Forecasting the winter wheat yield based on the vegetation index NDVI dynamic model

A G Bukhovets¹, E A Semin, M V Kucherenko and S I Iablonovskaia

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

¹E-mail: abuhovets@mail.ru

Abstract. The proposed vegetation index NDVI dynamic model allows us to move from describing the plants' development and growth phenomenological way to model ideas about the vegetation process as a whole. The biological research previously obtained experimental material is linked within a mathematical model framework with the vegetation indices values calculated from satellite images. This makes it possible to quantitatively evaluate the plant development characteristics and indicators in the crops observations framework carried out using spacecraft. This work main task is to explicate the mathematical model main provisions, which allows combining these two viewpoints obtained in different ways: experimentally and through visual analysis. This approach allows explaining how the processes occurring in plants are displayed in a data mathematical model obtained from space. In this paper, we consider an approach based solely on a dynamic model parameter for the current field season and an introduced new integral characteristic that depends on the vegetation index NDVI current values. This approach advantage to yield forecasting is that the forecast is made on the data basis coming from the spacecraft, i.e. is operational. At the same time, the forecast is based on a dynamic model that takes into account the agricultural plants physiological characteristics, in particular, winter wheat.

1. Introduction

The plant organs photosynthetic elements large contribution to the crop formation is known. Works in which these winter wheat organs (leaves, shoots, inflorescences, etc.) photosynthetic elements relationship was investigated with productivity were studied in [1]. However, as it is known, simultaneously with the green mass formation, there is a creating plastic substances process, the main material for creating grains and dry mass. The transition from the plants' photosynthetic elements' development models to the photosynthetic sowing surface as a whole is currently not a final solution either. This problem has become especially urgent in connection with the Earth remote sensing data use. The main indicator is usually the vegetation index NDVI, which, as it is known, is based precisely on the green crops' chlorophyll nature taken in the spectral range red part different areas. The growing green mass process is reflected in the Earth's surfaces images, received by space vehicles, in particular, of agricultural land massifs [2, 3]. The vegetative development NDVI index is based on the difference ratio between the spectrum red and infrared regions intensities to the same intensities sum. The grain crops assessing the state method based on NDVI readings are widely used in domestic and foreign practice [4-6].
Previously, we proposed a vegetation index NDVI dynamics mathematical model [7], which reflects the winter wheat plants development peculiarities, referring to relatively small areas - separate fields of 30-200 hectares in the Central Chernozem region. In this paper, an approach based solely on the model parameters for the current season and a new integral characteristic introduction used in forecasting the yield is considered.

2. Biological and physiological aspects of vegetation index modelling the dynamics

Previously, photosynthetic surface dynamic models were compiled for a wheat plant each vegetative organ [8]. In the future, we will consider NDVI as a generalized (integral) photosynthetic surface indicator without dividing and taking into account each vegetative organ contribution. In other words, we will consider sowing as a system that unites plants set, which properties largely determine the system properties as a whole. These properties are then reflected in the obtained empirical data.

The initial data consisted of NDVI values arrays for winter wheat crops in the Central Federal District, averaged according to the measurement results in the period from the vegetation renewal to maturation (March-August) for 2016–2017. These data were obtained using the MODIS spacecraft with a 250 m resolution. At the same time, the initial data were cleared of interfering factors (cloudiness, haze, etc.) using special masks [9].

In the constructing course and further analysing the differential vegetation index NDVI dynamics mathematical model, we assumed that the following basic assumptions take place:

- The plastic substances’ production process in plants through the photosynthesis process plays a decisive role in creating the future harvest [8]. This process basis is the so-called. photosynthetic surfaces.
- Within our model framework, we will consider the winter wheat sowing as a homogeneous environment, presented in the satellite images. We will assume that the plants’ totality in the fields is an integral formation of some kind, i.e. as a system object.
- We will consider the dynamic model NDVI parameters values as random variables. The photosynthetic surface size is positively correlated, as shown by the calculations carried out [4], with the vegetative index NDVI values.
- The function \( Y = Y(t) \) introduced into consideration by us describes the NDVI index dynamics [9]. Model parameters a and b characterize the growth and vegetation rates of both the plant itself and its organs. The parameters’ values a and b reflect all factors integral influence in individual plant development, as well as the culture biological properties: its genotype, the agrometeorological and geophysical conditions influence. These factors’ ratio is determined by the growing season, as well as by the time \( t_{max} \), at which the photosynthesizing surface area reaches its maximum value in the heading phase. We will use in our model the vegetation index NDVI numerical values, which were obtained during remote sensing [7].

3. Mathematical aspects of constructing the NDVI difference index model

Here are additional assumptions related to mathematical apparatus use.

- Thus, it is additionally assumed that the function \( Y(t) \) describing the vegetation index dynamics is not only a continuous function but also a differentiable function.
- The NDVI index value, as well as the photosynthetic surface magnitude, changes in proportion to the level already reached and depends on a certain function \( \varphi(t) \), i.e.
  \[
  Y'(t) = Y(t)\varphi(t)
  \]
  (1)
- The entered function \( \varphi(t) \) represents nothing more than the change relative rate in the NDVI vegetative index values. In other words
\[
\frac{Y'(t)}{Y(t)} = \varphi(t)
\]  

(2)

- Let's define the function \( \varphi(t) = \left( \frac{b}{t} - a \right) \). In this case, the function of the \( Y(t) \), which is the differential equation solution (1), will take the following kind

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

(3)

where the parameter \( C \) acts as a rationing multiplier, and the parameters' interpretation \( a \) and \( b \) are given in [7]. The built function graphic image is quite consistent with the available experimental data nature. The model parameters evaluation by experimental data was carried out by the smallest squares method [10]. Built with such assumptions, the growing index dynamics models, corresponding to different fields, have a good quality of fit the determinism factor is within 0.7÷0.85.

4. The NDVI dynamic model predictive properties

The yields \( U \) statistical relationship analysis with NDVI values will begin with the time when the vegetative index reaches its maximum value. To do this, it is enough to solve the equation \( Y'(t) = 0 \). It is easy to show that the point that defines this moment depends only on the parameters' ratio \( b \) and \( a \), i.e.

\[
t_{\text{max}} = \frac{b}{a}.
\]

Using a ratio (3), you can determine the timing of when the growing index growth rate changes. To do this \( Y''(t) = 0 \), you need to consider the equation, which roots will be the values \( t_1 = \left( b - \sqrt{b} \right)/a \) and \( t_2 = \left( b + \sqrt{b} \right)/a \). It's obvious that the values \( t_1 \) and \( t_2 \) are symmetrical relative \( t_{\text{max}} \), but at the same time \( Y(t_2) > Y(t_1) \).

Let's introduce a value \( S(a, b) \), numerically equal to the function integral relative at the interval \( Y(t) = C \cdot t^b \cdot e^{-at} \) to this interval \( [t_1, t_2] \) size, i.e.

\[
S(a, b) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} Y(t) dt = \frac{C}{t_2 - t_1} \int_{t_1}^{t_2} t^b e^{-at} dt.
\]  

(4)

The resulting value \( S(a, b) \) can be considered an average function \( Y(t) \) of the integration interval, or the NDVI dynamic model integral indicator average. This value is proportional to the total green mass, and therefore dry. In the future, for brevity, let's call the value \( S = S(a, b) \) an integral indicator. Figure 1 shows the NDVI dynamic model curve, the integration interval and the area under the integral curve corresponding to the mean value \( S \).

This interval choice is because the growing index average values, as seen from (4), depend, except on the parameters of \( b \) and \( a \), and also on the integration limits. To be able to compare the values corresponding to different fields \( S(a, b) \), limits were chosen that depended on the growing process characteristics, rather than on external factors, such as cleaning time.

Table 1 provides descriptive statistics of parameters of \( U, C, a, b \) and \( S \), and the same values correlations. As you can see, the parameters are \( a, b \) and \( C \), which are associated with plant growth, i.e. are intense characteristics, do not correlate, at least significantly, with the yield \( U \). At the same time, as can be seen from table 1, the parameters of the \( C, a \) and \( b \) are strongly related and correlate with each other and with the parameter \( S \). In turn, the parameter \( S \) correlates with the values of \( U \). Thus, the newly
introduced parameter of the $S$ is an intermediate parameter between the dynamic model characteristics and the yield of the $U$.

Figure 1. The value $S(a, b)$ represents the NDVI vegetation index average value in the interval $[t_1, t_2]$.

Table 1. Narrative parameters statistics and the NDVI dynamic model parameters correlations.

| Variable          | Correlations (Spreadsheet3 sta) Marked correlations are significant at $p < .05000$ |
|-------------------|-----------------------------------------------------------------------------------------|
|                   | N=105 (Casewise deletion of missing data)                                               |
|                   | Means | Std.Dev | U-Productiv | C        | a        | b        | S Integrat. |
| U-Productiv       | 51.520 | 15.523  | 1.0000      | -0.0611  | -0.0468  | -0.0024  | 0.4376      |
| C                 | 36.011 | 11.155  | -0.0611     | 1.0000   | 0.9985   | -0.9830  | 0.4810      |
| a                 | 8.835  | 2.776   | -0.0468     | 0.9985   | 1.0000   | -0.9898  | 0.4903      |
| b                 | -0.056 | 0.018   | -0.0024     | -0.9830  | -0.9898  | 1.0000   | -0.5221     |
| S - Integrated    | 0.682  | 0.052   | 0.4376      | 0.4810   | 0.4903   | -0.5221  | 1.0000      |

Further studies have found that the two variables $S$ - the average integral value and the yields $U$ - yields, as evidenced by the QQ-graphs, have the same distributions [11].

5. Results and discussion
The assumption is confirmed that the yield as a system indicator is integral, and many random factors that are not taken into account directly in the model affect this indicator value. The built model allows estimating a random factors combination contribution value not included in the model (3).

The integral indicator we have introduced, due to the close relationship with the yield indicator, can be used to predict the yields' size. The resulting equation will have the appearance of:

$$U = -37.06 + 129.83 \cdot S, \quad (R^2 = 0.19) \quad (5)$$

Equation (5) will be used for forecasting. Let, for example, the magnitude $S$ takes a value of 0.75. Then the forecast value $U$ will be 60.31 c/h, 95% confidence interval will be ($55.86 \div 64.76$); 90% confidence interval ($56.58 \div 64.03$). Other forecast calculation results and corresponding confidence intervals for 90% and 95% reliability can be seen in table 2.

The regression equation that determines the dependence of the yield ($U$) parameters $C$, $a$ and $b$ has $R^2=0.122$ ($R^2_{adj}=0.095$), while the paired regression of yield ($U$) versus $S$ has $R^2=0.192$ ($R^2_{adj}=0.184$).
These results support the claim that the integral indicator is more closely related to $S$ and the yield $U$ than with the model intensive characteristics.

### Table 2. Results from the average yield forecasts calculation by the integral index size.

| № | $S$ (integral indicator) | $U$ (yield) (forecast) | 90% reliability level $U_{bottom}$ | 90% reliability level $U_{top}$ | 95% reliability level $U_{bottom}$ | 95% reliability level $U_{top}$ |
|---|-------------------------|------------------------|----------------------------------|---------------------------------|-----------------------------------|----------------------------------|
| 1 | 0.60                    | 40.83                  | 36.59                            | 45.08                           | 35.76                             | 45.91                            |
| 2 | 0.65                    | 47.33                  | 44.65                            | 50.00                           | 44.13                             | 50.52                            |
| 3 | 0.70                    | 53.82                  | 51.42                            | 56.22                           | 50.95                             | 56.68                            |
| 4 | 0.75                    | 60.31                  | 56.58                            | 64.03                           | 55.86                             | 64.76                            |
| 5 | 0.80                    | 66.80                  | 61.19                            | 72.41                           | 60.09                             | 73.51                            |
| 6 | 0.85                    | 73.29                  | 65.63                            | 80.95                           | 64.14                             | 82.44                            |

6. Conclusion

The integral indicator $S$ presented in the article, which is obtained on the vegetation index NDVI dynamic model basis, can be used to predict the winter wheat yield. Unlike the previously proposed forecasting approaches [7], this approach does not use (dummy) variables and does not require the multicollinearity removal procedures implementation [10]. This approach advantage to yield forecasting is that the forecast is made on the data basis coming from the spacecraft, i.e. is operational. At the same time, the forecast is based on a dynamic model, which, as it was shown earlier, takes into account the crops physiological characteristics, in particular, winter wheat.

### References

[1] Andrianova Yu E and Tarczewski I A 2000 *Chlorophyll and Plant Productivity* (Moscow, Russia: Science) 134

[2] Stepanov A S 2019 Forecasting the yield of crops based on remote sensing data of the Earth (for example, soybeans) *Computational technologies* 24(26) 125-32

[3] Storchak I G, Eroshenko F V and Shestakova E O 2019 The vegetation index NDVI dynamics features in different soil and the Stavropol Territory climatic zones *Urals agrarian Bulletin* 9(188) 12-8

[4] Eroshenko F V, Bartalev S A, Storchak I G and Plotnikov D E 2016 The winter wheat yield remote assessment possibilities based on the photosynthetic potential vegetation index *Modern problems of remote sensing of the Earth from space* 13(4) 99-112

[5] Eroshenko F V, Storchak I G, Engovatova I V and Likhovid N G 2020 Using Earth remote sensing data for winter wheat grain quality regional assessment *Interkarto.Intergis* 26(3) 240-50

[6] Zinchenko V E, et al. 2013 Agricultural lands space monitoring in the Russia south *Research of the Earth from space* 3 33-44

[7] Bukhovets A G, Semin E A, Kucherenko M V and Yablonskaya S I 2020 Dynamic model of crops’ normalized difference vegetation index in central federal district environment *IOP Conf. Series: Earth and Environmental Science* 548 042019 DOI: 10.1088/1755-1315/548/4/042019

[8] Shatilin I S and Zamaarev A M 1987 Winter wheat sowing photosynthetic activity mathematical model *TAA News* 2 31-9

[9] Bukhovets A G, Semin E A, Kostenko E I and Yablonskaya S I 2018 Modeling the winter wheat vegetation index NDVIdynamics in the Central Federal District *Voronezh State Agrarian University Bulletin* 2(57) 186-99

[10] Yanovsky L P and Bukhovets A G 2017 *Introduction to econometrics* (Moscow, Russia: KNORUS) 256

[11] Khalafyan A A 2010 *Statistica 6 Statistical data analysis* (Moscow, Russia: LLC "Binom-Press") 528