Dynamic model of crops' normalized difference vegetation index in central federal district environment

A G Bukhovets, E A Semin, M V Kucherenko and S I Yablonovskaya

Voronezh State Agricultural University named after Emperor Peter the Great.
1, Michurina Street, Voronezh, 394087, Russian Federation

E-mail: abuhovets@mail.ru

Abstract. The article describes a mathematical model which represents dynamics of the normalized difference vegetation index for winter wheat plantings in central black soil areas. As opposed to the approaches considered during our theoretical study, which, as a rule, are based on averaged data related to respectively vast territories (districts, regions, agricultural enterprises), this article describes a model which relates to rather small areas, namely to particular fields measuring 30-200 ha. The multiplicative model under consideration takes into account two opposite tendencies in the development of winter wheat: the process of phytomass increase and the process of plastic substances production. Parameters of the suggested model were estimated for winter wheat plantings in central black soil area on the fields with different levels of productivity for 2017 in accordance with normalized difference vegetation index data. We estimated the parameters by least square method. We performed model functional tests on the basis of the data received during remote sounding of the soil at more than one hundred of fields in Central Federal District. The test results are very promising. The suggested model allows for estimation of ripening period and the time of harvesting. The model can be applied for approximation of normalized difference vegetation index missing values, as well as for estimation of time required to attain maximum index value and, consequently, for forecasting of harvesting terms.

1. Introduction

There is a growing tendency of satellite technologies application in agriculture. Data received from space will provide agricultural producers with valuable information on land condition. Based on this information, a producer will decide on the necessity to make adjustments into the process of plant growth and development to increase their productivity.

Remote sounding is a method used to receive information on the object or phenomena without immediate physical contact. In the modern sense, the term usually refers to the technologies of space sounding of territories in order to detect, classify or analyse objects on earth surface using signal propagation [1, 2]. The advantage of using such technologies is the possibility to quickly receive credible information from vast territories with necessary periodicity.

Nowadays pictures taken from space allow for calculation of numerous parameters which characterize plant growth and development [8]. Such qualitative characteristics are called vegetation indices. One of such biomass parameters is NVDI (Normalized Difference Vegetation Index). It is based on the analysis of plant reflectivity in different spectrum areas. Plant reflectivity is determined by different pigments, water content and leaf structure. It allows to monitor the dynamics of phytomass.
accumulation during the whole vegetation period [5]. In many countries around the world it is used to monitor the condition and productivity of plantings. Despite the fact that nowadays the development of information technologies for processing and visualization of satellite information represents a relevant area of research [6, 7], some authors believe that the attempts to establish the connection between vegetation index and productivity of plants still have no considerable result [2, 7]. To sum up, modelling of vegetation index dynamics is a relevant issue, which has substantial practical importance.

2. Methods and information background

Key objective of our study was to develop a mathematical model allowing to describe the dynamic process of winter wheat plantings development, on the basis of NDVI statistical data. Having had analysed scientific articles on the subject (for example, [4]), we came to the conclusion that the connection between NDVI and plant growth characteristics is not direct and explicit. In our study, we proceeded from the hypothesis that dynamic parameters of vegetation index curve are connected with phytomass of agricultural plants.

In order to achieve the objective, we suggested the following mathematical model of vegetation index dynamics. Assume that \( Y(t) \) is the level of phytomass achieved by the time point \( t \) and estimated by the value of NDVI. Then \( Y'(t) \) is the rate of phytomass change at the time point \( t \). It is logical to assume that \( Y'(t) \) is proportionate to the achieved \( Y(t) \). Further, following the analysis of empirical data, we can observe that NDVI value not only stops growing over time, but, after reaching a certain point, begins to decrease. In this connection we introduced the following function to describe NDVI change rate:

\[
Y'(t) = Y(t) \cdot \left( \frac{b}{t} - a \right),
\]

where \( b \) is the parameter which determines the rate of phytomass (biomass) increase; \( a \) is the parameter used to characterize caryopsis growth rate, which slows down the phytomass growth.

The influence of \( a \) is insignificant in the period of plant development, as intensive growth of plant biomass conceals its effect. However, as the plant growth slows down, the influence of this parameter becomes more and more apparent. At the same time \( b/t \) value determines the vegetation component [7].

Having divided the variables in the equation (1), we will receive the following:

\[
\frac{dY}{Y} = \left( \frac{b}{t} - a \right) \cdot dt.
\]

Having integrated, then taken antilogarithm for this differential equation, we will receive the following correlation:

\[
Y(t) = C \cdot t^b \cdot e^{-at},
\]

where \( C \) is integration constant responsible for the scale of value \( Y(t) \) measurement; \( a \) and \( b \) are earlier introduced parameters which characterize the dynamic process of vegetation index change; \( t \) stands for time (in days).

Figure 1 represents the example of formulated model for NDVI values approximation using available statistical data. We shall note that the technologies of receiving satellite data are not faultless; one of the faults is intrinsic error, namely the occurrence of unwanted noise, which is the result of weather conditions, etc. This is the reason why the statistical data was processed before using: a part of the received values was filtered out.
Having analysed the bracketed mathematical expression in the formula (1), we can conclude that there shall be such a moment of time \( t^* \), in which both the opposite tendencies balance each other. Then, if \( t < t^* \), the process of biomass increase will prevail (NDVI value increases); if \( t^* < t \), the process of ripening, related to caryopsis growth, will prevail. Thus, we can assume that \( Y(t) \) attains its maximum value at the moment when \( t \) equals \( t^* = \frac{b}{a} \).

3. Results

Having taken the logarithm of the equation (3), we receive linear equation with respect to the parameters, the values of which were estimated by the least square method.

\[
\ln Y(t) = \ln C + b \cdot \ln t - a \cdot t
\]

The calculation results can be further applied in study of connection between particular fields productivity and dynamic function parameters \( C, a \text{ and } b \) value sets.

The study of connection between the vegetation index dynamic curve parameters and productivity comes down to comparing values of parameters with the observed values of productivity. For the analysis, we have selected fields with different level of productivity. For our purpose, we divided the levels of productivity into the following: high (more than 70 centners/ha), mid-level (from 27 to 41 centners/ha), low (less than 25 centners/ha). For each group we calculated the parameters of vegetation index dynamic model. Descriptive statistics of the received population are represented in Table 1.

Table 1. Descriptive statistics of vegetation index parameters array.

| Variable | Valid N | Mean | Confidence 95.000% | Confidence 95.000% | Minimum | Maximum | Std.Dev. |
|----------|---------|------|-------------------|-------------------|---------|---------|---------|
| Productivity | 105 | 51.52 | 48.52 | 54.52 | 25.00 | 97.01 | 15.52 |
| C         | 105 | 36.01 | 33.85 | 38.17 | 13.50 | 62.48 | 11.16 |
| a         | 105 | -0.06 | -0.06 | -0.05 | -0.10 | -0.02 | 0.02 |
| b         | 105 | 8.84  | 8.30  | 9.37  | 3.27  | 15.49 | 2.78  |

We will consider the set of variables \{Productivity, \( C, a, b \)\} to be a Euclidean space \( \mathbb{R}^4 \) with standard, unweighted metrics. All the variables are pre-standardized. We use methods of cluster analysis for analysing the array structure [3]. The input parameter of k-means clustering will be the amount of clusters. We assume it to be three. We will apply complete linkage for result visualization. Figure 2
displays that cluster division is stable on the levels from 4.5 to 6.0, which corresponds to division of the original population into three clusters.

Thus, the division of the original population into three clusters is justified by the results obtained during several statistical procedures: variance analysis and different algorithms of cluster analysis. All the results are statistically significant on the standard 5% level. They are well-confirmed by respective geometric plotting.

Figure 2. Dendrogram of complete linkage hierarchical cluster analysis.

In order to consider the specified clusters in forecasting productivity tasks, we will introduce dummy variables [9, p.58] in the following way:

\[
D_{n1} = \begin{cases} 
1, & \text{if the field belongs to cluster 1,} \\
0, & \text{otherwise,}
\end{cases}
\]

\[
D_{n2} = \begin{cases} 
1, & \text{if the field belongs to cluster 2,} \\
0, & \text{otherwise.}
\end{cases}
\]

With the consideration of the introduced regressor variables, we will build a Productivity- \{\text{\(C_{\text{st}}, a_n, b_{\text{st}}, D_{n1}, D_{n2}\)}\} regression model. As regression equation is represented in stochastic form of multicollinearity [9], we applied forward stepwise procedure. Final results are represented in table 2. Respective regression equation will be the following:

\[
\text{PRODUCTIVITY} = 68.66 - 33.86 \cdot C_{\text{st}} + 25.40 \cdot b_{\text{st}} - 21.62 \cdot D_{n1} - 23.32 \cdot D_{n2},
\]

(5)

Table 2. Regression Equation received as a result of forward stepwise procedure.
4. Discussion of results

As an example, we will calculate the estimated values of productivity basing on the equation (5). It is important to note that for our test, we will take vegetation index data for a field with the productivity of 46.8 centners/ha. The data was not used to determine the parameters of forecast equation. The calculation can be considered as check-up test of the suggested scheme. According to the data received during vegetation index measurement, the following estimations were deduced: $C=29.5$ and $b=-0.0043$. The estimated value of productivity will amount to 42.07 centners/ha. At the same time, 95% confidence interval will amount to (35.85 ÷ 48.28) and 90% confidence interval will amount to (36.86 ÷ 47.27) centners/ha. It is obvious that the received result demonstrates an agreement between the calculated and empirical data.

5. Conclusion

In this article we presented the analysis of NDVI values related to winter wheat growing on the fields of central black soil area for 2017. Based on statistical estimation of empirical data, we formulated the assumption on dynamic character of biological processes, associated with the vegetation index values. The model assumptions were tested on the basis of the statistical data, which allowed us to confirm the stated hypothesis on vegetation index formation mechanism. Those assumptions further provided the foundation for the mathematical model of NVDI dynamics. The suggested model clearly reflects specific features of winter wheat, as a period of green mass rapid growth and its gradual decrease in the ripening period. The vegetation index dynamic model can be applied for approximation of NDVI missing values, as well as for estimation of time required to attain maximum index value and, consequently, for forecasting of harvesting terms.

References

[1] Bukhovets A G 2018 Proc. Int. Conf. on Mathematical Modelling and Information Technologies of Engineering and Business Applications (Voronezh, Russia: VSAU) p 174-84
[2] Eroshenko F V, Bartalev S A, Storchak I G and Plotnikov D E 2016 Modern Problems of Remote Space Earth Sounding (Moscow, Russia: Space Research Institute) 13(4) 99-112
[3] Bukhovets A G, Semin E A, Kostenko E I and Yablonovskaya S I 2018 NVDI Dynamic Mathematical Modelling for Winter Wheat in the Conditions of Central Federal District Vestnik VSAU 57(2) 186-99
[4] Pisman T I, Botvich I Yu and Sidko A F 2014 Determining Seasonal Dynamics of Agrosystems on the basis of Satellite Information and Mathematical Models RAS Journal. Biological Series 2 196-202
[5] Semin E A and Bukhovets A G 2018 Relevant Issues of Agricultural Business Industry and Rural Territories (Voronezh, Russia: VSAU) p 39-43
[6] Modern Technologies of Remote Earth Sounding Data Processing 2015 ed V V Ereemeva (Moscow, Russia: Fizmatlit)
[7] Terekhin E A 2014 Analysis of Seasonal Dynamics of NDVI and Reflective Properties of Corn Plantings on the Territory of the Belgorod Region Modern Problems of Remote Space Earth Sounding 11(4) 244-53
[8] Cherepanov A S 2011 Vegetation Indices Geomatika 2 98-102
[9] Yanovsky L P and Bukhovets A G 2017 Introduction to Econometrics. Study Guide (Moscow, Russia: KNORUS)