С.Сейфуллин атындағы Қазақ агротехникалық университетінің **Ғылым жаршысы (пәнаралық)** = **Вестник науки** Казахского агротехнического университета им. С.Сейфуллина (**междисциплинарный**). - 2019. - №4 (103). - С.33-42

ASSESSMENT OF USING LAND SURFACE TEMPERATURE (LST) AND SATELLITE REMOTE SENSING FOR WHEAT YIELD MODELLING IN THE NORTH KAZAKHSTAN REGION

Bekbayeva A.
S. Seifullin Kazakh Agro Technical University,
62 Zhenis Avenue, Nur-Sultan, 010111, Kazakhstan

Abstract

Remote sensing technology provides spatial and temporal information about objects from satellites that may contribute to crop management. The aim of study was to illustrate the application of remote sensing data for prediction wheat yield in "Agricultural experimental station" farm in the North Kazakhstan region.This study was conducted to estimate in-season yield using NDVI and land surface temperature (LST) using linear regression analysis. The land surface temperature was applied in the multiple linear regression analysis with NDVI. The accuracies were compared in terms of RMSE (root mean square error) for two models. As a result, the simple linear regression model was more accurate in comparison to the multiple regression model with LST and NDVI (4,15 and 6,78 respectively). The yield prediction using surface parameters will be further improved with the use of neural network modelling and more datasets. The simple linear model can be used for different crops of other locations.

Keywords: vegetation index, NDVI, wheat yield prediction model, remote sensing, land surface temperature (LST), North Kazakhstan

INTRODUCTION

Remote sensing is the technology of getting information about objects from remote platforms such as aircrafts, satellites. This information has spatial, spectral and temporal resolutions [1]. Remote sensing data is valuable information for crop management. Advances in remote sensing are currently providing the possibility to investigate, measure and model environmental processes. The normalized difference vegetation index was proposed as an average to calculate green biomass [2]. The NDVI describes the reflectance in the red and NIR regions to vegetation variables (NDVI= (NIR-Red)/ (NIR+Red)). There are studies on comparison of different vegetation indices for crop monitoring, but NDVI remains as the main productive index for yield prediction. The NDVI has been applied actively in crop yield evaluation and predicting [3]. There are studies with moderate success of using different methods such as numeral crop yield model [4], leastsquare regression [5], autoregressive (AR) state-space models [6],etc. Accurate yield prediction of crops is important issue in agriculture planning and national food security. The crop productivity relies on various factors, such as soil, topography, and management [7]. The sharply continental climate in Kazakhstan influences significantly on crops production. In the north part region, the irrigation technology is not developed and not recommended to apply it for chernozem types of soil. However, climate change in the region influences comparatively significantly. The shift in the start of the season particularly for the study area is seen. In the study, it was focused on wheat due to its widely cultivation in the region. The aim of this research was to assess the possibility to use the land surface temperature with vegetation indices to predict wheat yield in-season using linear regression models.

DATA AND STUDY AREA. *SITE DESCRIPTION*

The study area is located in the North part of Kazakhstan (Figure 1). The climate in the North part of Kazakhstan is sharply continental with a long snowing winter and warm summer. The average temperature of the coldest month (January) is about −18,6 °C, the hottest month (July) is +19,0 °C. The precipitation amounts to 350 mm per year, 80-85 % of these falls in the warm season (April-October). Snow cover lies about 5 monthsfrom November to March, by the end of winter has an average capacity of 25 cm [8]. The distinctive characteristic of the region is a highly productive soil type called chernozem. Therefore, agriculture in the region plays an important role in itseconomy. The main type of crop is cultivated in the region is wheat.

The private farm LLP "Agricultural experimental station" was considered in this research. The overall size of the study area is 20 000 ha (Figure 2). The LLP "Agricultural experimental station" is located on the North Kazakhstan plain. The relief of the territory is mostly flat without strong elevation changes. The average elevation is about 133,7 meters with the gradual rise of altitude from the North-West to South-East. The slopeis small the maximum is $1,59^0$occurs only in some areas. The types of soil in the farm are common chernozems, solonetzicchernozems and a complex type of meadow chernozems with semihydromorphic solonetz [9]. All soil types are characterized as the quiet high potential of productivity but yield depends on climate conditions during the vegetation period.

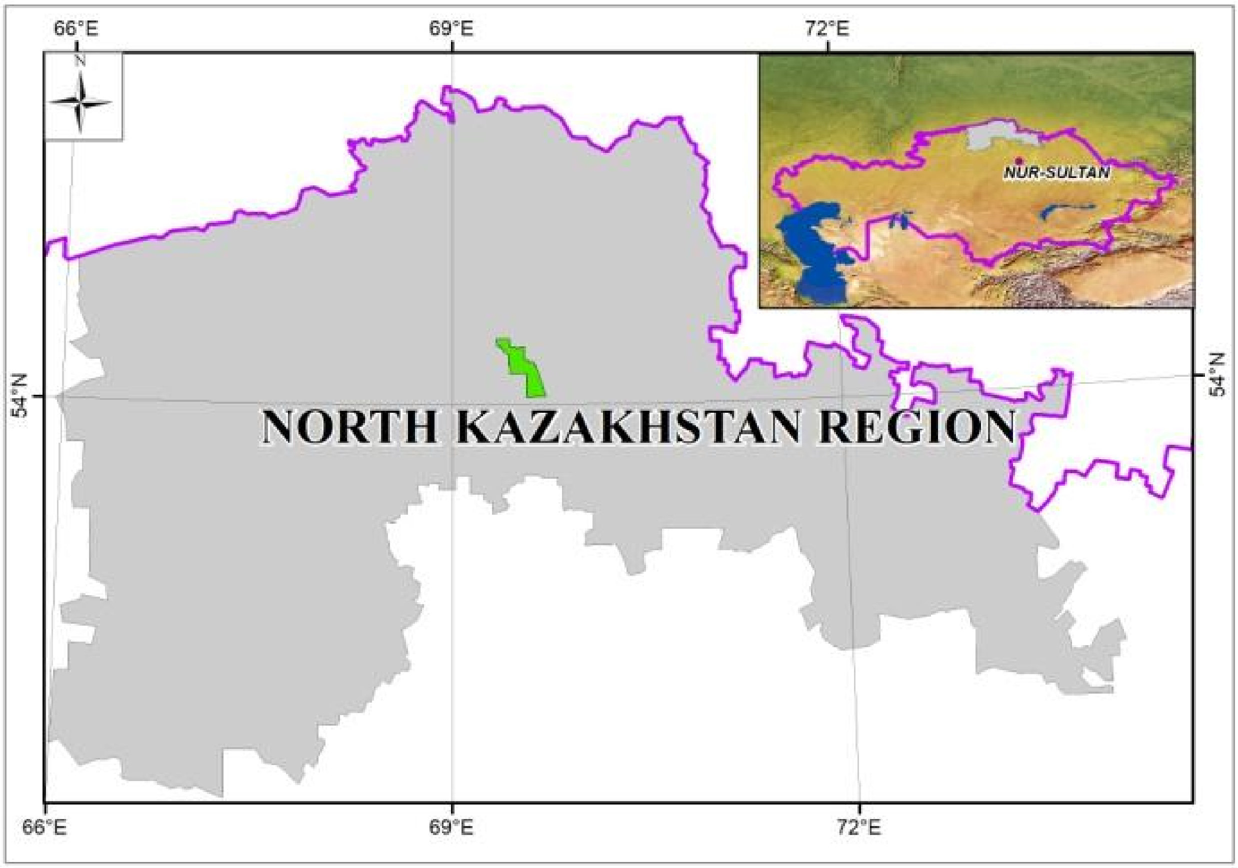


Figure 1. North Kazakhstan region

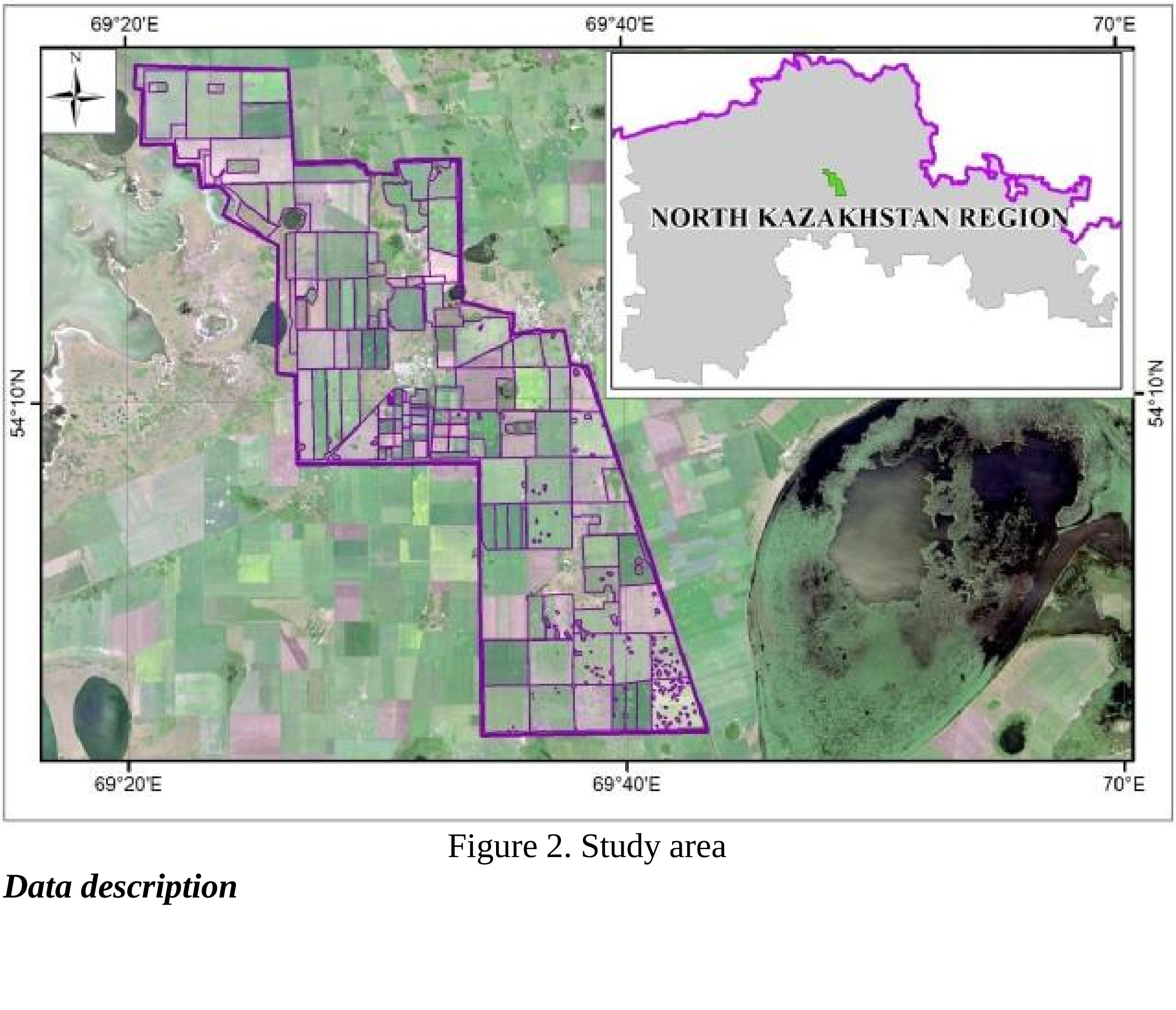


Figure 2. Study area

Data description

Crop yield data for 2018 for wheat have been used. The electronic map of fields was developed. The sowing dates of wheat are starts from the middle of May. Growth inbiomass observed from June to September with duration in average of 90 days. In September, harvesting campaigns are happening. Therefore, available satellite images and land surface temperature data was downloaded for the period from May to September.

Yield data

Average crop yield data for filed was provided by the farmer for 2018. The dataset provides data of sowing and yield in centers per hectare.

Satellite remote sensing data:

- Sentinel 2 13 bands surface reflectance data with 10-60 meters spatial resolution in 3-10 days periodicity;
- Landsat 8 OLI and TIRS 11 bands surface reflectance data with 15-100 meters spatial resolution in 16 days periodicity;
- MODIS MOD09A1 is an average 8-daythe surface spectral reflectance of Terra MODIS Bands 1-7 corrected for atmospheric conditions with 500-meter (m) spatial resolution;
- MODIS MOD11A2 is an average 8-day per-pixel land surface temperature (LST) data with a 1-kilometer (km) spatial resolution.

METHODOLOGY

The core of the methodology in this research to model yield is the use of linear regression analysis using remote sensing data. Several correlation analyses between all calculated NDVIs, average NDVI, sum NDVI and yield were analyzed. The analysis showed that the highest NDVI value in the season has a stronger correlation. Therefore, a simple linear regression model was applied. Additionally, the multiple regression model was applied using average LST data. An overview of the methodology is illustrated in Figure 3.

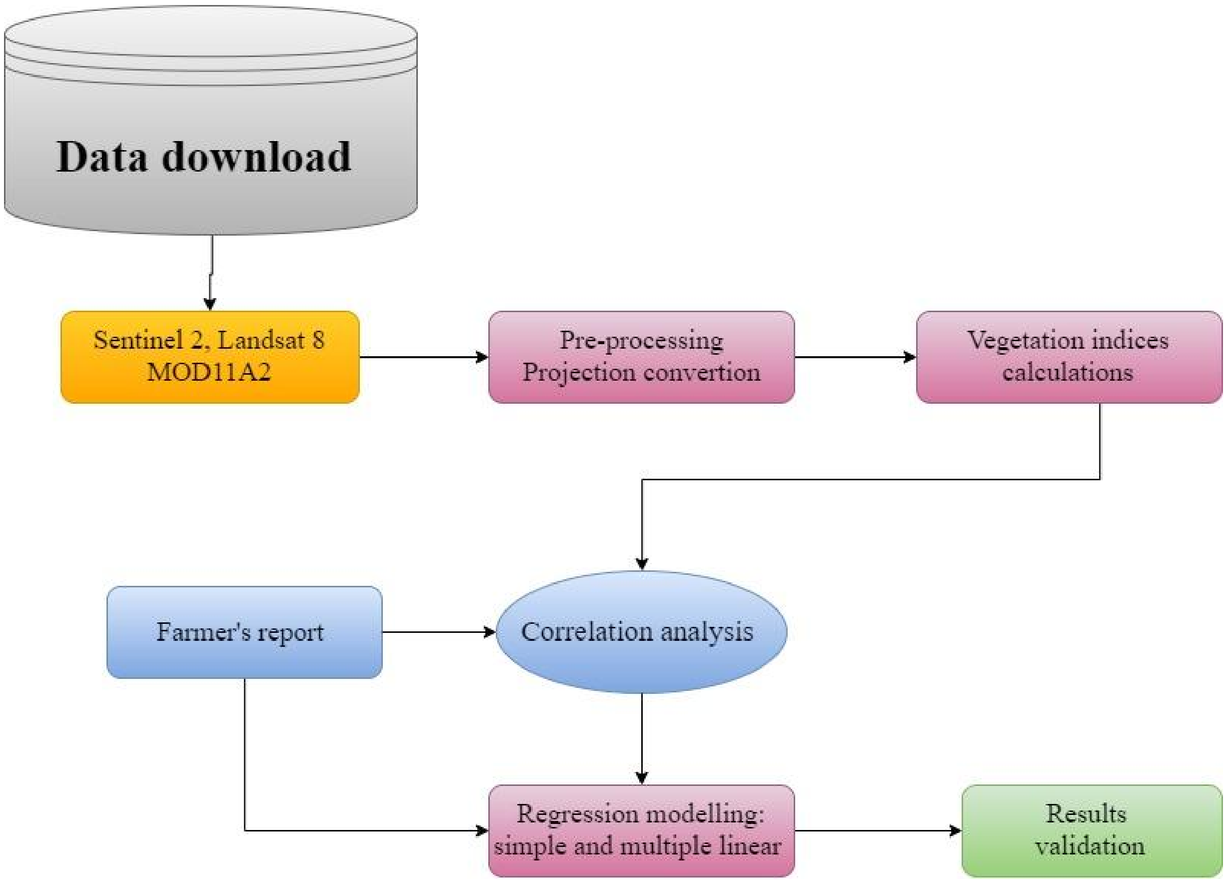


Figure 3. Flowchart of the overall methodology

Pre-processing

Sentinel 2A and Landsat 8 were atmospherically and radiometrically calibrated using ENVI software.

Radiometric correction

Radiometric correction is the first step in the normalization of raw data and it is a mathematical operation of translating the brightness values of the image pixel into the values of radiation received by the satellite sensor [10]. The original images are processed in radiance units using 32-bit floating-point values, and then these values are converted to a 16-bit integer value, which is the value in the finished L1 level image. Users can convert these values to spectral radiance values by using the luminance coefficients given in the metadata file. For such a correction, a file *_MTL.txt is present in the Landsat data set, the limit values from which are used at this stage of image processing. The following formula was used for radiometric correction [11]:

$$L_{\lambda} = M_L*Q_{cal} + A_L \quad \mathit{(15)}$$

where: L_{λ} = Spectral radiation coming to the satellite sensor (W/(m2 * sr * µm)); M_L = Radiometric channel gain (RADIANCE_MULT_BAND_nfrom metadata) (*Gainλ); *Q_{cal} = The brightness values of the pixel raw image (L1 pixel value in DN); A_L = The radiometric offset of the channel (RADIANCE_ADD_BAND_nfrom metadata) (*Baisλ).*

Then the spectral brightness values are converted into values of Top of Atmosphere Reflectance (TOA reflectance)

This procedure is necessary if more than one image is used, since some parameters, such as the angle of the sun's illumination of the surface, are unique during each shooting. Therefore, comparing any spectral values and calculating indices between two non-normalized images is incorrect [12].

Top of Atmosphere Reflectance (TOA reflectance):

$$\rho\lambda = L\lambda / \mathrm{Sin}\,(\theta)$$

where $\rho\lambda$ = TOA Reflectance; θ = Solar Elevation Angle from metadata.

Atmospheric correction

The next step in the normalization of satellite images is to reduce the influence of the atmosphere on the image and translate the values of radiation that reached the sensors of the satellite (TOA radiance) into the values of the spectral radiation of sunlight actually reflected from the earth. In the Earth's atmosphere, there is a variety of interference that can be got on the images. In the future, such interferences make it difficult to analyze the data.The influence of the atmosphere on the satellite image is manifested in a several number of factors: the angle of incidence and reflection of sunlight, the transparency of the atmosphere, the gas factor and haze [13].

Coordinate system transformation

In order to handle with MODIS product, the MCTK toolkit was used to convert the sinusoidal coordinate system of HDF files into Geographic projection. The Modis conversion toolkit was designed by D.White and it is free to download [14].

Vegetation index calculations

The normalized vegetation indices (NDVI = (NIR-RED)/(NIR+RED)) were calculated for each occurred satellite image for Sentinel 2, Landat 8, and Modis. The NDVI value growth as vegetation growth reaching the peak at the earing stage (Figure 4). Therefore, the maximum NDVI value occurs usually at the end of July. In this study, average and sum NDVI values were calculated for the identification of the relationship between crop productivity. After the earing phase the chlorophyll content in the plant decreases, this is reflected in the decrease in the vegetation index, which continues until the phase of full ripeness.

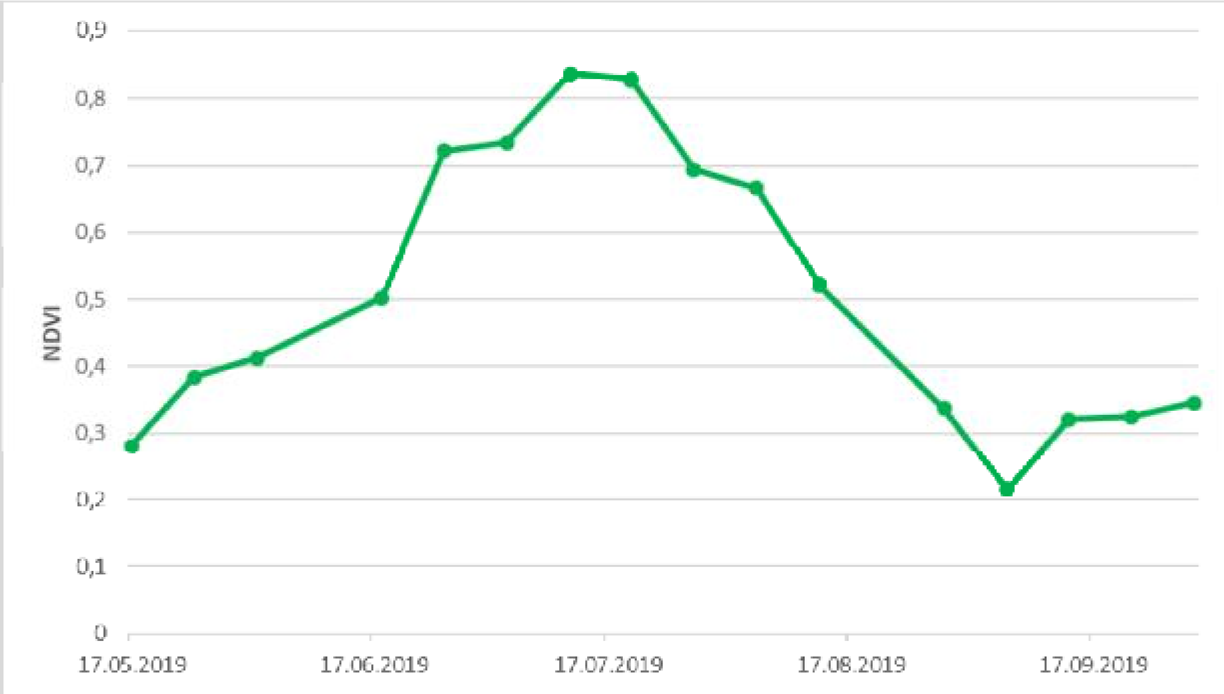


Figure 4. General NDVI trend of wheat during the vegetation season

Land surface temperature

8 days average land surface temperature datain Kelvin degrees per pixel with 1 km spatial resolution were converted to Celsius degrees. Also, datawere averaged to field size using the Zonal statistics toolbox in ArcGIS.

Correlation

To estimate the relationship between yield and remote sensing data a correlation analysis has been produced. The correlation analysis was produced in SPSS software. For this, null and alternative hypotheses were determined.

H_0: Data (NDVI and yield) are not correlated

H_A: Data (NDVI and yield) are correlated significantly.

There are two possible results might be:

1) If the probability value is greater than 0.05 hence the null hypothesis is accepted, which means the data are not correlated;

2) If the probability value is less than 0.05 hence an alternative hypothesis is accepted, which means the data are significantly correlated.

The same threshold (0.05) for the probability (P) was used in regression analysis also.

Results for each NDVI showed the presence of correlation with yield, but Pearson's r coefficients were high only for the maximum NDVI value of wheat vegetation period. The maximum NDVI was in the end of July for Sentinel 2 image. Pearson's r coefficient is equal to 0,46.

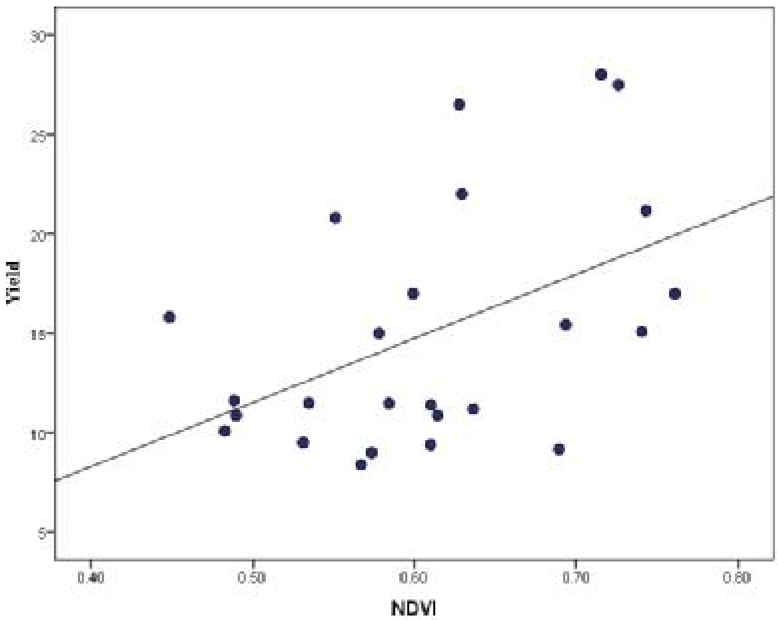


Figure 5. Correlation between yield and maximum NDVI

Therefore, this value was applied for the regression analysis.

Regression modeling

Simple linear regression

Regression analysis should be used to determine whether one variable affects another variable and to illustrate how much of an impact it has. The variable Y is fixed (or dependent) for general analysis, but the choice of variable X (or independent) is explained from the correlation analysis. Formula for simple linear regression: $y=a+bx+\varepsilon$. Crop yield was considered as the dependent variable - Y, which changes based on changes of independent variable the maximum NDVI - X.

Null and alternative hypotheses:

H_0: the slope of wheat yield regression on the max. NDVI does not differ from zero;

H_A: the slope of wheat yield regression on the max. NDVI is different from zero.

Similarly, to correlation analysis, the probability threshold was establishedto 0.05.

Multiple linear regression

To consider the influence of LST the multiple regression analysis was applied. The formula for multiple regression is $y=a+b_1*X_1+b_2*X_2 +\varepsilon$, where X_1 and X_2 are two independent variables NDVI and LST respectively.

The null and alternative hypotheses would be:

H_0:b_1 (slope of the first independent variable) does not differ from 0

H_A: b_1 (slope of the first independent variable) differs from 0

H_0:b_2 (slope of the second independent variable) does not differ from 0

H_A: b_2 (slope of the second independent variable) differs from 0.

RESULTS AND VALIDATION

The P probability values for NDVI in both models are less than 0.05 hence alternative hypotheses were accepted. At the same time, the probability for the LST slope in multiple regression was equal to 0.05. Nevertheless, the yield for both models was calculated. Additionally, the validation analysis has been produced by estimation the accuracy and root mean square error.

The produced models:

y =- 3,51+ 29,8461 *NDVI - Simple linear regression

y =18,35+31,4*NDVI -0,8 LST – Multiple linear regression

R square is 0,2 and 0,23 respectively.

Acceptance of regression models

To accept the results of regression models, it is necessary to examine the assumptions: data are independent, data are a random sample, and residuals are normally distributed and homoscedastic.

As for data used in regression models, it is a random selection and they are independent. As for residuals, the distribution the Kolmogorov-Smirnov (K-S) test was applied to test normality. The null and alternative hypothesis for K-S test.

H_0: Data are normally distributed

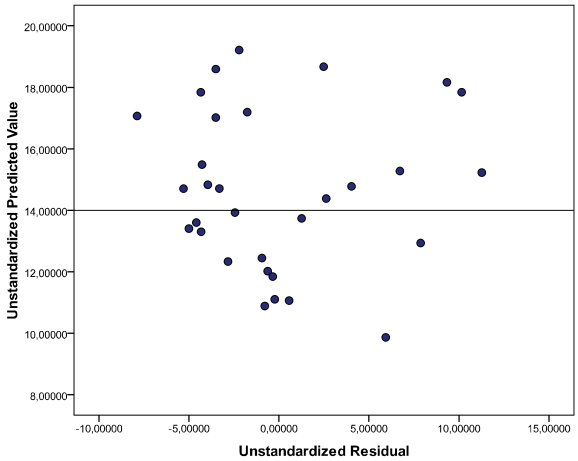
H_A: Data are not normally distributed.

The P probability value is greater than 0.05 hence the null hypothesis is accepted, which means the data are normally distributed (Table 1).

Table 1. Kolmogorov-Smirnov tests results

Data	**K-S statistic**	**P value**
Residuals (Simple regression model)	0,74	0,645
Residuals (Multiple regression model)	0,8	0,54

In Figure 6, it is illustrated that residuals from both models are homoscedastic.



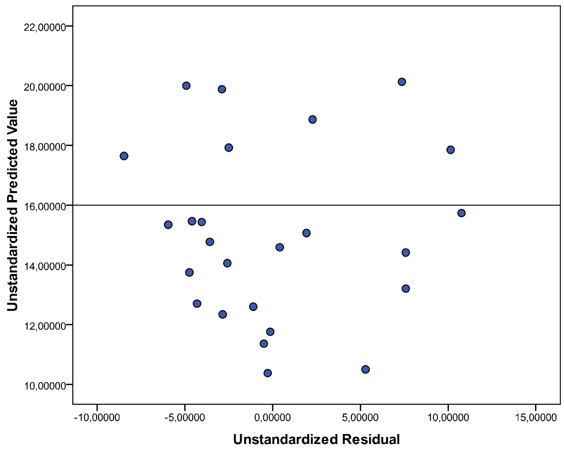


Figure 6. Homoscedasticity of residuals for simple (left) and multiple (right) regression models

The type of crop in the field for previousyear plays an important role in crop production. In table 2, the actual yield in 2019 and predicted results of the simple linear regression model are illustrated.

Table 2. Predicted and actual wheat yield

N	Square (ha)	Crop in 2018	Crop in 2019	Predicted yield Simple linear regression	Actual yield in 2019 provided by farmer
1	333,05	fallow	Wheat	14,93447544	26
2	245,24	Rape	Wheat	12,70028399	14,2
3	167,64	Linen	Wheat	9,558740467	10
4	172,66	Wheat	Wheat	9,842039894	14,2
7	451,15	Corn	Wheat	12,02044893	15,7
17	158,08	Annherbs	Wheat	10,74395997	12,8
25	684,65	Wheat	Wheat	9,034612649	9,5
34	366,01	Wheat	Wheat	10,80105561	13,1
47	232,43	fallow	Wheat	17,56624713	11,3
49	311,51	Rape	Wheat	13,96531203	18,9
50	305,01	Wheat	Wheat	11,45534238	12
52	399,69	Wheat	Wheat	11,77979962	13,7
53	395,64	Wheat	Wheat	13,09261131	15,8
54	115,35	fallow	Wheat	18,27968892	30,7
55	116,24	fallow	Wheat	18,11102846	30,7
56	117,95	fallow	Wheat	18,81411365	30,7
57	41,01	fallow	Wheat	18,26733262	30,7
58	401,47	Wheat	Wheat	14,16217708	12,5
60	217,99	Wheat	Wheat	12,55702258	12,5
61	260,33	fallow	Wheat	17,95209784	26,43
62	161,04	Wheat	Wheat	13,01748861	12,02
63	217,09	Wheat	Wheat	10,78628178	12,02
64	139,01	Wheat	Wheat	14,56139886	16,97
66	397,82	fallow	Wheat	17,05247591	24,2
68	365,43	Wheat	Wheat	11,40693197	12,2
70	259,62	Wheat	Wheat	12,43581747	12,3
72	210,85	fallow	Wheat	14,33958245	19,4
83	395,56	Rape	Wheat	10,74213936	11,76
84	399,99	Wheat	Wheat	12,38388521	12,68
86	136,36	fallow	Wheat	19,32146794	22,76
88	98,88	fallow	Wheat	19,16602932	22,76
89	106,96	fallow	Wheat	17,87924345	22,76

Validation

The model comparison was made calculating accuracy and root mean square error of each model (Table 3). Both average accuracy and RMSE showed that the lowest error followed by the simple linear regression model. The use of LST information slightly deteriorated prediction results. It should be noted, that the number of fields predicted in the multiple regression model decreased due to the spatial resolution of the MODIS dataset.

Table 3. Accuracy of prediction and root mean square error results of wheat yield.

Models	Accuracy	RMSE
Simple linear regression	88%	4,15
Multiple linear regression	71%	6,78

DISCUSSION AND CONCLUSION

This study was conducted to predict wheat yield for the North Kazakhstan region within the season. Vegetation index (NDVI) and land surface temperature were usedfor the linear regression modeling. The simple statistical analysis produces quite good results when there are not many independent variables are applied in the analysis. The important fact that should be noted that in 2019 the fallow fields were implemented by fertilizers, while last year fallow fields have not been fertilized. The average RMSE for prediction models are distorted by fields which were the fallow predecessor. As we can see from the study, the use of multiple variables produces prediction errors. On the other hand, important climatic factors should not be ignored for the yield prediction. This research will be improved with the use of machine learning technologies such as artificial neural networks, convolution neural networks, and support vector machines. It should be noted, thatresearch has limitations in predicting the yield by exceptional factors such as pests, diseases.

REFERENCES

1 Shanahan, J., Schepers, J., Francis, D., Varvel, G., Wilhelm, W., Tringe, J., Schlemmer, M. and Major, D. Use of Remote-Sensing Imagery to Estimate Corn Grain Yield. //Agronomy Journal - 2001. - № 93(3), p.583.

2 Tucker, C. Red and photographic infrared linear combinations for monitoring vegetation // Remote Sensing of Environment - 1979. - № 8(2), pp.127-150.

3 Quarmby, N., Milnes, M., Hindle, T. and Silleos, N. The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction // International Journal of Remote Sensing - 1993. - 14(2), pp.199-210.

4 Hayes, M. and Decker, W. Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. // International Journal of Remote Sensing - 1996. - № 17(16), pp.3189-3200.

5 Jones, D. A statistical inquiry into crop-weather dependence // Agricultural Meteorology - 1982. - № 26(2), pp.91-104.

6 Wendroth, O., Reuter, H. and Kersebaum, K. Predicting yield of barley across a landscape: a state-space modeling approach.// Journal of Hydrology - 2003. - № 272(1-4), pp.250-263.

7 Chlingaryan, A., Sukkarieh, S., & Whelan, B. Machine learning approaches for crop yield prediction andnitrogen status estimation in precision agriculture: A review // Computers and Electronics in Agriculture - 2018. - № 151, pp. 61-69.

8 Kazhydromet.kz. [Electronic resource] Kazgidromet. Режим доступа: http://www.kazhydromet.kz/.

9 Etomesto.ru. [Electronic resource] / Soil map of Kazakhstan 1976. Режимдоступа: http://www.etomesto.ru/map-kazakhstan_pochva-1976/.

10 Barsi, J., Alhammoud, B., Czapla-Myers, J., Gascon, F., Haque, M., Kaewmanee, M., Leigh, L. andMarkham, B. Sentinel-2A MSI and Landsat-8 OLI radiometric cross comparison over desert sites// European Journal of Remote Sensing - 2018. -№ 51(1), pp.822-837.

11 Thome K.J., Biggar S.F., Gellman D.L., Slater P.N. Absolute-radiometric calibration of Landsat-5 Thematic Mapper and the proposed calibration of the Advanced Spaceborne Thermal Emission and Reflection Radiometer. Paper presented at the Geoscience and Remote Sensing Symposium - 1994. IGARSS'94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysisand Interpretatio

12 USGS.gov. [Electronic resource] Using the USGS Landsat Level-1 Data Product. [online] Режим доступа: https://www.usgs.gov/land-resources/nli/landsat/using-usgs-landsat-level-1-data-product.

13 Ju, J., Roy, D., Vermote, E., Masek, J. and Kovalskyy, V. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. RemoteSensingofEnvironment- 2012. -№ 122, pp.175-184.

14Yceo.yale.edu. [Electronic resource] Using the MODIS Conversion Tool Kit | Center for Earth Observation. [online] Режим доступа: https://yceo.yale.edu/using-modis-conversion-tool-kit.

ASSESSMENT OF USING LAND SURFACE TEMPERATURE (LST) AND SATELLITE REMOTE SENSING FOR WHEAT YIELD MODELLING IN THE NORTH KAZAKHSTAN REGION

Bekbayeva A
S.Seifullin Kazakh Agro Technical University,
62 Zhenis Avenue, Nur-Sultan, 010111, Kazakhstan,

Summary

The results of the study illustrated that the simple linear regression model using the NDVI index produces better results in comparison to the multiple regression model using NDVI and LST (88% and 71% comparatively). Similarly, the RMSE also showed surpassed results of the simple linear regression model (4,15 to 6,78). Additionally, the fact that the implementation of fertilizers in predicted year for fallow fields has significantly increased wheat yield and influenced to the prediction accuracy. The study illustrates that the accuracy in statistical models with the use of several variables is distorted.The simple linear regression model might be applied to different crop types and territories.

Keywords vegetation index, NDVI, wheat yield prediction model, remote sensing, land surface temperature (LST), North Kazakhstan

ОЦЕНКА ИСПОЛЬЗОВАНИЯ ДАННЫХ ТЕМПЕРАТУРЫ ПОВЕРХНОСТИ И СПУТНИКОВОГО ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ДЛЯ МОДЕЛИРОВАНИЯ УРОЖАЙНОСТИ ПШЕНИЦЫ В СЕВЕРОМ КАЗАХСТАНЕ

Бекбаева А.
Казахский Агро Технический университет им. С. Сейфуллина, Пр. Женис 62, Нур-Султан, 010111, Казахстан,

Резюме

Результаты исследования показали, что простая линейная регрессионная модель с использованием индекса NDVI дает лучшие результаты по сравнению с моделью множественной регрессии с использованием NDVI и LST (88% и 71% соответственно). Аналогичным образом, среднеквадратическое отклонение также показало превосходящие результаты простой линейной регрессионной модели (4,15 и 6,78 соответственно). Кроме того, факт внесения удобрений в прогнозируемый год для паровых полей значительно повысило урожайность пшеницы и повлияло на точность прогноза. Исследование показывает, что точность статистических моделей с использованием нескольких переменных искажается. Простая линейная регрессионная модель может быть применена к различным типам культур и территориям.

Ключевые слова вегетационный индекс, NDVI, модель прогнозирования урожайности пшеницы, дистанционное зондирование, температура поверхности Земли (LST), Северный Казахстан

СОЛТҮСТІК ҚАЗАҚСТАН ЖАҒДАЙЫНДА БИДАЙДЫҢ ӨНІМДІЛІГІН МОДЕЛЬДЕУ ҮШІН ЖЕР БЕТІ ТЕМПЕРАТУРАСЫНЫҢ ЖӘНЕ СПУТНИКТІК ҚАШЫҚТЫҚТАН ЗОНДТАУ ДЕРЕКТЕРІН ПАЙДАЛАНА ОТЫРЫП БАҒАЛАУ

Bekbayeva A.
С.Сейфуллин атындағы Қазақ агротехникалық университеті, Жеңіс даңғылы 62, Нұр-Сұлтан, 010111, Қазақстан,

Түйін

Зерттеу нәтижелері көрсеткендей, NDVI индексін пайдалана отырып жасаған, қарапайым сызықтық регрессиялық моделі NDVI және LST индекстерін қолдана отырыпжасаған көптік регрессиялықмоделімен

салыстырғанда өте жақсы нәтижелер берді (сәйкесінше 88% және 71%). Осыған орай, орташа квадраттық ауытқу, қарапайым сызықтық регрессиялық моделінің нәтижелерінен әлдеқайда асып түсті(тиісінше 4,15 және 6,78).Сонымен қатар, болжалынып отырған жылы, сүрі жер алқаптары үшін тыңайтқыш енгізу бидайдың өнімділігін едәуір арттырып, болжамның дәлдігіне септігін тигізді.Зерттеу нәтижелері бойынша, бірнеше айнымалыларды қолданған кезде статистикалық модельдердің дәлдігі бұрмаланады. Қарапайым сызықтық регрессиялық модель әртүрлі дақылдар түрлеріне және әртүрлі аумақтарға қолданылуы мүмкін.

Кілттік сөздервегетациялық индекс, NDVI, бидай өнімділігін болжау моделі, қашықтықтан зондтау, жер бетінің температурасы (LST), Солтүстік Қазақстан

ACKNOWLEDGMENT

This work was carried out under the scientific project “Transfer and adaptation of precision agriculture technologies in the production of crop products on the principle of"demonstrativefarms (fields)" in the North Kazakhstan region” of S. Seifullin Kazakh Agro Technical University, Kazakhstan. I express my gratitude LLP “Agricultural experimental station” for your cooperation and assistance in data collection.