Method of statistical tests in solving problems of food production management

T S Buzina, A Yu Belyakova and Ya M Ivanyo
Department of informatics and mathematical modeling, FSBEI HE Irkutsk SAU, Irkutsk district, Irkutsk region, Russia

E-mail: buzinats@mail.ru

Abstract. The article discusses some of the problems of forecasting and planning the production of agricultural products and harvesting of wild plants in conditions of uncertainty. Variants of their solution using the method of statistical tests are proposed. The factorial regression equations for modeling the yield of grain crops depending on precipitation in the initial growing season and time are given. Two dependencies are considered for different agrolandscape areas. In one case, the relationship between the yield of grain crops and precipitation and time was determined, and in the second, a dependence on precipitation was found. The direct and inverse problems are considered. The conditions for obtaining high and low yields were determined. A linear model for optimizing the placement of grain crops is proposed using the Monte Carlo method for assessing random indicators. The extreme problem is applied to a municipal district with developed agriculture. The results of modeling are presented, which make it possible to estimate the volume of production of grain crops for different levels of productivity, which correspond to the reference probabilities. A multi-criteria parametric model of the functioning of clusters for obtaining food wild-growing products with probabilistic estimates in constraints and a parameter included in the objective function is described and implemented. The yield of wild plants was used as a parameter, on which the sales price depends.

1. Introduction
The Monte Carlo method or statistical test method refers to numerical methods for solving mathematical problems by simulating random samples.
A large number of tasks related to forecasting and managing food production include uncertain indicators that can be estimated using interval and random variables. In such cases, it is effective to use the method of statistical tests [1, 2, 3].

From the set of models describing the variability of crop yields, one can single out factor models that characterize the relationship of the successful indication with meteorological parameters, which are random variables or random variables with low significant autocorrelation coefficients [4, 5].

In a number of works [4, 5, 6, 7], models of the dependence of the yield of grain crops on precipitation and air temperature during the growing season are presented. In this case, to model the successful indication, it is effective to use the Monte Carlo method to determine the combinations of meteorological parameters that affect the values of crop yields. The presence of significant regression expressions allows solving inverse problems, determining the combination of factors for obtaining high or low yields [8, 9]. Solving such problems is especially effective for the relationship between yield and rainfall, since rainfall is a controllable factor due to artificial sprinkling [6, 10, 11].
Knowledge of probable situations contributes to the identification of possible risks and the use of adequate management decisions for various simulated situations [12]. In this case, the method of statistical tests is applicable in the case of knowledge of the probability distribution laws for the studied quantities, as well as in the case of their description using interval estimates [13, 14, 15, 16].

The use of the Monte Carlo method for modeling uncertain parameters is described in the problems of obtaining optimal sowing plans, crop production [17, 18], livestock products and their combinations. In addition, this method was used in the tasks of optimizing the interaction of participants in agro-industrial clusters [19, 20] and clusters for obtaining wild-growing products [21]; in models for optimizing the receipt of agricultural products in conditions of risks [22] and ecological and mathematical models [23].

Many authors pay attention to climate variability. Analysis of rare hydrometeorological events on the territory of the Irkutsk region shows that they continue to bring significant damage to the region's economy. In [4], algorithms are proposed for modeling rare events using the method of statistical tests at a given repeatability interval of a certain extreme value. In this case, the probability of the event manifestation and the scattering interval are estimated. In the development of models for optimizing agricultural production, it is necessary to take into account the best distribution of sales by season. To calculate the optimal coefficients describing the distribution, we use the Monte Carlo method.

In addition, in the tasks under consideration, it is necessary to take into account natural and man-made risks, depending on the characteristics of the territory and external conditions. A review of the tasks in which the statistical test method is used shows its wide possibilities for application in various tasks of agricultural production management.

2. Materials and methods
The work used the data of hydrometeorological observations for the growing season 2000 - 2019, the yield of agricultural crops in the cultivar plots of Kachug and Usol'e-Sibirskoye, as well as the yield in the municipal districts of the Irkutsk region. In addition to this, sb attracted information about the development of agriculture in the Usolsky District, one of the best in this respect in the region.

When solving a multicriteria problem of parametric programming to optimize the harvesting and processing of wild-growing products, information about the yield of wild plants, harvesting areas, labor costs, processing capacities, and the amount of resources were used.

When obtaining the results, the methods of statistical data processing, the Monte Carlo method, methods of forecasting and mathematical programming under conditions of uncertainty were used.

The work performed is the development of scientific research carried out by various authors at different times on the problems of optimization and forecasting of food production in conditions of uncertainty of climatic and production and economic indicators [4, 5, 17, 18, 19, 21, 22, 23].

3. Results and discussion
The yield of agricultural crops is used to plan and manage the activities of an agricultural enterprise.

Long-term studies on the statistical assessment of the yield of agricultural crops, the determination of the influence of the main meteorological factors on it, the assessment of the climatic background of different territories of the region made it possible to identify some regularities for constructing models for predicting the yield of grain crops in the Irkutsk region.

A significant number of tasks related to forecasting and managing food production include uncertain parameters that can be estimated using interval and random variables. In this case, the parameters included in the model can be dependent and independent. In the first and second cases, a large number of solutions are obtained, which are necessary to identify possible situations for the development of agricultural production.

In [19], to assess the influence of meteorological factors on the yield of grain crops, the results of field studies and data on air temperature and precipitation by months, from May to September, for 2000-2018, were used.
In continuation of these studies, the following algorithm is proposed. Based on the obtained regression expressions describing the wheat yield depending on climatic factors, a direct problem was solved using the method of statistical tests. According to a given law of probability distribution, precipitation values are randomly determined. Then, according to the obtained dependence, the yield is calculated. Since the series of precipitation are asymmetric, a three-parameter power-law gamma distribution was adopted as the distribution law [20]. Its application was confirmed using the Kolmogorov agreement test. Table 1 shows the results with reference probabilities: 0.05, 0.50, 0.95.

In addition to assessing the yield by the Monte Carlo method, the remainder of the series is determined to clarify the yield values. In this table, \( y_t \) is the simulated yield, \( y_t + \varepsilon_t \) is the yield with a random addition, \( p \) is the probability, \( z \) is the June rainfall.

**Table 1.** The results of solving the direct problem of finding the yield using the Monte Carlo method for the Kachuga and Usolye-Sibirskoye variety plots.

| \( p, \%) | z, mm | \( y_t, \text{ c/ha} \) | \( y_t + \varepsilon_t, \text{ c/ha} \) |
|---|---|---|---|
| Kachug | | | |
| 0.05 | 17.2 | 8.04 | 7.08 |
| 0.5 | 38.5 | 12.2 | 12.2 |
| 0.95 | 83.6 | 20.9 | 21.8 |
| Usolye-Sibirskoye | | | |
| 0.05 | 22.7 | 9.11 | 5.82 |
| 0.5 | 66.9 | 17.6 | 17.6 |
| 0.95 | 120.4 | 28.0 | 31.2 |

In the case of the implementation of the inverse problem at the first step, the statistical parameters of the yield series and factors were estimated [17]. Then, based on the selected probability distribution law, the yield series are modeled using the Monte Carlo method. According to the data obtained, using a certain regression model, the values of the factor are calculated.

To optimize the placement of the structure of crops, deterministic models are often used, in which the parameters are averaged values. In real conditions, they are uncertain and can be described in some cases by the distribution law, in others, with insufficient information, by upper and lower estimates.

If the parameters are probabilistic values, then the model can be written in the following edition. The objective function characterizes the profit:

\[
\sum_{i\in I} \sum_{s\in S} d_{is} y_{is} (t_{is}, t) x_{is} - \sum_{i\in I} \sum_{s\in S} c_{is} x_{is} \rightarrow \max ,
\]

under conditions:

limited production resources

\[
\sum_{s\in S} y_{is}^p x_{is} \leq V_i \quad (l \in L, i \in I);
\]

limited size of the crop industry

\[
\sum_{i\in I} \sum_{s\in S} (1 + \eta_s) x_{is} \leq \bar{n}
\]

production of final products of a given volume

\[
\sum_{i\in I} \sum_{s\in S} y_{is} (t_{is}, t) x_{is} \geq Y_s \quad (s \in S);
\]

limited application of fertilizers and plant protection products...
\[ \sum_{s \in S} w_{mis} x_{is} \leq W_{mi} \quad (m \in M, i \in I) \];

non-negativity of variables

\[ x_{is} \geq 0. \]

Here \( d_{is} \) is the selling price of culture \( s \) of field \( i \) (rubles/c); \( y_{is}^p \) is the output of products per unit area of the crop \( s \) of the field \( i \) (c/ha), corresponding to a certain probability \( p \); \( x_{is} \) is the cultivation area of the crop \( s \) on the field \( i \) (ha); \( c_{is} \) is the cost per 1 ha of field \( i \) of crop \( s \) (rubles / ha); \( v_{il}^m \) is the consumption of resource \( l \) per unit of crop area \( s \) field \( i \) (thousand man-h/ha, thousand rubles/ha), corresponding to a certain probability; \( V_{li} \) is the presence of a resource of the form \( l \) of field \( i \); \( Y_s \) is the guaranteed (minimum) volume of production of culture products \( s \) (c); \( \bar{n} \), \( n \) is the maximum and minimum possible cultivation area (ha); \( \eta_s \) is a coefficient that takes into account the area of sown seeds of culture \( s \); \( w_{mis} \) is the presence of fertilizer \( m \) (plant protection products) per unit area of the crop \( s \) field \( i \) (centner/ha); \( W_{mi} \) is the presence of fertilizer of the form \( m \) of the field \( i \) (c).

Since Eastern Siberia belongs to the zones of risky farming with a sharply continental climate, crop yields are subject to significant fluctuations, which must be taken into account when planning agricultural production.

In situations where the yield of agricultural crops is described using one- and multifactor dependencies, a model for optimizing the placement of crops is proposed, the left side of constraint (4) of which has the following form:

\[ y_{is} = \alpha_{is} + \alpha_{is}^p t_{is}^p + \alpha_{is}^t t, \]

where \( t_{is}^p \) are precipitation at the beginning of the growing season, \( t \) is time, \( p \) is likelihood.

The study of the series of the factor shows that it is random, obeying the three-parameter power-law gamma distribution. In this case, when solving problem (1) - (7), the method of statistical tests can be used to simulate climatic factors.

The task of optimizing the structure of grain crops with a probabilistic parameter is implemented for the Usolsky district of the Irkutsk region. The maximum profit from the sale of grain products is taken as an optimality criterion.

The solution to the problem was carried out on the basis of the simulated values of the yield (table 1) using the three-parameter gamma distribution function for the probabilities of 0.05, 0.5 and 0.95. Table 2 shows the results of modeling the placement of grain crops.

**Table 2.** The results of solving the problem of optimizing the placement of grain crops for the probabilities of 0.05, 0.5 and 0.95 (according to the Usolsky district).
When solving the problem of optimizing the placement of crops of grain crops with high yields (the probability of exceeding 0.95), the profit from the sale of products was obtained, equal to 208955 thousand rubles. At the same time, production costs amounted to 217281 thousand rubles. In the case when the yield values are minimal (the probability of exceeding is 0.05), the value of the objective function decreases by 38%. The difference between the values of the objective function for the worst and best variants relative to the mean is 57% and 34%, respectively.

The proposed approach to optimizing the placement of agricultural crops using the simulation method allows the decision-maker to obtain various options for solving problems depending on the variability of yield and other parameters.

In continuation of developments on the creation and implementation of models for optimizing the harvesting, processing and sale of food wild forest products, a multi-criteria model of parametric programming with probabilistic estimates is proposed. Incomes of purveyors and processors were determined as criteria of optimality. When constructing a model for optimizing the interaction of cluster members, the coefficients for variables of the objective function are taken to be dependent on the parameter characterizing the yield of wild plants. At the same time, the yield of wild-growing resources is a random value with the assumption that the statistical estimates of the series of bioproductivity of wild plants correspond to similar indicators of the volumes of harvesting.

Therefore, to optimize the functioning of cluster members for obtaining wild-growing products, taking into account sales markets and consumer demand, models with probabilistic estimates can be used. The given assumptions allow using some simplified model for modeling.

The lognormal law and gamma distribution are used to describe the samples of mushrooms and nuts. The long-term row of berry harvesting obeys the Gaussian distribution. Let us present the problem of multicriteria parametric programming with probabilistic estimates of the productivity of wild plants.

The objective function has the form

\[ f = f_1 + f_2 = \sum_{v \in V} \sum_{l \in L} d_{vl}(\alpha_{vl})x_{vl} + \sum_{i \in I} \sum_{j \in J} d_{ij}x_{ij} \rightarrow \max. \]  

(8)

In this case, the revenue per unit of production \( d_{vl} \) is linearly related to the yield of wild plants \( \alpha_{vl} \), which is a random variable with probability \( p \):

\[ d_{vl} = \zeta_{0vl} + \zeta_{1vl} \alpha_{vl}. \]  

(9)

The task constraints are related:

with areas where wild products are harvested

\[ \sum_{v \in V} d_{vl}^p x_{vl} \leq A_l, \]  

(10)

the receipt of guaranteed income by the cluster members

\[ \sum_{v \in V} \sum_{l \in L} b_{vl} x_{vl} + \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \geq B, \]  

(11)

labor costs

\[ \sum_{v \in V} \sum_{l \in L} \beta_{vl} x_{vl} + \sum_{i \in I} \sum_{j \in J} \gamma_{ij} x_{ij} \geq G, \]  

(12)

availability of capacities necessary for processing products

\[ \sum_{i \in I} \sum_{j \in J} x_{ij} \leq W. \]  

(13)
In addition, the ratio of the volumes of procurement and processing of products should be taken into account:

\[ \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij} \leq \sum_{v=L}^{V} \sum_{l=L}^{L} x_{vl} \]  \hspace{1cm} (14)

and the condition for the volume of sales of processed products

\[ \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij} = \sum_{v=L}^{V} \sum_{l=L}^{L} x_{vl} k_{l} \]  \hspace{1cm} (15)

The model must take into account the interests of the procurers of wild-growing products:

\[ \sum_{v=L}^{V} \sum_{l=L}^{L} d_{l} x_{vl} \geq R \]  \hspace{1cm} (16)

All model variables must be non-negative

\[ x_{vl}, x_{ij} \geq 0. \]  \hspace{1cm} (17)

Here \( d_{l} \) is the revenue of the supplier \( l \) from the unit of production \( v \); \( x_{vl} \) is the volume of products \( v \) of the supplier \( l \); \( d_{i} \) is the profit of processing enterprise \( i \) from unit \( j \); \( x_{ij} \) are the volumes of the received product \( j \) of the processing enterprise \( i \); \( a_{ij} \) is the area of the product type \( v \) of the supplier \( l \); \( A_{i} \) is the total area of production \( v \); \( b_{ij} \) is the proceeds from the sale of harvested products \( v \) of the supplier \( l \); \( c_{ij} \) is the proceeds from the sale of processed products \( j \) of processing enterprise \( i \); \( B \) is a guaranteed income; \( \beta_{vl} \) are the labor costs for the procurement of products of the type \( v \) of the supplier \( l \); \( \gamma_{ij} \) are labor costs for processing type of product \( j \) by processing plan \( i \); \( G \) is the limitation of labor costs; \( W \) is the total power; \( k_{l} \) is the share of sold products of type \( v \) from the volume of processed products of type \( j \) by the purchaser \( l \); \( R \) is the profit from the sale of products.

The multicriteria model is implemented on the example of the Irkutsk cluster. The results of calculations of the problem of optimizing the production and processing of wild-growing products for the probabilities of 0.05, 0.5 and 0.95 are shown in Table 3. The model was applied under the assumption of equally probable values of the yield of three species of wild plants.

The highest yields for all types of wild-growing products correspond to the maximum values of the probability. In this case, the proceeds of procurers will amount to 216797 thousand rubles. With a probability equal to 0.05, we can assume that all three types of products under study will turn out to be barren. In the event of such a situation, the revenue of procurers will amount to 7460.8 thousand rubles with a guaranteed profit of the processing enterprise equal to 3902 thousand rubles.

Table 3. Solution of the multicriteria problem of optimizing the procurement and processing of wild food products with probabilistic parameters for the Irkutsk cluster.

| Probability | \( f_{i} \), the profit of the processing enterprise, thousand rubles | \( f_{i} \), revenue of procurers, thousand rubles | \( a^{p} \), indicator of biological productivity of wild plants associated with probability, ha / c |
|-------------|-------------------------------------------------|---------------------------------|-----------------------------------------------|
| 0.05        | 3902                                           | 7460.8                          | Berries: 1.88, Mushrooms: 1.66, Pine nut: 7.95 |
| 0.5         | 7500                                           | 208985.6                        | Berries: 0.33, Mushrooms: 0.39, Pine nut: 0.50 |
| 0.95        | 8500                                           | 216797                         | Berries: 0.18, Mushrooms: 0.15, Pine nut: 0.11 |

In the development of models for optimizing agricultural production, it is necessary to take into account the best distribution of sales by season. To calculate the optimal coefficients describing the distribution, you can apply the Monte Carlo method.

In addition to this, in the tasks under consideration, it is necessary to take into account natural and man-made risks, depending on the characteristics of the territory and external conditions. A review of the
problems in which the method of statistical tests is used shows the broad possibilities of applying the method in the management of agricultural production.

4. Conclusions
The Monte Carlo method in solving problems of agricultural production management in conditions of uncertainty of many indicators shows high efficiency.

The paper presents the results of modeling the yield of grain crops using the method of statistical tests. Precipitation in the initial growing season and time were used as factors. When solving the problem, the rest of the series was simulated.

The proposed dependences of the yield of grain crops on factors are used to optimize the placement of crops. Various solutions were obtained depending on the meteorological conditions affecting the production of grain crops.

The problem of multicriteria parametric programming is solved for the optimization of harvesting and processing of wild plants according to the data of the forest territory of the south of Pre-Baikal region.

References
[1] Prasolova E A, Prasolova E A and Kopyrina T O 2017 Application of the Monte Carlo method in economics Modern scientific research and innovations 2
[2] Raychaudhuri S 2008 Introduction to Monte Carlo simulation Winter Simulation Conference 91-100 doi: 10.1109/WSC.2008.4736059
[3] Olgarenko V I, Yurchenko I F and Olgarenko I V 2017 Application of the Monte Carlo method for modeling the risks of planned irrigation Scientific life 211-9
[4] Ivanyo Ya M and Petrova S A 2015 Optimization models of agricultural production in solving problems of assessing natural and man-made risks (Irkutsk: Irkutsk State Agrarian University) p 180
[5] Polkovskaya M N and Ivanyo Ya M 2016 Optimization of the crop structure taking into account the variability of climatic parameters and crop bioproductivity (Irkutsk: Irkutsk State Agrarian University) p 150
[6] Bernardi Michele and Delince Jacques et al 2016 Crop Yield Forecasting: Methodological and Institutional Aspects p 239
[7] Bryksin V M and AV Evtyushkin 2005 Evaluation of the yield of grain crops based on remote sensing data and a model of bioproductivity Geo-Siberia 5 218-22
[8] Oskin S V and Tarasenko B F 2014 Simulation modeling in the analysis of the efficiency of tillage machines Polythematic network electronic scientific journal of the Kuban State Agrarian University 1025-50
[9] Lyapunova I A and Tetraushvili E V 2014 The value of simulation in agriculture NovaInfo.Ru 25 3-5
[10] Mozol A V and Mozol A A 2018 Assessment of the influence of weather and climatic factors in predicting yield Scientific works of the Belarusian State Economic University (Minsk: Belarusian State Economic University) 288-96
[11] Parfenova V E 2018 Multi-criteria fuzzy choice of the best alternative Bulletin of the International Academy of Agricultural Education 43 135-40
[12] Luzgin A N 2014 Investigation of the influence of the parameters of the adaptive probabilistic-statistical cluster model on the results of interval forecasting Safety of regions - the basis of sustainable development 1(2) 208-12
[13] Lubentsova E V and Rudakova T A 2014 Statistical estimation of the parameters of the interval dynamic model of a biological object Methods and algorithms of applied mathematics in technology, medicine and economics: materials of the 14th international scientific and practical conference (Novocherkassk: South-Russian State Polytechnic University) 12-6
[14] Tsybry L V and Mulyar S S 2017 Interval estimates of the decision of the statistical model of
management *Bulletin of the Transnistrian State Academy of Economics and Architecture* **6** 78-84

[15] Gagarin Yu E and Gagarina S N 2018 Interval estimation of resource consumption volumes with stochastic initial data *University Bulletin* **12** 64-70

[16] Vasiliev O I, Akhmadeev B A and Boychenko I A 2017 Analysis of resource consumption in the system of serial production of standard planting material based on simulation System analysis in design and management: *Collection of scientific Proceedings of the XXI International Scientific and Practical Conference* (St. Petersburg: Federal State Autonomous Educational Institution of Higher Education "Peter the Great St. Petersburg Polytechnic University") 403-11

[17] Ivanyo Ya M, Polkovskaya M N and Fedurina N I 2016 Inverse problems of assessing production and economic parameters in conditions of uncertainty. *Climate, ecology, agriculture of Eurasia: Materials of the V international scientific - practical conference* (Irkutsk: Irkutsk State Agrarian University) 56 - 64

[18] Ivanyo Ya M and Khogoeva E A 2013 Regional economic and mathematical models of agricultural production with interval natural and production and environmental parameters *Izvestiya IGEA* **6(92)** 138-43

[19] Buzina T S, Ivanyo Ya M and Fedurina N I 2015 Multicriteria problem of optimization of interaction of participants of agroindustrial clusters with interval parameters *Scientific and practical journal "Bulletin of IrGSKhA"* 115-23

[20] Kritskiy S N and Menkel M F 1981 *Hydrological foundations of river runoff management* 255

[21] Buzina T S, Ivanyo Ya M and Polkovskaya M N 2020 Multicriteria problem with uncertainty parameters for modeling obtaining forest food products *International Multi-Conference on Industrial Engineering and Modern Technologies* 9271204

[22] Belyakova A Yu, Ivanyo Ya M and Petrova S A 2020 Modeling of rare rain floods on the example of middle rivers of Angara basin *IOP Conference Series: Earth and Environmental Science* (Krasnoyarsk: Krasnoyarsk Science and Technology City Hall) 42103

[23] Ivanyo Ya M and Stolopova Yu V 2019 Climatic variability and agrometeorological conditions of Cisbaikalia: experimental research and modeling of grain yield *Meteorology and hydrology* **10** 117-24