Relationship between yield and NDVI was derived using regression analysis. As this model performs regression analysis on location data it can be considered as a geospatial regression model. A lineal regression model similar to Equation 2 is used as a sample model for analysis.It represents yield as a function of the NDVI at a given time period (*n-i*) until n time period for *i*time windows. ε is the deviation from the regression line. For the best fit regression, we try to minimize$\sum \varepsilon^2$.

$$Yield = NDVI_{n-i}x_{n-i} + \cdots + NDVI_n x_n + \varepsilon \quad (2)$$

This study connects the simplified satellite data available to the geographical regression modeling. This is important for farmers as well as modelers as they will have a simple tool to perform yield forecasting without carrying outan elaborate remote sensing and image processing. This will also help to make the logistical decisions at the end of the harvest season. Analysis section presents the data analysis part with the case study. The logistical cost of the sugarbeet processing can be further calculated in the research article by Farahmand, Dharmadhikari and Khiabani[18].

IV. ANALYSIS

Analysis is performed with the help of GIS and statistical analysis. Sugar beet crop from North Dakota's Cass County is chosen for the analysis step. Sugar beet production in Cass County is a co-op operation which is managed by American Crystal Sugar Company (ACSC) [19]. The location and yield data for the sugar beet farms were received from the company. The analysis is carried out in three steps. First step is of NDVI data collection with GIS analysis. Second step consists of statistical analysis using regression modeling. Third and final step tests the model developed in the second step.

Step 1: NDVI data collection with GIS analysis

The NDVI data collection is carried out with the GIS analysis. Using location based NDVI data helps with the geographical regression process. The sugar beet locations are added in the GIS document. The locations are shown in Figure 3.NDVI data is downloaded from VegScape website. Cass County is selected as the area of interest as shown in Figure 4.The information for the area of interest is entered as weekly NDVI data for selected year. The ranges of dates were selected based on the plant time and harvest period of the sugar beet crop.

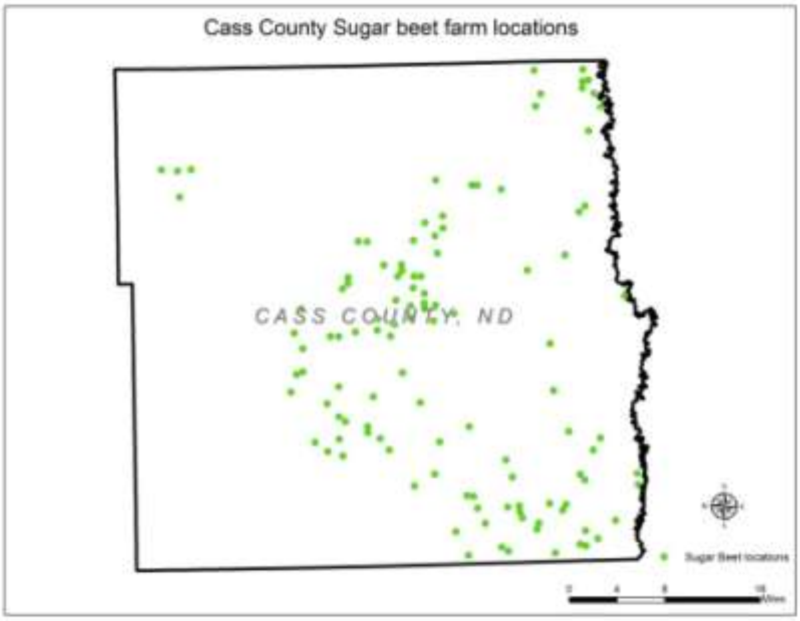


Figure 3.Sugar beet locations in Cass County of North Dakota.

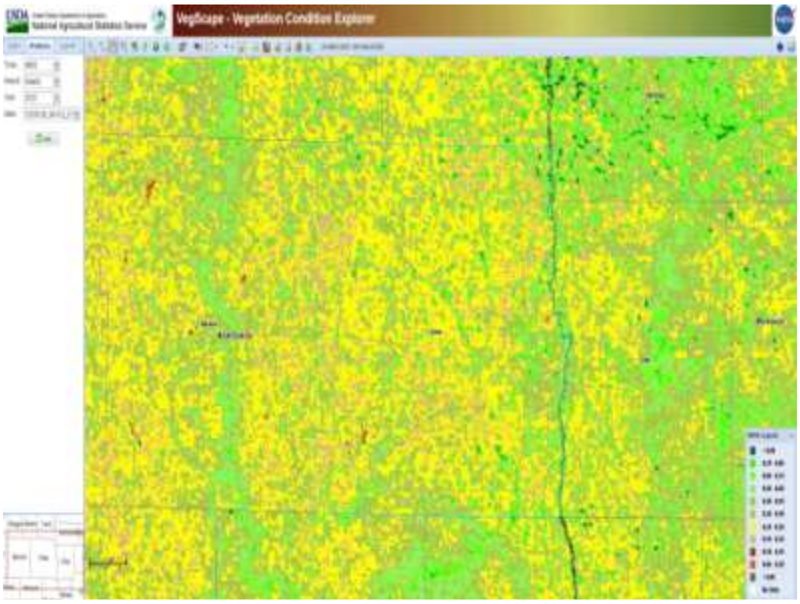


Figure 4.Screen shot of the area of interest in VegScape[16].

The downloaded data comes in the raster form. The raster data is converted into polygons. A polygon in GIS is a vector object which can be considered as the boundary of each field in this case. These polygons are used to attach the locations to the NDVI values. The polygons are created using ArcGIS software and 'raster to polygon' tool. The polygons are further joined with the locations of the sugar beet farms. This gives timely

NDVI data for the sugar beet farm locations. This process can be seen in Figures 5, 6, and7. The NDVI values from the polygons are joined to the farm location datasets. The NDVI values are collected for the months in which sugar beet farming takes place. They are collected twice for each month.For example,theNDVI values in early September are *N9A*. These values are then added with the dataset of previous records of the yield from the respective farms. This provides a big dataset for regression analysis.

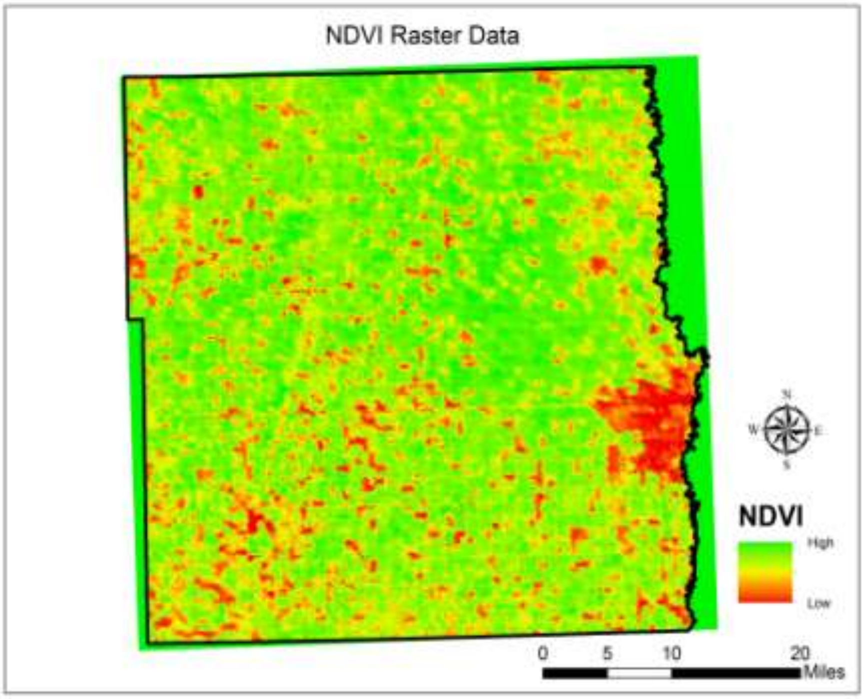


Figure 5. NDVI data in Raster form.

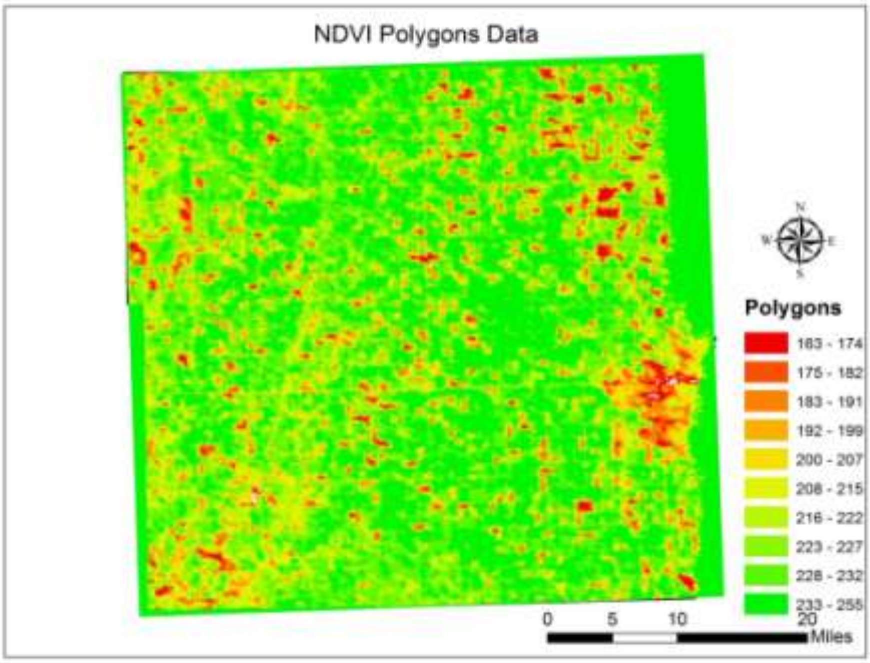


Figure 6. NDVI data in Polygon form.

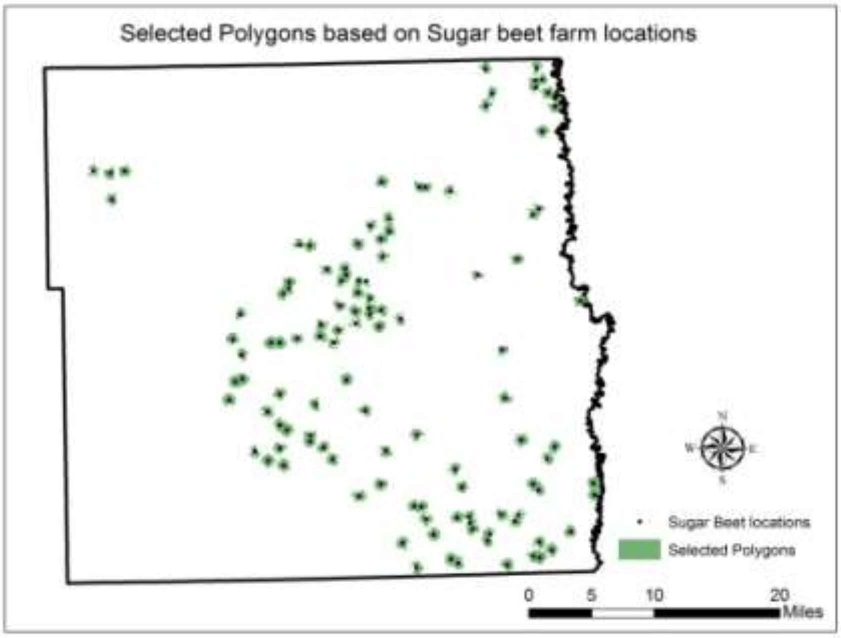


Figure 7. Selected polygons based on locations of Sugar beet farms.

Step 2: Regression analysis

As mentioned in the methodology we use linear regression model for performing this analysis. Regression model is similar to the model shown in Equation 2. SAS® Enterprise Guide software is used to perform the regression analysis. The results of regression analysis are presented in Table 2.

Table 2. Preliminary Results from Regression Analysis

Number of Observations Read	4180
Number of Observations Used	3585
Number of Observations with Missing Values	595

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	69113	23038	2027.79	<.0001
Error	3581	40684	11.36095		
Corrected Total	3584	109797			

Root MSE	3.37060	R-Square	0.6295
Dependent Mean	20.75227	Adj R-Sq	0.6292
CoeffVar	16.24208		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-27.41441	0.69801	-39.28	<.0001
N8A_mean	1	32.53307	1.07348	30.31	<.0001
N9_mean	1	9.63945	0.57729	16.70	<.0001
N8B_mean	1	24.70714	1.37598	17.96	<.0001

The adjusted R-square value of 0.6292 is within the acceptable range. Also, residual plot in Figure 8 shows the distribution of residuals for yield is normal. Using the parameter estimates from Table 2 a preliminary regression for forecasting can be developed. Equation 3 represents the forecasting equation for yield for each farm. This regression equation can be used to forecast the yield at the farm in different years. The validation of this equation is important. The step 3 is the validation process.

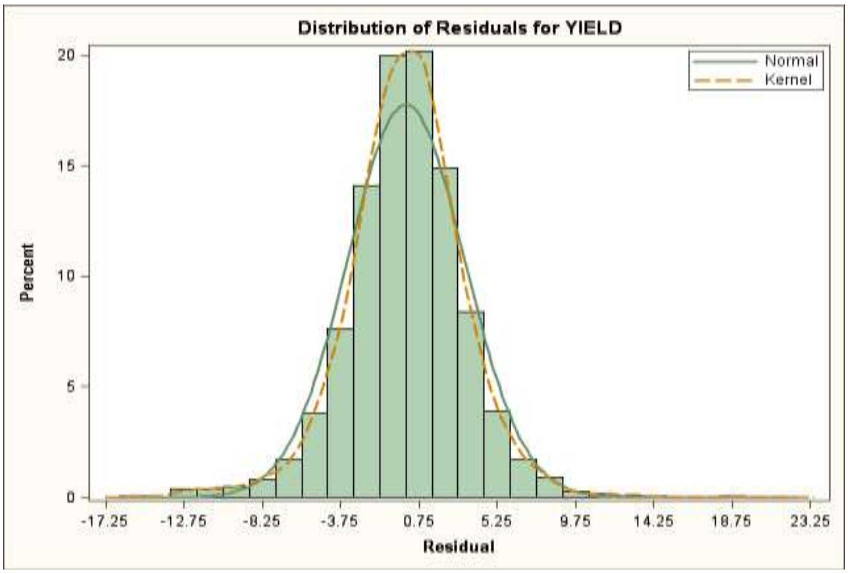


Figure 8 Distribution of residuals (ε) for yield

Preliminary equation:
$\widehat{yield} = -27.41 + 32.53(N8A) + 9.63(N9) + 24.70(N8B)(3)$

Where,
yield: Yield at a farm.
N8A: NDVI Value reading at that farm for early August
N9: NDVI Value averaged for month of September
N8B: NDVI Value reading at that farm for late August

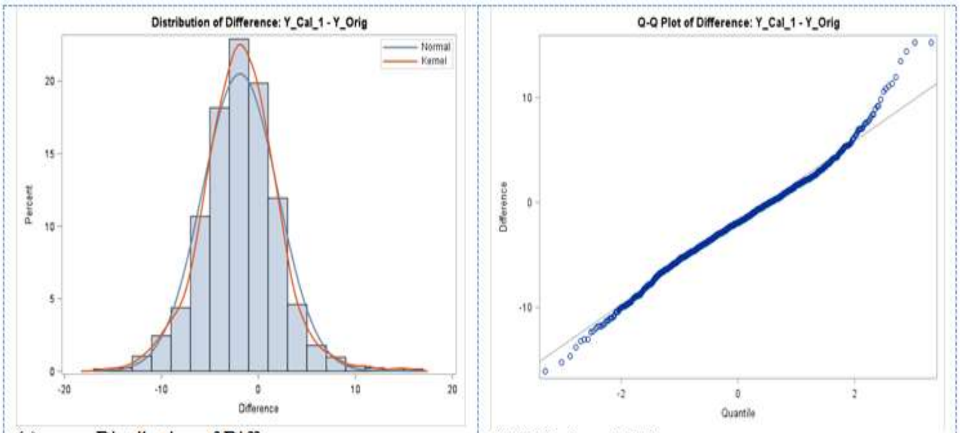
Step 3: Validation

The validation process includes forecasting and comparison parts. A different year'sNDVI data is collected similar to previous process. This data is added for similar farms as earlier year's farms. The Equation 3 is used to forecast the yield using NDVI data. This yield population is compared with the actual yield numbers. We assume that the difference between two populations is small. The t-test results for comparing two samples are presented in Table 3. In the table, *Y_Orig* is the actual yield and *Y_Cal_1* is the yield calculated based on model.

Table 3.t-Test results

	Y_Orig	*Y_Cal_1*
Mean	24.90600445	23.00976276
Variance	14.00129181	4.882805396
Observations	1349	1349
Pearson Correlation	0.226423724	
Hypothesized Mean Difference	2	
df	1348	
t Stat	-0.979419744	
P(T<=t) one-tail	0.163774183	
t Critical one-tail	1.645984801	
P(T<=t) two-tail	0.327548366	
t Critical two-tail	1.961725384	

P-value is 0.3275 thus we can't reject the null hypothesis. Plots in Figure 9 depicts that the distribution of difference is normal. We can say that the difference between two populations may be small. Thus it can be said that the forecasting equation provides results closer to the actual yield. These results can be used in the making of logistical decisions in the future to reduce the logistical costs.



(a) Distribution of Difference (b) Q-Q plot of difference

Figure 9. Normality test using (a) distribution ofdiffernce and 9b0 Q-Q plot of difference.

V. CONCLUSION AND FUTURE RESEARCH

This paper provides insight about using yield forecasting techniques to sustain agricultural transportation. Yield forecasting can be carried out with the help of GIS and statistical analysis. USDA's VegScape product can be effectively used to acquire NDVI data. The analysis part shows that the yield NDVI data can be successfully used to predict the yield. This yield prediction is important for the logistical operations. If the yield is predicted with some part of certainty, the logistical operational decisions at the harvest season can be made with the ease. This will help to reduce the logistical cost. The forecasted data can be used in demand analysis as disaggregated or aggregated based on township or county. The methodology developed in this paper can be used by farmers, co-operatives, logistical operators, or producers. This methodology is based on the free data which is available readily from USDA.

The model tested here is an ordinary least squares (OLS) regression model. As the accuracy and precision of the data gathered increases, the forecasted model can be improved. Addition of other variables such as soil type, fertilizer with the VegScape NDVI data can help the model to perform better. Other analysis such as crop mix analysis and crop rotation analysis can be used to provide insights to the model. Drones or unmanned aerial vehicles (UAV) can be used to gather different data that can be used in such model. This will help to fine tune the model and generate more useful and timely results. The results from this model can be used in the optimization model which can optimize the total logistical cost based on planting, fertilizing, and harvesting cost, in addition to transportation and storage cost.

Software Availability

Name of the software: ArcMap, ArcGIS
Developer: ESRIhttp://resources.arcgis.com/en/home/
Current Version: 10.5
Hardware requirements:
CPU Speed: 2.2 GHz minimum, Memory/RAM: Minimum 4 GB, Recommended 8 GB, Screen resolution: 1024x768 recommended minimum at normal size (96 dpi), Video/Graphics adapter: 64 MB RAM minimum; 256 MB RAM or higher recommended. NVIDIA, ATI, and Intel chipsets supported.
Name of the software: SAS 9.4, SAS Enterprise Guide
Developer: SAS Institute Inc. https://www.sas.com/en_us/software/sas9.html
Current Version: 9.4
Hardware requirements: Basic desktop/laptop PC

REFERENCES

[1] U.S. Department of Transportation, "Freight Facts and Figures 2015," 2015. [Online]. Available: http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/FFF_complete.pdf. [Accessed 13 12 2016].

[2] E. A. Beimborn, "A Transportation Modeling Primer," June 2006. [Online]. Available: https://www4.uwm.edu/cuts/utp/models.pdf. [Accessed 12 December 2016].

[3] G. B. Senay and J. Verdin, "Characterization of Yield reduction in Ethiopia using GIS-based crop water balance model," Canadian Journal of Remote Sensing, pp. 687-692, 2003.

[4] W. J. Murphy, "Tables for Weights and Measurement: Crops," 1993. [Online]. Available: http://extension.missouri.edu/publications/DisplayPub.aspx?P=G4020. [Accessed 13 02 2013].

[5] M. Donatelli, G. Russell, A. E. Rizzoli, M. Acutis, M. Adam and I. Athanasiadis, "A Component-Based Framework for Simulating Agricultural Production and Externalities," in Environmental and Agricultural Modeling, Springer Netherlands, 2010, pp. 63-108.

[6] W. G. M. Bastiaanssen and S. Ali, "A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan," Agriculture, Ecosystems and Environment, p. 321–340, 2003.

[7] A. J. Challinor, J. M. Slingo, T. R. Wheeler, P. Q. Craufurd and D. Grimes, "Toward a Combined Seasonal Weather and Crop Productivity Forecasting System: Determination of the Working Spatial Scale," American Meteorological Society, pp. 175-192, 2003.

[8] P. Cantelaube and J. M. Terres, "Seasonal weather forecasts for crop yield modeling in Europe," Tellus A, pp. 476-487, 2005.

[9] A. D. Baez-Gonzalez, P.-y. Chen, M. Tiscareno-Lopez and R. Srinivasan, "Using Satellite and Field Data with Crop Growth Modeling to Monitor and Estimate Corn Yield in Mexico," Crop Science, p. 1943–1949, 2002.

[10] K. Louhichi, "A Generic farming system simulator," in Environmental and Agricultural Modelling, Springer, 2010, pp. 109-132.

[11] J. Liu, "A GIS-based tool for modelling large-scale crop-water relations," Environmental Modelling & Software, pp. 411-422, 2009.

[12] J. Liu, J. R. Williams, X. Wang and H. Yang, "Using MODAWEC to generate daily weather data for the EPIC model," Environmental Modelling & Software, p. 655 – 664, 2009.

[13] D. R. Chavas, R. C. Izaurralde, A. M. Thomson and X. Gao, "Long-term climate change impacts on agricultural productivity in eastern China," Agricultural and Forest Meteorology, p. 1118–1128, 2009.

[14] X. Mo, S. Liu, Z. Lin, Y. Xu, Y. Xiang and T. R. McVicar, "Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain," Ecological Modelling, p. 301–322, 2005.

[15] F. Ewert, M. K. van Ittersum, I. Bezlepkina, O. Therond, E. Andersen, H. Belhouchette, C. Bockstaller, F. Brouwer, T. Heckelei, S. Janssen, R. Knapen, M. Kuiper, K. Louhichi, J. A. Olsson, N. Turpin, J. Wery, J. E. Wien and J. Wolf, "A methodology for enhanced flexibility of integrated assessment in agriculture," Enviornmental Science and Policy, pp. 546-561, 2009.

[16] National Agricultural Statistic Service, "VegScape - Vegetation Condition Explorer," 2016. [Online]. Available: https://nassgeodata.gmu.edu/VegScape/. [Accessed 13 12 2016].

[17] P. A. Propastin and M. Kappas, "Reducing Uncertainty in Modeling the NDVI–Precipitation Relationship: A Comparative Study Using Global and Local Regression Techniques," GIScience & Remote Sensing, vol. 45, no. 1, pp. 47-67, 2008.

[18] K. Farahmand, N. Dharmadhikari and V. Khiabani, "Analysis of Transportation Economics of Sugar-Beet Production in the Red River Valley of North Dakota and Minnesota using Geographical Information System," Journal of Renewable Agriculture, 2013.

[19] K. Farahmand, V. Khiabani, N. Dharmadhikari and A. Denton, "Economic Model Evaluation of Largest Sugar-beetProduction in U.S. States of North Dakota and Minnesota," International Journal of Research in Engineering and Science, vol. 1, no. 4, pp. 11-24, 2013.

Nimish Dharmadhikari1*. "Yield Forecasting to Sustain the Agricultural Transportation Under Stochastic Environment." International Journal of Research in Engineering and Science (IJRES) 05.08 (2017): 51-60.