Vegetative indices dynamics modeling based on multispectral distant monitoring data analysis

K E Tokarev 1,2
1 Volgograd State Agricultural University, 26, Universitetskiy Avenue, Volgograd, 400002, Russia
2 Volgograd State Technical University, 28, Lenina Avenue, Volgograd, 400005, Russia
E-mail: tke.vgsha@mail.ru

Abstract. The article deals with the problems of monitoring the dynamics of vegetation indices on the basis of multispectral data obtained using technical means of remote monitoring. The effectiveness and validity of the application of vegetation indices was proved by comparing the values of plant indices with reference measurements at test sites of various types of plantings. A review of foreign sources and the results of statistical analysis allowed us to conclude that there is a non-linear relationship between productivity indices and crop yields.

1. Introduction
Agriculture is a major industry and it is vital to the economy of many countries. Due to the importance of agriculture there has been a wide variety of studies conducted in this field. The purpose of this literature review is to justify the research objectives of this study in light of previous work by investigating past crop yield prediction studies and determining where further research is needed. Literatures related to remote sensing, prairie climate and the various types of machine learning methods used in this study are reviewed. Machine learning theory and past applications are examined to show that these methods will likely perform well in a crop forecasting application. The influence of this material on the field of crop yield forecasting will be discussed and how it can be applied to this specific project is explained.

The NDVI theory is based on the fact that living plants absorb solar radiation in the range of photosynthetically active radiation, which they use as an energy source in the process of photosynthesis. Plants have also evolved to diffuse solar radiation in the near-infrared region of the spectrum, and so living plants appear relatively dark in the paired and relatively bright in the near-infrared region of the spectrum.

The effectiveness and validity of the plant indices were demonstrated by comparing the plant indices with the measured test sites, which were different types of plantings. An analysis of foreign studies has shown the suitability of using vegetation indices for agricultural monitoring and found that the results are very similar to the actual sowing conditions [1-3].

2. Materials and methods
The effectiveness and validity of the plant indices were demonstrated by comparing the plant indices with the measured test sites, which were different types of plantings. An analysis of foreign studies has shown the suitability of using vegetation indices for agricultural monitoring and found that the results are very similar to the actual sowing conditions [4-6].

The vegetation index data set under study is presented as a time series of composite values for NDVI and composite values for EVI covering the growing season. However, the lead time should still be long enough for the forecast to be useful, and therefore, to ensure a significant lead time, only data up to August were considered as predictors.

Figure 1a shows a preliminary correlation analysis between the yield of each crop and the smoothed MODIS-EVI, MODIS-NDVI, and NOAA-NDVI time series for the growing season.
The product term MODIS-EVI MODIS NDVI was considered as input to multiple linear regression models, and a correlation between this cross-term and each culture studied is also shown [7].

Figure 1. Examples of statistical analysis of vegetation index parameters

The correlation between MODIS-AVI and MODIS-NDVI is shown in Figure 1, where the correlation drops around week 27-30. Thus, at this stage of the growing season, EVI and NDVI can provide different information to the model.

The data sets studied were time series for each climate index, covering the period from December of the previous year to the end of the year. Due to the large fluctuation periods and the scale of these climate indices, the data was smoothed to a 3-month moving average and a similar correlation analysis was performed between each climate index and the yield of various crops.

Based on the correlation analysis, it was clear that climate indices do not show much potential for predicting crop yields.

For further study, the set of statistical data on crop yields was expanded and a correlation analysis was performed over a longer period. The results of the correlation analysis are shown in Figure 2.

Figure 2. Visualization of correlation analysis results.

After analyzing the results of the correlation analysis, it can be concluded that there is a non-linear relationship between some climate indices and crop yields, and to confirm that none of the climate indices will be useful as predictor variables, some preliminary non-linear models were run using different climate indices as predictors, and it was found that these models have almost no skills.

3. Results and discussion

Using the MATLAB mathematical package, multiple linear regression predictive models were developed using various combinations of predictors. During training, the multiple linear regression model finds a
linear relationship between the target data and multiple predictor variables. The model was trained using a leave one year out cross-validation scheme.

Example of a multiple linear regression model with several different combinations of predictors is shown in Table 1. Cross-terms are used in regression equations to reflect the interaction between two variables, and adding this term may allow the model to obtain additional information from vegetation indices [7].

**Table 1.** Predictor combinations for multiple linear regression model.

| MLR model | Predictor 1 | Predictor 2 | Predictor 3 |
|-----------|-------------|-------------|-------------|
| 1         | EVI         |             |             |
| 2         | MODIS-NDVI  |             |             |
| 3         | AVHRR-NDVI  |             |             |
| 4         | EVI         | MODIS-NDVI  |             |
| 5         | EVI         | MODIS-NDVI  | EVI x MODIS-NDVI |

The model-based recursive partitioning package is a non-linear tree-based method. The model-based recursive partitioning model splits the data into sections, like a standard Cart model, but then a linear model is developed for the data in each section. The main steps are: - fitting the parametric model to the data set, - checking the instability of the parameters against the set of partitioning variables, - if there is some general instability of the parameters, split the model with respect to the variable associated with the greatest instability, - repeat the procedure in each of the child nodes [7-12].

**Table 2.** Regression and partitioning variable combinations for MOB.

| MOB model | Predictor 1 | Predictor 2 | Partitioning Variable 1 | Partitioning Variable 2 |
|-----------|-------------|-------------|-------------------------|-------------------------|
| 1         | EVI         |             | EVI                     |                         |
| 2         | MODIS-NDVI  |             | MODIS-NDVI              |                         |
| 3         | AVHRR-NDVI  |             | AVHRR-NDVI              |                         |
| 4         | EVI         | MODIS-NDVI  | EVI                     | MODIS-NDVI              |
| 5         | EVI         | MODIS-NDVI  | MODIS-NDVI              |                         |
| 6         | MODIS-NDVI  |             | EVI                     | MODIS-NDVI              |

The model-based recursive partitioning requires a hyperparameter called minsplit that gives the minimum number of data points that can be in any terminal node. For this application the values considered for clusterization were 10%, 20%, 30%, 40% and 50% of the training data [7, 13-15]. The MOB model was developed using various different combinations of regression and partitioning variables as shown in table 2 [7].

**4. Conclusion**

The results of the study showed that the use of correlation analysis to select the optimal time during the growing season is the preferred tool for predicting crop yields and determining predictor variables. Model-based multiple linear regression and recursive partitioning prediction models were developed using different sets of predictors, and each model type was trained on data from individual clusters defined by hierarchical clustering and agroclimatic zone schema.

**5. Acknowledgments**

The research was conducted as a part of the grant of the President of the Russian Federation MK 592.2020.11.

**References**

[1] Zeileis A, Hothorn T and Hornik K 2008 Model-based recursive partitioning *Journal of Computational and Graphical Statistics* 17 492-514

[2] Atkinson P M and Tatnall A R L 1997 Neural networks in remote sensing. *International Journal of Remote Sensing* 18(4) 699-709
[3] Blackburn G A and Milton E 1997 An ecological survey of deciduous woodlands using airborne remote sensing and geographical information system (GIS). *International Journal of Remote Sensing* **18**(9) 1919-1935
[4] Célallos J C and Bottino M J 1997 Technical note: The discrimination of scenes by principal components analysis of multi-spectral imagery *International Journal of Remote Sensing* **18**(11) 2437-2449
[5] Huete A, Justice C and Van Leeuwen W 1999 Modis vegetation index (MOD13): Algorithm theoretical basis document *USGS Land Process Distributed Active Archive Center* 129 p
[6] Fensholt R and Proud S R 2011 Evaluation of Earth observation based global long term vegetation trends – Comparing GIMMS and MODIS global NDVI time series *Remote Sensing of Environment* **119** 131–147
[7] Johnson M D, Hsieh W W, Cannon A J, Davidson A and Bédard F 2016 Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods *Agricultural and Forest Meteorology* **218-219** 74-84
[8] Garge N R, Bobashev G and Eggleston B 2013 Random forest methodology for model-based recursive partitioning: the mobForest package for R *BMC Bioinformatics* **14** 125
[9] Chang D-H and Islam S 2000 Estimation of soil physical properties using remote sensing and artificial neural network *Remote Sensing of the Environment* **74**(3) 534-544
[10] Earl R, Wheeler P N, Blackmore B S and Godwin R J 1996 Precision farming – The management of variability *Landwards* **51**(4) 18-23
[11] Mair C, Kadoda G, Lefley M, Phalp K, Schofield C, Shepperd M and Webster S 2000 An investigation of machine learning based prediction systems. *Journal of Systems and Software* **53**(1) 23-29
[12] Osborne S L, Schepers J S, Francis D D and Schlernrner M R 2002 Use of spectral radiance to estimate in-season biomass and grain yield in nitrogen- and water-stressed corn *Crop Science* **42** 165-171
[13] K E Tokarev 2021 *J. Phys.: Conf. Ser.* **1801** 012030
[14] Plant R E, Munk D S, Roberts B R, Vargas R L, Rains D W, Travis R L and Hutmacher R B 2000 Relationship between remotely sensed reflectance data and cotton growth and yield *Transaction of the ASAE* **43**(3) 535-546
[15] K E Tokarev 2021 *J. Phys.: Conf. Ser.* **1801** 012031