AGRICULTURAL BIOLOGY,
2019, V. 54, № 1, pp. 84-90
(SEL'SKOKHOZYAISTVENNAYA BIOLOGIYA)

ISSN 2412-0324 (English ed. Online)
ISSN 0131-6397 (Russian ed. Print)
ISSN 2313-4836 (Russian ed. Online)

UDC 633: 57.087: 51-76

doi: 10.15389/agrobiology.2019.1.84eng
doi: 10.15389/agrobiology.2019.1.84rus

EXPERIENCE WITH THE USE OF MATHEMATICAL STATISTICS METHODS FOR ASSESSMENT OF AGRICULTURAL PLANTS STATUS

V.M. BURE[1, 2], A.F. PETRUSHIN[1], E.P. MITROFANOV[1], O.A. MITROFANOVA[1, 2], V. DENISOV[3]

[1]*Agrophysical Research Institute*, 14, Grazhdanskii prosp., St. Petersburg, 195220 Russia, e-mail vlb310154@gmail.com (✉ corresponding author), apetrushin@agrophys.ru, mjeka@agrophys.ru, omitrofa@gmail.com;
[2]*Saint Petersburg State University,* 7/9, Universitetskaya nab., St. Petersburg, 199034 Russia;
[3]*Klaipeda University*, Herkaus Manto 84, LT-92294 Klaipeda, Lithuania, e-mail vitalij.denisov@ku.lt
ORCID:
Bure V.M. orcid.org/0000-0001-7018-4667
Petrushin A.F. orcid.org/0000-0002-6482-8611
Mitrofanov E.P. orcid.org/0000-0002-1967-5126
Mitrofanova O.A. orcid.org/0000-0002-7059-4727
Denisov V. orcid.org/0000-0002-0325-1193
The authors declare no conflict of interests
Received October 22, 2018

Abstract

Solving problems related to the assessment of the status of agricultural plants during the growing season, allows us to effectively use fertilizers, obtain favorable yields, improve the quality characteristics of plants, as well as the ecological condition of the field. To solve such problems of precision farming, the use of various methods of mathematical statistics is becoming an increasingly promising direction. The aim of our work was to assess the state of agricultural plants using an approach based on the combined use of kriging and binary regression methods, as well as the determination of nitrogen planting using the NDVI (Normalized Difference Vegetation Index) index. The studies were carried out at the site of an experimental agricultural field located on the territory of the branch of the Agrophysical Institute (Menkovo, Leningrad region) in 2015. With the help of aerial photographs taken from the automatized unmanned aerial vehicle complex Geoscan-401 (Geoscan Group of Companies, Russia), a set of NDVI (Normalized Difference Vegetation Index) vegetation index values was obtained at arbitrary points of the plot. A number of ground-based measurements were also conducted on the studied area of the field. The proposed approach to assessing the state of agricultural plants consisted in the joint use of two methods of mathematical statistics: ordinary kriging and logistic regression. A preliminary variogram analysis was carried out, and a variogram model was constructed. After this, the kriging method was used to calculate a series of predicted values of the parameter being studied. At the next stage, the threshold value of the parameter for the study area was established, and also a dummy variable was entered, taking the value 1 if the parameter value exceeded the threshold, and 0 otherwise. Then a logit model was built, in which one of the factors was a series of estimates of the parameter of interest, obtained using the ordinary kriging method. The input data for building logit models were as follows: $N(x_i)$ is the NDVI value at the location x_i, $i = \overline{1.78}$; variable $T = 1$, if $N(x_i) \geq 0.46$, otherwise $T = 0$; the variables X and Y are the coordinates of the observations, are considered as explanatory variables; $N_{pred}(x_i)$ is parameter values, predicted using the kriging method at the observed points. All calculations were performed using the R programming language. As a result of the experiment, three logit models were built with the dependent variable T: in the first model, the explanatory variables X and Y; in the second model — X, Y and N_{pred}; in the third model N_{pred}. Testing showed that when adding the N_{pred} variable, the logit model works better (2 times less than the erroneous determination of the level of the parameter under study). The results obtained suggest that adding in the binary regression factors a set of values predicted by the kriging method can significantly improve the accuracy of calculations.

Keywords: plant status, Normalized Difference Vegetation Index, NDVI, kriging, binary regression, language R

Evaluating state of crops during the growing season (availability of nutrients, watering parameters, weeds, diseases, etc.) is necessary for using fertilizers efficiently and producing a great and high-quality yield of [1-3]. In recent years, statistical testing and remote sensing data processing are becoming increasingly more effective ways to address these challenges [4-6].

One of the new approaches in agrophysics is based on binary regression methods. Thus, Bure [7] describes the application of binary regression to yield forecasting. Norwegian and Dutch scientists have proposed methods for predicting the spatial distribution of soil types by means of multinomial logistic regression using digital terrain analysis [8, 9]. More sophisticated and advanced areas of precision agriculture include geostatistics, which helps map the soil content of nutrients (nitrogen, phosphorus, potassium, etc.) [10, 11], and soil electrical conductivity, pH, density, and humidity estimates [12-14] used to optimize land management. In the geostatistical approach, the soil is treated as a set of spatially continuous variables, changes wherein are described in terms of spatial dependency [15, 16]. It is only economics that combines geostatistics and binary regression methods [17]; no such methodology has yet been described in detail in relation to precision agriculture.

This paper is the first to predict the spatial distribution of the Normalized Difference Vegetation Index (NDVI) in an experimental field using Gaussian process regression (kriging) in combination with logistic regression as a subtype of binary regression. Test results show that the proposed approach allows a sufficiently accurate evaluation of the test-site parameter of interest.

The goal was to characterize the condition of crops by combining kriging and binary regression, as well as to find the availability of nitrogen to crops in terms of NDVI.

Techniques. Studies were carried out in 2015 (a test field of the Institute of Agrophysics in Menkovo, Leningrad Province). Aerial photographs taken by a Geoskan-401 (Geoskan, Russia) unmanned aircraft were used to obtain the Normalized Difference Vegetation Index (NDVI) values at arbitrary points, 78 in total.

The condition of agricultural plants was evaluated by a combination of two methods of mathematical statistics: ordinary kriging and logistic regression. A logit model was used as an approach that enabled simple parametric evaluation.

The spatial distribution of the parameter of interest was predicted by ordinary kriging for a set of measurements [18, 19]:

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i), \quad \sum_{i=1}^{n} \lambda_i = 1, \tag{1}$$

where n is the number of observations, $Z(x_i)$ is the value of the observed parameter at the location x_i, λ_i is the unknown weight for the parameter, $\hat{Z}(x_0)$ is the parameter value predicted for the location x_0.

The unknown weight was found by variogram analysis and constructing a theoretical variogram model $\gamma(h)$ based on the obtained experimental curve $\gamma(\hat{h})$.

To run a logistic regression, the value d (the threshold) was recorded for the test site and a dummy variable was inserted:

$$y(x) = \begin{cases} 1, Z(x) \geq d, \\ 0, Z(x) < d. \end{cases}$$

The values of $y(x_i)$ were known for the observed points, as the point-specific values of the parameter of interest, as well as its level in relation to the threshold, were known. At the preceding stage, a set of kriging-predicted values of the parameter was produced. This gave a set of inputs for logistic regression that would reflect that correlation between the threshold exceedance probability and the explanatory variables [20, 21]:

$$P(y(x_i) = 1 \mid \varphi_i) = p_i = \frac{1}{1 + \exp(-\varphi_i^T \beta)}, \tag{2}$$

where φ_i are the factors that explain the dummy variable $y(x_i)$.

The set of kriging-predicted values was used as one of the factors in the logit model. The adequacy of the build logit model [2] was tested by the classical statistical tests i.e. the Walt test, W, and the likelihood ratio test, LR [22].

For each point of the test site, the research team computed the probability $P(y(x) = 1)$, which, when tending to 1 indicates that the parameter of interest exceeds the threshold d, while when tending to 0, it indicates that the parameter is below the threshold.

Calculations were run in R software (https://www.r-project.org), which is a popular solution used in precision agriculture [23].

Results. Fig. 1 presents an aerial photograph of the test field, as well as a data distribution map (circle parameters are proportional to the original values) made in R-statistics. The result is a set of NDVI values, which are known to correlate with the point-specific in-plant nitrogen content [24, 25].

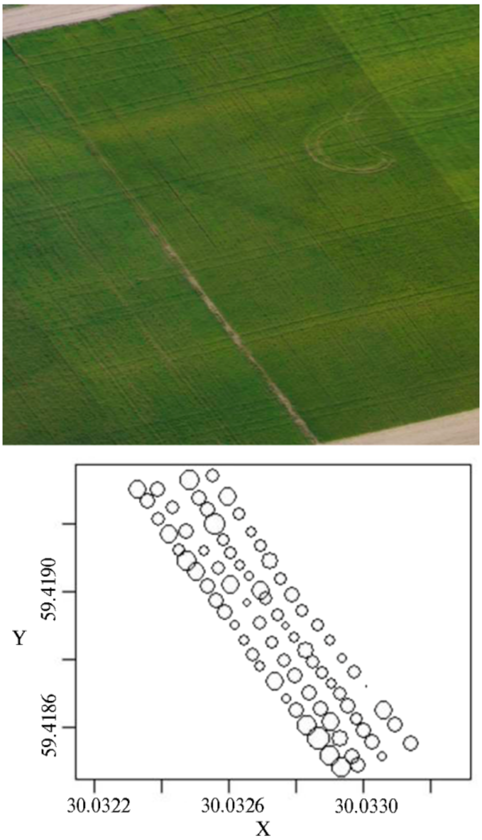

Fig. 1. Aerial view of the test site and observation distribution map (map location)**:** X and Y are the observation coordinates. The diameters of the circles are proportional to the value of the analyzed indicator, NDVI, as measured in Menkovo, Leningrad Province, 2015.

The approach proposed herein is to use a set of kriging-predicted values as one of the model factors. Accordingly, the first stage (predicting the spatial distribution of the parameter of interest) was to check whether the geostatistical conditions of stationarity and multinormality are held [26]. The detected outliers were cut at 2.5% bilateral quantiles. Besides, the research team would evaluate the linear correlation of the parameter with the coordinates. No spatial trend was identified. Verification by the Kolmogorov-Smirnov test did not allow rejecting the hypothesis of normal distribution (the attained significance was 89.75%).

The next stage was to run variogram analysis and to build a variogram model using the vgm function. Fig. 2 shows an experimental variogram of four directions (0, 90, 135, and 270°) for the configured variogram model. It was used for ordinary kriging (1): from a set of input observations, point-specific values were removed one-by-one, each time predicting the removed value by kriging using the krige function. As a result, the glm function produced three logit models. The value d = 0.46 was set as a threshold. The significance of the built models was evaluated by the LR test. The inputs for building logit models were as follows: $N(x_i)$ was the NDVI value at x_i, $i = \overline{1.78}$; the variable $T = 1$ if $N(x_i) \geq 0.46$, else $T = 0$; the variables X and Y were the coordinates of observations used as the explanatory variables; $N_{pred}(x_i)$ were the kriging-predicted point-specific values of the parameter.

The estimated coefficients of the second logit model were largely insignificant. The value α was assumed to equal 0.05, see Table 1. Statistical testing proved that the equation of this logit model is generally not significant, whereas

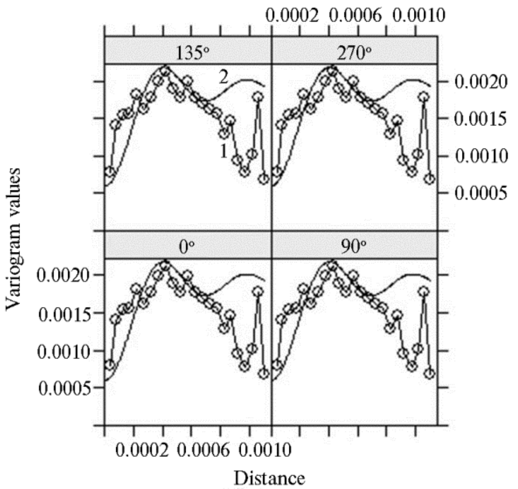

Fig. 2. Experimental variogram of the $\hat{\gamma}(h)$ **spatial NDVI distribution on the test site** (1), **with the theoretical model superimposed** (2) **in four directions (0, 90, 135, 270)** (Menkovo, Leningrad Province, 2015).

the equations of Models 1 and 2 are; testing those showed the third model was better.

The next stage was to test the adequacy of the three obtained models; to that end, points were removed from the input data one by one, the logit models were built again and revaluated to find how accurately each model would predict the probability of exceeding the threshold in the removed point. Testing showed that the first model made errors for 26 points (33.3%), the second one was wrong in 12 points (15.38%), see Table 2.

Similar experiments on simulated data show that the second complete model is better as long as it is statistically significant.

1. Results of constructing the logit models of spatial NDVI distribution; the models use different explanatory variables based on the aerial photography data (Menkovo, Leningrad Province, 2015).

Logit model 1 (the dependent variable is T, the explaining variables are X and Y)	
$$P(T=1)=\frac{1}{1+e^{-479071+5717X+5173Y}}$$	
Coefficient χ^2	5.53
Significance of	
coefficient 1 (constant term)	0.0263
coefficient 2 at X	0.0458
coefficient 3 at Y	0.0236
Logit Model 2 (the dependent variable is T, the explanatory variables are X, Y, and N_{pred})	
$$P(T=1)=\frac{1}{1+e^{-326585{,}69+3587{,}44X+3683{,}28Y-26{,}65N_{pred}}}$$	
Coefficient χ^2	11.049
Significance of	
coefficient 1 (constant term)	0.2424
coefficient 2 at X	0.3488
coefficient 3 at Y	0.2052
coefficient 4 at N_{pred}	0.0238
Logit Model 3 (the dependent variable is T, the explanatory variable is N_{pred})	
$$P(T=1)=\frac{1}{1+e^{14{,}299-32{,}068N_{pred}}}$$	
Coefficient χ^2	9.207
Significance of	
coefficient 1 (constant term)	0.00416
coefficient 2 at переменной N_{pred}	0.00439

N o t e. The observation coordinates X and Y, as well as the set of the kriging-predicted values $N_{pred}(x_i)$, were used as the explanatory variables for logistic regression.

Similar results were obtained by Fernandes et al. [17] who studied a credit scoring logit model using a spatial variable as an explanatory one. They compared two models, one that contained a spatial variable and one that did not. The results showed that the author-proposed method had better performance than conventional methods. In this paper, we studied an approach for predicting the spatial distribution of the parameter of interest, which is based on the combined use of kriging and binary regression; the complete model (where the logit model incorporates a set of kriging-predicted parameter values) was better than the alternatives. Notably, it was only in the 2000s that binary regression found application in precision agriculture in Russia. Some reports [7, 20] give detail upon the opportunities to use logit and probit models in plant growing; however, those do not take into account the spatial variable.

2. Sample of the NDVI logit model testing results as obtained on the basis of aerial photography data (Menkovo, Leningrad Region, 2015)

Coordinate point No.	*X*	*Y*	*N*	*T*	N_{pred}	P	
						Model 1	Model 3
1	30.032934	59.418484	0.527	1	0.4894404	6.187555e-11	0.7519646
2	30.032902	59.418514	0.517	1	0.5037567	4.848053e-12	0.8326184
3	30.032835	59.418605	0.527	1	0.4917005	0.9999989	0.7661754
4	30.032695	59.418778	0.407	0	0.4396790	0.9999876	0.3652577
5	30.032673	59.418811	0.455	0	0.4261240	4.863455e-12	0.2654931
6	30.032588	59.418940	0.461	1	0.4387105	6.46559e-13	0.3366613
7	30.032477	59.419087	0.517	1	0.4614526	0.8979254	0.5382949
8	30.032327	59.419302	0.496	1	0.4600064	2.212942e-11	0.5256631
9	30.032472	59.419176	0.468	1	0.4688632	2.652915e-10	0.6007014
10	30.032528	59.419119	0.411	0	0.4656420	1.943197e-06	0.5943595

Note. *X* and *Y* are the observation coordinates, *N* are the values of input observations, *T* is the dependent variable, $N_{pred}(x_i)$ are the kriging-predicted parameter values. For models 1 and 3, the probabilities of threshold exceedance, predicted for the removed observation points, are presented.

Thus, the approach proposed herein is to use a set of the parameter-of-interest values predicted by ordinary kriging as one of the binary regression factors in the logit model. In general, combining kriging and binary regression to evaluate the plant condition seems to be promising and relevant. However, the experiments sometimes produced statistically insignificant models, this is why it is recommendable to use more examples to evaluate the proposed approach.

REFERENCES

1. Bure V.M., Mitrofanova O.A. Analysis of aerial photographs to predict the spatial distribution of ecological data. *Contemporary Engineering Sciences*, 2017, 10(4): 157-163 (doi: 10.12988/ces.2017.611175).
2. Kim Y., Reid J.F., Han S. On-the-go nitrogen sensing and fertilizer control for site-specific crop management. *International Journal of Agricultural and Biosystems Engineering*, 2006, 7(1): 18-26.
3. Blackmer T.M., Schepers J.S. Aerial photography to detect nitrogen stress in corn. *J. Plant Physiol.*, 1996, 148(3-4): 440-444 (doi: 10.1016/S0176-1617(96)80277-X).
4. Graeff S., Pfenning J., Claupein W., Liebig H.P. Evaluation of image analysis to determine the N-fertilizer demand of broccoli plants. *Advances in Optical Technologies*, 2008, Article ID 359760 (doi: 10.1155/2008/359760).
5. Thenkabail P.S., Lyon J.G., Huete A. *Hyperspectral remote sensing of vegetation*. CRC Press, MA, USA, 2011.
6. Franzen D.W., Reitmeier L., Giles J.F., Cattanach A.C. Aerial photography and satellite imagery to detect deep soil nitrogen levels in potato and sugarbeet. *Proc. of the 4th International Conference «Precision Agriculture»*. St. Paul, MN, 1999: 281-290.
7. Bure V.M. *Materialy Mezhdunarodnogo seminara, posvyashchennogo pamyati professora Ratmira Aleksandrovicha Poluektova (Poluektovskie chteniya)* [Proc. International Seminar dedicated to the memory of Prof. Ratmir Alexandrovich Poluektov (Poluektov reading)]. St. Petersburg, 2014: 118-121 (in Russ.).
8. Debella-Gilo M., Etzelmüller B. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. *Catena*, 2009, 77(1): 8-18 (doi: 10.1016/j.catena.2008.12.001).
9. Kempen B., Brus D.J., Heuvelink G.B.M., Stoorvogel J.J. Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. *Geoderma*, 2009, 151(3-4): 311-326 (doi: 10.1016/j.geoderma.2009.04.023).
10. Yakushev V.P., Zhukovskii E.E., Kabanets A.L., Petrushin A.F., Yakushev V.V. *Variogrammnyi analiz prostranstvennoi neodnorodnosti sel'skokhozyaistvennykh polei dlya tselei tochnogo zemledeliya* [Variogram analysis of agricultural fields spatial heterogeneity for precision farming]. St. Petersburg, 2010 (in Russ.).

11. Claret M.M., Urrutia R.P., Ortega R.B., Best S.S., Valderrama N.V. Quantifying nitrate leaching in irrigated wheat with different nitrogen fertilization strategies in an Alfisol. *Chilean Journal of Agricultural Research*, 2011, 71(1): 148-156 (doi: 10.4067/S0718-58392011000100018).
12. Krasilnikov P., Sidorova V. Geostatistical analysis of the spatial structure of acidity and organic carbon in zonal soils of the Russian plain. In: *Soil geography and geostatistics: concepts and applications*. P. Krasilnikov, F. Carré, L. Montanarella (eds.). European Communities, Luxembourg, 2008: 55-67.
13. Dulaney W.P., Lengnick L.L., Hart G.F. Use of geostatistical techniques in the design of an agricultural field experimental. *Proc. of the Survey research methods section*. Alexandria, VA, 1994: 183-187.
14. Adamsen F.J., Pinter P.J., Barnes E.M. Measuring wheat senescence with a digital camera. *Crop Science*, 1999, 39(3): 719-724 (doi: 10.2135/cropsci1999.0011183X003900030019x).
15. Panayi E., Peters G.W., Kyriakides G. Statistical modelling for precision agriculture: a case study in optimal environmental schedules for Agaricus Bisporus production via variable domain functional regression. *PLoS ONE*, 2017, 12(9): e0181921 (doi: 10.1371/journal.pone.0181921).
16. Isaaks E., Srivastava M. *An introduction to applied geostatistics*. Oxford University Press, NY, USA, 1989.
17. Fernandes G.B., Artes R. Spatial dependence in credit risk and its improvement in credit scoring. *Eur. J. Oper. Res.*, 2016, 249(2): 517-524 (doi: 10.1016/j.ejor.2015.07.013).
18. Dem'yanov V., Savel'eva E. *Geostatistika. Teoriya i praktika* [Geostatistics. Theory and practice]. Moscow, 2010 (in Russ.).
19. Webster R., Oliver M.A. *Geostatistics for environmental scientists, Second edition*. John Wiley and Sons Ltd, Chichester, UK, 2007 (doi: 10.1002/9780470517277).
20. Yakushev V.P., Bure V.M., Parilina E.M. *Binarnaya regressiya i ee primenenie v agrofizike* [Binary regression and its application in agrophysics]. St. Petersburg, 2015 (in Russ.).
21. Hosmer D., Lemeshow S. *Applied logistic regression, Second edition*. Wiley, NY, 2000 (doi: 10.1002/0471722146).
22. Bure V.M., Parilina E.M. *Teoriya veroyatnostei i matematicheskaya statistika: uchebnoe posobie* [Probability theory and mathematical statistics: tutorial]. St. Petersburg, 2013 (in Russ.).
23. Plant R.E. *Spatial data analysis in ecology and agriculture using R*. CRC Press, Boca Raton, 2012.
24. Yakushev V.P., Kanash E.V., Konev A.A., Kovtyukh S.N., Lekomtsev P.V., Matveenko D.A., Petrushin A.F., Yakushev V.V., Bure V.M., Rusakov D.V., Osipov Yu.A. *Teoreticheskie i metodicheskie osnovy vydeleniya odnorodnykh tekhnologicheskikh zon dlya differentsirovannogo primeneniya sredstv khimizatsii po opticheskim kharakteristikam poseva: prakticheskoe posobie* [Theoretical and methodological basis for the allocation of homogeneous technological zones for the differentiated use of chemicals according to the optical characteristics of sowing: practical guide]. St. Petersburg, 2010 (in Russ.).
25. Kanash E.V., Osipov Ju.A. Optical signals of oxidative stress in crops physiological state diagnostics. *Proc. 7th European conference on precision agriculture*. Wageningen, Netherlands, 2009: 81-89.
26. Goovaerts P. *Geostatistics for natural resources evaluation*. Oxford University Press, NY, 1997.