Prediction of agricultural land use

A Rasskazova and Yu Sinits
State University for Land Management, 15, Kazakova St., Moscow 105064, Russia

Abstract. In the article the authors examined the role of forecasting in the system of agricultural land use management. Particular attention was paid to the choice of a method for predicting agricultural land use. The classification of formalized methods of agricultural land use forecasting was highlighted, which include extrapolation and modeling methods. The authors examined the features of each forecasting method and the possibility of its use for predicting agricultural land use. The experience of forecasting agricultural land use in India is considered. The study was based on the work of Russian and foreign scientists like A.A. Varalamov, S.A. Galchenko, A. Pankrats, S. Makridakis and others.

1. Introduction
The economic development of any agricultural region directly depends on its land and resource potential. The management of agricultural land use is a complex system that includes the forecasting and planning of land use, land supervision, land monitoring and so on.

The forecasting with agricultural land use is the most important pre-project activity and initial stage of the agricultural land use management process the main goal of which is rational use and protection of land.

To achieve goals regarding the prediction of agricultural land use, it is necessary that the chosen method of forecasting should correspond to the maximum extent.

2. Materials and methods
Among the forecasting methods used in the agricultural land use management system, the extrapolation and modeling methods should be highlighted.

The idea of the extrapolation methods is expressed in the analysis of changes in the objects of research in time and the distribution of the revealed regularities into the future. Inside the extrapolation methods, there are as follows: simple extrapolation methods, least squares, exponential smoothing, moving averages and others [1].

The method of simple extrapolation involves calculation of the average value of indicator, which forms the basis of the short-term forecast. So, for example, it is necessary to substantiate a short-term forecast of products of a particular agricultural land use, i.e. it is necessary to determine the arithmetic mean value:

$$X_n = \frac{\sum X_i}{n}$$  \hspace{1cm} (1)

where $X_n$ is the predicted value; $X_i$ is the volume of agricultural products; $n$ is the number of years under consideration.

Analytical alignment in forecasting is a finding of mathematical function that most accurately describes the change trend.
Among the modeling methods aimed at forecasting the agricultural land use the most commonly used methods are the methods of network, structural and statistical modeling and economic and mathematical methods [2].

With the help of structural models, it is possible to distinguish qualitatively homogeneous groups based on a large number of features. For example, land use in similar conditions.

The network modeling is based on the construction of the activity network. Each type of work in the activity network can be represented by an arrow that connects the initial and final event (Figure 1).

![Figure 1. Activity network](image)

Each type of work is shown by two numbers on the graph, corresponding to the initial and final events. In this example, the type of work (1.2) is indicated by A1; the type of work (1.3) is A2. Each type of work is characterized by two parameters: the duration of its implementation and the number of resources necessary for its implementation.

Neural network modeling is used along with other methods of forecasting of agricultural land use.

A neural network is a set of special mathematical functions with a variety of parameters that are configured in the process of learning from past data. Then, the neural network processes the initial real data and gives its forecast of future behavior of the system under study.

When applying this method for the purpose of predicting the use of land resources, the layer of input signals will be the values of the factors that have the greatest impact on the indicator being studied, i.e. values of independent variables. The values of the predicted indicators (dependent variables) corresponding to the existing set of values of the input variables are used as the output layer.

Let’s consider the methods for predicting agricultural land use abroad, in particular, the experience of India. The following indicators are considered as predictable, i.e. the output of products from the area of agricultural crops, the spread of pests of cultivated plants and the diseases, as well as the dynamics of the volumes of agricultural crops.

For agricultural forecasting in India, the following statistical forecast models are applied: regressive, time series, and probabilistic.

The regressive models are multiple linear regression (for predicting output per unit area), meteorological indices (for predicting output per unit area, and the number of crop pests), as well as a regressive logic model (for predicting low or high yields, the presence or absence of pests/diseases, etc.) [3].

The regressive model which is very close to the model of multiple linear regression has the following form:

$$ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon, $$

(2)

where $\beta_0$ and $\beta_i$ ($i = 1, 2, \ldots, p$) are evaluated variables and $\varepsilon$ is random errors of representativeness; $Y$ is the predicted variable; $X_1, X_2, \ldots, X_p$ are forecast variable; $p$ shows the number of variable forecast.

The obtained assumptions (limitations, conditions) of the inverse reaction of $Y$ to the predicted $X_1, X_2, \ldots, X_p$ are linear regression $\beta_0, \beta_1, \ldots, \beta_p$. The drop-down errors are independent of each other and equally distributed (idd) random variables with an average zero value can fluctuate within $\sigma^2$. The errors
are independent from each other (pairwise mixed moment of the second order is zero), according to which the predictor variable $X_1, X_2, ..., X_p$ is not accidental and error-free [4-6].

In practical application, the indicator of output per unit area is designated as $Y$, and biometric indicators of crops as $X$ ($i = 1, 2, ..., p$).

Depending on the crop, the biometric indicators of seedlings are plant growth, the number of spike-heads of panicles, etc. In addition to the “growth index” the indicators of a certain period of time are taken during the harvest period.

To build weather indices, the parameters of changing weather forecast, using the following formulas are taken [6]:

$$Y = a_0 + \sum_{i=1}^{p} a_i Z_i + \sum_{i, j}^{p} a_{ij} Z_{ij} + e,$$  \hspace{1cm} (3)

$$Z_i = \sum_{w=1}^{m} r_{iw} X_{iw},$$ \hspace{1cm} (4)

$$Z_{ij} = \sum_{w=1}^{m} r_{ijw} X_{iw} X_{jw},$$ \hspace{1cm} (5)

where $Y$ is the dependent variable which shows the output per unit area / pest; $e$ is a random error; $r_{iw}$ is the correlation coefficient; $Y \in 1$ denotes the variable value of weather parameter; $r_{ijw}$ shows the resulting volume of production; $i^{th}$ and $j^{th}$ are variables of the weather parameter during the week $w^{th}$; $m$ indicates the weekly forecast; $p$ indicates several weather parameters; $Z_i, Z_{ij}$ the predictors that depend on the basic weather indicators of maximum and minimum temperatures, precipitation in the form of rain, relative humidity, etc.

These models use the weather parameters for a week for a year or several years.

In cases when the predictor is a qualitative indicator, a regressive logical model is used. In the absence of data on the number of pests of agricultural crops / diseases of agricultural crops, such qualitative indicators as incidence / non-occurrence, low / high morbidity, etc. are taken. The preferred statistical model for analyzing such dichotomous responses is the binary logistic regression model. It is used to describe the relationship of several independent variables to a binary dependent variable. The logistic regression is used to obtain the probabilities of occurrence of various classifications [7, 8].

$$P(E = 1) = \frac{1}{1 + \exp(-z)},$$ \hspace{1cm} (6)

where $z$ is the weather parameter function.

If $P(E = 1)$ is greater than 0.5, then the probability of the disease is higher, and if it is less, then the probability of occurrence of the disease is minimal. For more accurate forecast, the value of up to 0.7 is taken.

Time series models reflect the analysis of data for a specific time interval. The basic principle of the model is to predict what will happen after certain period of time, without determining the factors that have an impact on this.

Time series models are used in short term forecasting. For this type of forecasting it is enough to have from 20 to 50 indicators. The more indicators we have, the more accurate the forecast is.

Time series models include exponential smoothing and an autoregressive integrated moving average.

A probabilistic model is built on a Markov chain to predict the yield of agricultural products per unit area.

The above mentioned models are effectively applied in agricultural forecasting. Depending on the results obtained, the course of the direction of the state policy in the field of agriculture is determined.

3. Results and Discussion

As studies show the application of the method of analytical alignment for the purpose of agricultural land use, the analytical functions align well the initial dynamic range (the correlation coefficient exceeds 0.7). This significantly reduces the autocorrelation (the autocorrelation coefficient decreases four times or more), its actual value is much less than the critical table value, which at 5% probability level of significance and the length of the original time series (16 years) is equal to 0.323.
All selected functions are suitable regarding the forecast, i.e. with their help one can extend the current trend for the future and find the most likely limits of deviations of the predicted values, for example, the yield of winter wheat [9].

Besides, the method of exponential smoothing is used to determine the effectiveness of agricultural land use. This alignment is particularly strongly fluctuating time series for subsequent prediction. This method allows you to make reasonable predictions based on the time series, having a moderate relationship in time, and provides a greater account of the indicators achieved in recent years. Let’s consider the application of this method and define the predicted value of crop yields. We presume that the calculation period will be qualitatively similar to the last segment of the period under study. The coefficients of equation (estimates of the coefficients), found with the exponential smoothing of the last value of the index in the original time series, are used for the subsequent prediction.

Statistical modeling for the purpose of determining the effectiveness of land use is performed on the basis of a single regression equation and on the basis of a system of regression equations. The forecast options can be viewed as both one equation and a set of them [10].

The studies conducted using the regressive analysis methods at agricultural enterprises of the Ryazan region allowed determining the list of main independent variables for the inclusion into the statistical models for such important indicators as crop yield Y1, productivity of farm animals Y2, labor productivity Y3, and production cost Y4 (Table 1).

### Table 1. Statistical model data

| Indicators            | Independent variables                                                                                                                                 |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Crop yield Y1,        | Evaluation points, differential income per 1 ha, cost recovery, the level of production mechanization, capital supply, capital intensity, irrigation level, forms of labor and production organization, the amount of mineral and organic fertilizers applied; coefficients characterizing the degree of development of the management system |
| Productivity of farm animals Y2 | Feed area, quality of feed, security of the main ingredients of nutrition; breed composition of animals, labor costs, coefficients of application of progressive zootechnical measures, costs of veterinary measures |
| Labor productivity Y3  | The level of mechanization of labor and production, indicators of the level of specialization of production, indicators of social infrastructure, wages, working conditions, provision of social infrastructure |
| Production costs Y4    | Crop yields, productivity of farm animals, the share of individual cost items, product quality indicators, the share of mechanized labor and production |

In order to effectively manage agricultural land use, the cluster analysis can be applied. This method of forecasting involves the selection of compact and distant groups of objects, looking for a “natural” partition of the aggregate into the areas of clusters of objects. The clustering process is a system consisting of a series of interconnected elements. An example of the use of the clustering method in forecasting the agricultural land use can be the clustering of municipalities in the region according to the following indicators: the efficiency ratio of agricultural land use, the level of profitability of all activities,%; net income of agricultural products to the total cost, rub./rub.; profits of agricultural enterprises from all activities, rub/ha..

### 4. Conclusion

To sum up, it should be noted that when choosing a method for predicting agricultural land use, the following factors are of particular importance, i.e. the purpose of the forecast, the characteristics of the available data, the forecast period, the permissible probability of errors, the cost of the forecast.

For effective management of agricultural land use at various administrative and territorial levels, it is necessary to create a reliable and effective system of forecasting and planning indicators, from various aspects, characterizing the state of land resources and the effectiveness of their management. The
requirements for the composition and structure of indicators are formed in accordance with the types of the subject and the tasks set.

References

[1] Komarov S I and Rasskazova A A 2018 *Forecasting and planning the use of land resources and real estate* (Moscow)

[2] Galchenko S, Varlamov A and Bogdanova O 2018 Theoretical and methodological foundations for formation of sustainable land management system *IOP Conf. ser. Materials Science and Engineering* 012141

[3] Sinitsa Yu S 2016 *Economic efficiency of land cadastral systems of the Russian Federation*, candidate thesis, State University of Land Management (Moscow)

[4] Agrawal R, Mehta S C, Bhar L M and Kumar A 2004 Development of weather based forewarning system for crop pests and diseases – Mission mode project (NATP) report IASRI publication

[5] Jain R C, Jha M P and Agrawal R 1985 Use of growth indices in yield forecast *Biometrical J.* 27(4) 435–39

[6] Makridakis S, Wheelwright S C and Hyndman R J 1998 *Forecasting: Methods and Applications* (New York: John Wiley)

[7] Misra A K, Prakash O and Ramasubramanian V 2004 Forewarning powdery mildew caused by Oidium mangiferae in mango (Mangifera indica) using logistic regression models *Indian J. of Agricultural Sci.* 74(2) 84–7

[8] Pankratz A 1983 Forecasting with univariate Box Jenkins models: concepts and cases (New York: John Wiley&Sins)

[9] Askolskaya É A 2010 Some aspects of methodology for the selection of experts based on assessing the degree of confidence in the level of professional judgment *News of the Orenburg State Agrarian University* 4(28–1) 170–2

[10] *Electronic textbook on statistics* StatSoft 2001 (Moscow: StatSoft) Available at: http://www.statsoft.ru/home/textbook/default.htm