

RESULTS DESIGNING AND ANALYSIS WHEN INTRODUCING NEW BEVERAGE IDENTIFICATION CRITERIA

Irina I. Ageikina¹, Ekaterina G. Lazareva², Irina Yu. Mikhailova¹, Vladislav K. Semipyatny¹

¹ All-Russian Scientific Research Institute of the Brewing, Non-Alcoholic and Wine Industry – Branch of the V. M. Gorbатов Federal Research Center for Food Systems of RAS, Moscow, Russia

² V.M. Gorbатов Federal Research Center for Food Systems of Russian Academy of Sciences, Moscow, Russia

KEY WORDS:
experiment designing, beverage identification criteria, evolutionary operation

ABSTRACT
Building a digital profile of food product with use of modern mathematical apparatus of basic matrices is a solution to the problem of designing innovative beverage recipes. In this regard, for the effective use of the food resource base, modeling and production of high-quality food products, there is an acute problem of developing a methodology for identifying food products using the full range of the currently available analytical base. The article discusses an algorithm for constructing a flexible experimental design for the new identification criteria development, taking into account the laboratory research peculiarities in the beverage industry. The application of software in experiment designing is considered and a practical example of integrated designing based on the construction of an identification criterion for wine materials is presented.

FUNDING: The article was published as part of the research topic No. 0585–2019–0011 of the state assignment of the V. M. Gorbатов Federal Research Center for Food Systems of RAS

1. Introduction

One of the priorities in Russia over the past decade has been to provide the population with high-quality and safe food. The development of technologies in the agro-industrial area makes it possible to produce food products with various consumer characteristics. In this regard, a steady trend has been established to introduce personalized nutrition elements.

The variety of food assortment and, to a greater extent, the dishonesty of some manufacturers do not allow the construction of robust evaluative quality criteria based on one or a narrow group of specialized methods. In this regard, for the effective use of the food resource base, modeling and production of high-quality food products, there is an acute problem of food products identifying methodology developing using the full range of the currently available analytical base.

The main document, regulating the relations of all counterparties of the food industry – manufacturers, processors, distributors, retail chains, consumers and regulatory authorities – is Federal Law No. 29-FZ “On the quality and safety of food products.” Within the framework of this document, the identification of food products is defined as “the activity to establish the compliance of certain food products with the requirements of regulatory, technical documents and information on food products contained in documents attached to them and on labels” [1].

In GOST R51293–99 “Product identification. General provisions” [2] the process of product identification is described as “establishing the conformity of a specific product to a sample and/or its description”, which can also be traced in other regulatory documents [3,4]. Despite the fact that this GOST comprehensively describes the identification process and the requirements for its initiation, and also contains recommendations on the hierarchical structure of the identification methods application, in this work we will move away from the commodity study view of the identification process and will use more theorized definition, which in further will call clustering or discriminating, which postulates identification as a methods set for establishing differences between food groups [5,6].

Since the organoleptic and many biochemical indicators of beverages are not linear, moreover, they are mathematically unstable characteristics, sensitive to minimal recipe changes, the most common linear regression identification method gives acceptable results only in a limited area of the factor space. The clustering methodology based on experiment designing is promising and uses the whole range of analytical methods when constructing discriminatory criteria. The methodology does not use aggregating information on laboratory analyzes, but dynamically operates with unmodified data of the entire multidimensional factor space, which will minimize the error percentage and accurately predict qualitative relationships in beverage technology [7].

2. Materials and methods

The starting point for the identification methods construction and the development of new beverage formulations is the design of new laboratory studies to find new discriminating factors. For this, a full-factorial design of the experiment or its analog is constructed [8,9,10].

In particular, when determining the factors significance, a design with a verification point is used. To build an experimental design, you need to calculate the number of units in a row, equal to the limiting factors number that can be tested for significance. If the number of factors is less than the number of characters in a row, the extra factors are called fictitious and are simply not taken into account. Hence follows a certain information surplus, which makes it possible to put all design tests in one replication, while insignificant losses occur in the description accuracy. If there are no fictitious factors, then the optimal solution is to use a verification test (verification point). This test, by default, is the center point, that is, the row in which all factors are at the average (in coded form – zero) level.

When constructing an experiment design, the most important stage is the choice of the variation interval. On the one hand, the variation interval should not be too small, otherwise, due to the small difference in responses, a significant factor

may be mistaken for insignificant. On the other hand, the interval of variation should not be too large, since there is always a risk of an extremum “missing”. The variation interval is mainly determined by the experiment tasks, the creative part is the discretization of the selected interval for conducting point tests.

Nevertheless, there are so-called “saturated” and “oversaturated” designs, for which the number of degrees of adequacy freedom is equal to zero or even less. In the theory of experiment designing, such designs are widely known [10]. In practice, these designs represent for insignificant factors “elimination” (Table 1).

Table 1

Standard columns for constructing Plackett-Berman designs with different number of factors

Number of factors	7	11	15	19	23
Factor values	1	1	1	1	1
	1	1	1	1	1
	1	-1	1	-1	1
	-1	1	1	-1	1
	1	1	-1	1	1
	-1	1	1	1	-1
	-1	-1	-1	1	1
		-1	1	1	-1
		-1	-1	-1	1
		1	1	1	1
		-1	1	-1	-1
			-1	1	-1
			1	-1	1
			-1	-1	1
			-1	-1	-1
				1	1
				1	-1
				-1	1
					-1
					-1
					-1

When constructing regression models, it should be noted that the description error, other things being equal, is determined by the main diagonal of the correlation-covariance matrix $(X * X)^{-1}$, where X is the experimental design matrix. In the case of using an irregular design, the errors of the coefficients can greatly increase due to the growth of numbers that make up the main diagonal of the correlation-covariance matrix.

It is possible to develop various designs if we take a B_2 design as a basis, that is, a full-factor design. The rest of the designs differ from him in the size of the “star” shoulder. Constructing an optimal design in all respects is a very difficult task and hardly solvable. In this regard, the literature has developed formal criteria for determining the design effectiveness (described in detail in [8]). Let’s consider the main ones:

- ❑ A – optimal design (from “average variance”). This is a design that provides the minimum mean variance of the regression equation coefficients.
- ❑ D – optimal design (from “determinant”). This design is characterized by the minimum determinant value of the correlation-covariance matrix, which provides the most reliable information. Practitioners find this design type as the most appropriate for the work.
- ❑ E – optimal design. This design minimizes the maximum variance of the regression equation coefficients.
- ❑ G – optimal design (from “general variance”). This design minimizes the maximum variance of one of the regression equation coefficients (almost always b_0).
- ❑ Rotatable design – the approximation error does not depend on the direction, but depends only on the distance to the center of the design.
- ❑ Orthogonal design – provides a relatively easy calculation of the regression equation coefficients with “manual” counting (without using a computer).
- ❑ B_n – design, close in properties to the D – optimal design. In practice, this design is convenient because the “star shoulder” is equal to one.

In reality, all these designs differ from each other both in the location of points in the factor space (in particular, in the size of the “star shoulder”), and in the tests number.

Often, the main designs diagonals have additional test in the center. Let consider such deviation feasibility from the principle of minimizing the cost of experiment conducting. Without additional test, the matrix $X * X$ of the design with a “star” shoulder 1.414 cannot be inverted, since it has a singularity. Accordingly, it is impossible to calculate the regression equation coefficients. An increase in the number of experiments to 9 allows a fairly noticeable decrease in the errors of the most regression equation members.

Adding one test to the design with shoulder 1 does not change the main diagonal of the correlation-covariance matrix and, accordingly, the accuracy of the coefficients calculating.

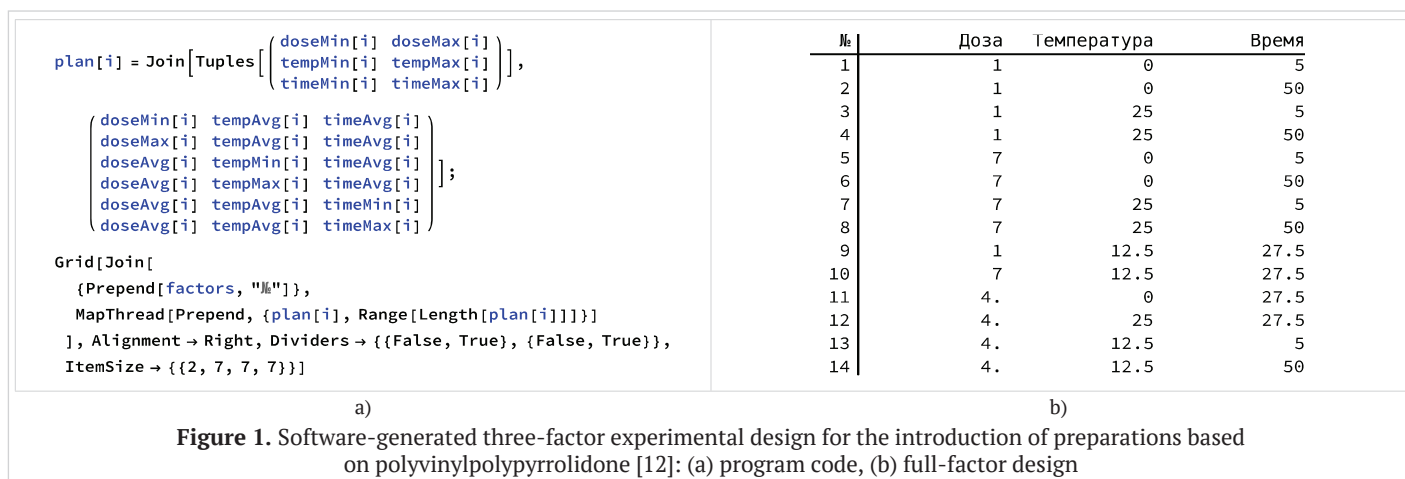
3. Results and discussion

In practice, in the alcohol industry, evolutionary operation (EVOP) [11], which is a kind of factor designing, shows itself well. It is used to find the optimum in conditions of strong “noise” and a small interval of factors variation, which is typical for the identification criteria of the beverage industry, expressed in limited scores and unstable interval estimates of laboratory studies. Under such conditions, it is very difficult to obtain a regression equation that contains significant terms.

When EVOP implementing, a factorial design for a linear model (with a central point) is used, rightly believing that with a small interval of variation, the investigated dependence may well be approximated by a straight line. When conducting a study, several such designs are built in a certain sequence, each of which is called a phase. Each replication of the design (phase) is called a cycle. The cycles are repeated until a significant effect is obtained at least for one factor. After that, a new design is built, taking as the central point the point of the previous design at which the optimal response was obtained. The search for the optimum continues until the desired result is achieved or until the maximum permissible values of the factors are reached.

The algorithm for finding the optimum is as follows (using the example of a two-factorial experiment):

- 1) Construct an experimental design, classical or variable, according to Table 1. For the case with three factors, a matrix of the form shown in Table 2 should be used. For convenience, Figure 1 depicts software written in the Wolfram Language to automate the experimental designs development.



At each point of the design, carry out the test in duplicate. Calculate the average for each test. Do not sort the data in order to exclude anomalous results.

Table 2

Three-factor design for evolutionary operation

No. of the test	Factors value						
	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$
1	0	0	0	0	0	0	0
2	-1	-1	+1	+1	-1	-1	+1
3	+1	+1	+1	+1	+1	+1	+1
4	+1	-1	+1	-1	+1	-1	-1
5	-1	+1	+1	-1	-1	+1	-1
6	-1	-1	-1	+1	+1	+1	-1
7	+1	+1	-1	+1	-1	-1	-1
8	+1	-1	-1	-1	-1	+1	+1
9	-1	+1	-1	-1	+1	-1	+1

2) Calculate the values of the regression equation coefficients:

$$b_1 = 0.5(y_3 + y_4 - y_2 - y_5),$$

$$b_2 = 0.5(y_3 + y_5 - y_2 - y_4),$$

$$b_{12} = 0.5(y_2 + y_3 - y_4 - y_5).$$

Essentially, you should first find the arithmetical mean of the responses obtained in those tests, where $x_1 = +1$. From this value, subtract the average value of the responses obtained in those tests, $x_1 = -1$.

3) Calculate the effect of mean changing:

$$\Delta y = 0.2(y_2 + y_3 + y_4 + y_5 - 4y_0)$$

This is a simplified general expression

$$\Delta y = \frac{1}{N} \sum_{i=2}^N (y_i - y_1)$$

This expression should be used when the number of factors exceeds 2.

4) Calculate the values of the reproducibility variance in each test and the average value of the s_r reproducibility variance.

5) Calculate the errors of the regression equation coefficients.

For b_i and b_{ij} (k — the number of replications):

$$\varepsilon_i = \frac{2s_r}{\sqrt{k}}$$

For Δy :

$$\varepsilon_i = \frac{1.78s_r}{\sqrt{k}}$$

If all the regression equation members are insignificant, that is, their absolute value is less than or equal to the corresponding error, all tests should be performed in one more replication. Then the calculations should be repeated, repeating the tests. Repeat these actions until the significance of at least one member of the regression equation.

If at least one member of the regression equation is significant, it is necessary to choose the point with the optimal response value and build a new design, taking this point as its center.

Continue actions to achieve significance either until the desired result is achieved, or until the maximum possible values of the factors are reached.

If at least one member of the regression equation is significant, then the value Δy indicates the shape of the response surface:

- if Δy is significant and greater than zero, there is a minimum response within the designing domain;
- if Δy is significant and less than zero, the maximum response takes place within the designing domain;
- if Δy is not significant, then the response surface is either a plane or a saddle.

This technique was implemented when comparing the effectiveness of industrial preparations based on polyvinylpolypyrrolidone with a decrease in the content of phenolic substances during wine materials processing. [12].

Recommended from the point of view of the constructed regression models dosage regimes, exposure time at room temperature (20 °C) for all drugs: Polyclar 10 — dose 3.2 g/dm³, time 70 min; Polyclar VT — dose 4.6 g/dm³, time 25 min; Polyclar V — dose 4.6 g/dm³, time 25 min.

When constructing discriminating equations, a three-factor evolutionary A-optimal design with a cycle of three was used, which made it possible to obtain statistically significant differences in the preparation's effectiveness.

Using the equations, the optimal processing temperature of wine materials was calculated for the minimum and maximum dosages of each preparation recommended by the manufacturer, as well as the recommended processing time for wine materials — 5 minutes. The calculated range of optimum processing temperatures varied within a fairly wide range, from 6.6 °C to 17.2 °C.

After determining a specific experimental design, suitable for a certain optimization criterion, within the framework of experiment designing with the introduction of new identification criteria, the following tasks should be implemented [13, 14]:

- aggregation of new analytical data on modern identification techniques in the beverage industry;
- development of a dynamic system of cluster identification;
- obtaining a metadata set for the most common similar beverages;

- software development for the practical implementation of the cluster identification system for beverage groups;
- development of methodological principles for determining the directions of falsification in the industry, as well as a methodology for adjusting assessment criteria based on historical data.

4. Conclusion

Providing the population with high quality, safe and at the same time affordable food products is one of the socio-economic priorities of the Russian Federation. In connection with these priorities in the field of nutrition, the development of functional products

direction is observed, which is the first step towards personalized nutrition. The paper presents an algorithm for constructing a flexible experimental design for the development of new identification criteria, taking into account the peculiarities of laboratory research in the beverage industry. As part of the task implementation of obtaining analytical data using modern identification methods, a method is proposed for constructing experiments designs to identify the main factors and technological modes that have a dominant effect on the qualitative and physicochemical characteristics of beverages. On the basis of experiments, in the future, it is planned to build local models of new developed beverages using statistical methods for subsequent cluster analysis.

REFERENCES

1. Federal law of January 02, 2000 № 29-FZ “On quality and safety of food products”. Sz RF. — 2000. — № 2. (In Russian)
2. GOST R51293-99 “Identification of products. General principles”. Moscow: Standardinform. — 2018. — 4 p. (In Russian)
3. Technical regulation of the Customs Union TR CU021/2011 “On safety of food products” (adopted by The decision of the Council of the Eurasian economic Commission of December 9, 2011 № 880). Moscow, 2011. (In Russian)
4. Oganesyants, L. A., Khurshudyan, S. A., Galstyan, A. G., Semipyatny, V. K., Ryabova, A. E., Vafin, R. R., Nurmukhanbetova, D. E., Assembayeva, E. K. (2018). Base matrices — invariant digital identifiers of food products. Base matrices — invariant digital identifiers of food products. *NEWS of National Academy of Sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences*, 6(432), 6–15. <https://doi.org/10.32014/2018.2518-170x.30>
5. The strategy for improving the quality of food products in the Russian Federation until 2030, approved by Order of the Government of the Russian Federation no. 1364–29.06.2016 (In Russian)
6. Khurshudyan, S.A. (2008). Identifying signs of food products. *Food industry*, 11, 14.
7. Semipyatny, V. K., Ryabova, A. E., Egorova, O. S., Vafin, R. R. (2018). Experimental Modeling Optimization: New Beverages Recipes Composition by Statistical Approach. *Beer and beverages*, 3, 48–51. (In Russian)
8. Fetisov, E.A., Semipyatny, V. K., Petrov, A.N., Galstyan, A.G. (2015). Planning and analysis of the results of technological experiments. Moscow: Stalingrad. — 98 p. ISBN978-5-906565-21-1 (In Russian)
9. Harrington E. S. (1955). The desirability function. *Industrial Quality Control*, 21(10), 494–498.
10. Batuner L. M., Pozin M. Ye. (1963). Mathematical methods in chemical engineering. Leningrad: Goskhimizdat. — 639 p. (In Russian)
11. Vilanova, M., Zamuz, S., Tardaguila, J., Masa, A. (2008). Descriptive analysis of wines from Vitis vinifera cv. Albariño. *Journal of the Science of Food and Agriculture*, 88(5), 819–823. <https://doi.org/10.1002/jsfa.3157>
12. Panasyuk, A.L., Kuzmina, E.I., Egorova, O.S., Semipyatny, V.K., Nurmukhanbetova D. E. (2019). Ensuring the Stability of White Grape Wines Taste Characteristics with Use of Products Based on Polyvinylpyrrolidone. *NEWS of the Academy of Sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences*, 2(434), 174–180. <https://doi.org/10.32014/2019.2518-170x.52>
13. Petrov, A.N., Khanferyan, R.A., Galstyan, A.G. (2016). Current aspects of counteraction of foodstuff’s falsification. *Voprosy Pitaniia*, 85(5), 86–92. (In Russian)
14. Khurshudyan, S.A. (2013). Identification of the Packaging in the Definition of Counterfeit and Falsified Food Products. *Food Industry*, 1, 10–11. (In Russian)

AUTHOR INFORMATION

Irina I. Ageikina — junior research scientist, Interdisciplinary scientific and technical center of food quality monitoring, All-Russian Scientific Research Institute of Brewing, Beverage and Wine Industry — Branch of V. M. Gorbатов Federal Research Center for Food Systems of RAS, 119021, Moscow, Rossolimo str., 7. Tel. +7-499-245-61-18, E-mail: agira_ro.ru
ORCID: <https://orcid.org/0000-0002-5667-1335>

Ekaterina G. Lazareva — junior research scientist, Laboratory of Molecular Biology and Bioinformatics, V. M. Gorbатов Federal Research Center for Food Systems of RAS, 109316, Moscow, Talalikhina street, 26. Tel.: +7-499-245-61-18, E-mail: lkg1996@mail.ru
ORCID: <https://orcid.org/0000-0002-8069-9661>

Irina Yu. Mikhailova — research scientist, Interdisciplinary scientific and technical center of food quality monitoring, All-Russian Scientific Research Institute of Brewing, Beverage and Wine Industry — Branch of V. M. Gorbатов Federal Research Center for Food Systems of RAS, 119021, Moscow, Rossolimo str., 7. Tel.: +7-499-245-61-18, E-mail: vogts-villa@mail.ru
ORCID: <https://orcid.org/0000-0003-0514-505X>

Vladislav K. Semipyatny — candidate of technical sciences, senior research scientist, Interdisciplinary scientific and technical center of food quality monitoring, All-Russian Scientific Research Institute of Brewing, Beverage and Wine Industry — Branch of V. M. Gorbатов Federal Research Center for Food Systems of RAS, 119021, Moscow, Rossolimo str., 7. Tel.: +7-499-245-61-18, E-mail: semipyatniy@gmail.com

*corresponding author

ORCID: <https://orcid.org/0000-0003-1241-0026>

All authors bear responsibility for the work and presented data.

All authors made an equal contribution to the work.

The authors were equally involved in writing the manuscript and bear the equal responsibility for plagiarism.

The authors declare no conflict of interest.

Received 23.06.2020 Accepted in revised 26.08.2020 Accepted for publication 15.09.2020