Content is avaliable at SCOPUS

Meat Technology — Special Issue 66/3

www.meatcon.rs ▪ www.journalmeattechnology.com

UDK: 664.8/.9:004.4

ID: 179349001

https://doi.org/10.18485/meattech.2025.66.3.75

Original Scientific Paper

Generation and Pareto optimization of sterilized meat-and-plant canned food formulations using the R programming language

Ivan Yurov[1*], *Boris Kapovsky*[1] *and Marina Nikitina*[1]

[1] *V. M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences, Talalihina 26, Moscow, 109316, Russia*

ARTICLE INFO

Keywords:
Pareto optimization
Sterilized meat-and-plant
Canned food
R programming language

ABSTRACT

This study outlines key trends in modern software for food formulation development. An algorithm that enables the generation of a set of formulations for meat-and-plant sterilized canned products and identifies a Pareto-optimal subset thereof based on selected criteria is presented. The generation is carried out based on a developed mathematical model of a formulation that complies with the requirements of current standards. The algorithm was used to generate 1,000 formulations, of which 44 were identified as Pareto-optimal.

1. Introduction

The development of multicomponent food formulations is a labor-intensive process that requires both creativity and knowledge of the chemical composition and technological properties of ingredients. Additionally, it involves numerous calculations, particularly when accounting for coefficients and biological value indicators. Currently, specialized software has been developed that significantly simplifies and automates the formulation process, increases its speed, and eliminates potential errors in mathematical calculations (*Lisitsyn et al.*, 2021; *Musina et al.*, 2017). Most modern recipe formulation systems typically include the following features and functionality: built-in standardized parameters from current regulatory documentation for specific product categories; integrated databases containing chemical composition of ingredients and selected technological properties of ingredients; capability to calculate chemical composition and economic indicators of formulations; functionality for predicting certain technological properties of designed products; optimizing existing formulations; ability to maintain a registry of active formulations (*Khabibullin et al.*, 2014). Some software solutions also include functionality for generating labeling information blocks. Modern-generation systems are typically characterized by implementation of cloud computing technology (*Foodworks.online, n. d.*), utilization of artificial intelligence for recipe optimization (*FoodSolver, n. d.*), and prediction of product taste characteristics (*NIQ, n. d.*).

In Russia, the first software enabling the design of multicomponent food formulations was the “System for Designing and Assessing the Quality of Multicomponent Food Compositions”, developed

***Corresponding author:** Ivan Yurov, *i.yurov@fncps.ru*

Paper received July 31st 2025. Paper accepted August 12th 2025.
The paper was presented at the 63rd International Meat Industry Conference “Food for Thought: Innovations in Food and Nutrition” – Zlatibor, October 05th-08th 2025.
Published by Institute of Meat Hygiene and Technology – Belgrade, Serbia.

in 1997 by Academician Lipatov N. N. (Jr.) in collaboration with Bashkirov O. I. It was Academician Lipatov who laid the foundations for computer-aided design of multicomponent food products based on the theoretical framework of food combinatorics (*Lipatov and Rogov*, 1987; *Lipatov et al.*, 1990; *Lipatov et al.*, 2001). This software allows for calculating the chemical composition of a product, the minimum amino acid score, the utility coefficient, and the comparable redundancy coefficient. However, the program has significant limitations, such as the lack of a formalized ingredient composition database, the manual data entry for each ingredient, its restricted functionality, and the inconvenient user interface. These constraints substantially limit its applicability for broader scientific research and practical applications. Currently available Russian software for designing multicomponent food formulations includes Generic 2.0, MultiMeat Expert, Recipe Constructor, and others (*Droficheva*, 2023; *Nikitina et al.*, 2016). Internationally, widely used solutions are NutriCalc, Genesis R&D, and Emydex Recipe Formulation (*Khabibullin et al.*, 2014). However, key limitations of existing solutions are their inability to adapt to specific niche requirements, the lack of an option to add custom quality assessment parameters not pre-programmed in the software, and the paid access model.

A promising solution to personalize software functionality for food product development is by using programming languages, such as Python or R. Both languages are widely employed in research due to their open-source nature, extensive libraries for optimization and data visualization, relative simplicity of use, and abundance of reference and educational materials. The ultimate choice of language and environment depends on the user's individual preferences and skill level (*Hackenberger*, 2020).

The objective of this study was to perform automated generation of a set of synthetic formulations of sterilized meat-and-plant canned products using the R programming language based on a constructed mathematical model, and, through multicriteria optimization, to identify a subset of Pareto-optimal formulations.

2. Materials and methods

The object of this study is the process of generating a set of formulations for sterilized meat-and-plant canned products compliant with the Russian GOST 32245-2013 (*Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing*, 2013) standard and its Pareto optimization. The subject is an algorithm in the R programming language that implements the generation and optimization of formulations.

Formulation generation, optimization and data visualization were performed in the R programming language (R Core Team, 2025) version 4.4.3 (2025-02-28 ucrt) in the R-Studio environment (*Ushey et al.*, 2024). The following packages were used: NutrienTrackeR (*Rodriguez-Martinez et al.*, 2023), readxl (*Wickham and Bryan*, 2025), writexl (*Ooms*, 2025), ggplot2 (*Wickham*, 2016), and plotly (*Sievert*, 2020).

As the database (DB) of ingredient chemical composition, we used the Canadian Nutrient File (CNF) database of Health Canada, accessed through the R package NutrienTrackeR (*Rodriguez-Martinez et al.*, 2023).

Since it was necessary to find a subset of optimal formulations by more than one criterion in the generated set of formulations, we used Pareto optimization (*Nikitina and Chernukha*, 2018; *Becker et al.*, 2023). According to the definition of the Pareto-optimal set, in the set of formulations Y, a Pareto-optimal (dominant) formulation $y_i \in Y$ relative to formulation $y_k \in Y$ is one for which all evaluation criteria are no worse, and one of the criteria is better (Equation 1). In the study, we used Pareto dominance without weight coefficients for the criteria.

$$(y_i, y_k) \in Y \rightarrow \forall_j : [f_j(y_i) \geq f_j(y_k)] \Delta [f(y_i) \neq f(y_k)], \begin{cases} f_1(x) \rightarrow max \\ f_2(x) \rightarrow max \end{cases} \quad (1)$$

For optimization we selected the following three criteria: 1) amino acid adequacy criterion (the rationality coefficient of amino acid composition, RAAC (*Lisitsyn et al.*, 2021; *Lisin et al.*, 2012)). For the calculations, we used the amino acid composition reference standard from the *Food and Agriculture Organization [FAO]*, 2013); 2) the fatty acid adequacy criterion (Euclidean distance of fatty acids, EDFA) (Equation 2):

$$EDFA = \sum_{i=1}^{5} \sqrt{(j_{ref,i} - (i_{ji} \cdot 9 \cdot 100 \div E_j)^2} \quad (2)$$

where $j_{ref,\,i}$ – reference content of the i-th fatty acid fraction, %, according to MR 2.3.1.0253-21 (*Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing*, 2021); i_{ji} – content of the i-th fatty acid fraction in the j-th formulation, g/100g; E_j – energy value of the j-th formulation, kcal; 9 – energy conversion factor for fats; 100 – percentage conversion factor. Here, we calculate the distance between the reference point $j_{ref,i} = (p_1, p_2, p_3, p_4, p_5)$, where $p_1, \cdots, p_5$ are the reference percentages of fatty acid fractions relative to daily caloric intake, and the formulation point $i_j = (q_1, q_2, q_3, q_4, q_5)$, where $q_1, \cdots, q_5$ are the actual percentages of fatty acid fractions relative to the formulation's caloric value. Thereby, formulations with minimal EDFA values are nutritionally closer to the reference fatty acid ratio; 3) formulation cost criterion, calculated as the sum of average ingredient costs. For each ingredient, at least three price values were obtained.

Since equation (1) represents a solution to a two-criterion maximization problem, we reformulated the Pareto-optimality condition for the three selected minimized criteria (equation 3):

$$(y_i, y_k) \in Y \rightarrow \forall_j : [f_j(y_i) \leq f_j(y_k)] \Delta [f(y_i) \neq f(y_k)], \begin{cases} f_1(x) \rightarrow min \\ f_2(x) \rightarrow min \\ f_3(x) \rightarrow min \end{cases} \quad (3)$$

where f_1 – RCAAC criteria, which ought to be minimized; f_2 – EDFA criteria, which ought to be minimized (2); f_3 – Formulation cost criteria, which ought to be minimized.

The mathematical model for meat-and-plant canned food formulations (equation 4) was developed in compliance with GOST 32245-2013 (*Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing*, 2013).

$$\begin{cases} \sum_{i \in I} x_i = 1, \\ 0{,}3 < \sum_{i \in I_{meat}} x_i \leq 0{,}6, \\ \sum_{i \in I} x_i \cdot p_i \geq 0{,}06, \\ \sum_{i \in I} x_i \cdot f_i \leq 0{,}25, \\ x_i \geq 0 \ \forall i \in I. \end{cases} \quad (4)$$

where I – set of all canned food formulation ingredients; x_i – mass fraction of the i-th ingredient in the formulation, %; $I_{meat} \subset I$ – subset of all meat ingredients; p_i – protein mass fraction of the i-th ingredient in the formulation, %; f_i – fat mass fraction of the i-th ingredient in the formulation, %.

3. Results and discussion

The algorithm for generating a set of formulations and their Pareto optimization consisted of three blocks (Figure 1). The input data for the algorithm is an XLSX file containing a list of selected ingredients and their chemical composition (protein, fat, carbohydrates, vitamins, micro- and macronutrients, fatty acids, etc.; 152 parameters total). In Block 1, data are prepared for subsequent calculations: the contents of methionine with cysteine and phenylalanine with tyrosine are summed, as required by the reference standard (*FAO*, 2013).

Block 2 starts with the determination of the number of recipes to be generated and the ingredient group type, which specifies the mechanism for assigning mass fraction to an ingredient. There are two types of this mechanism. For some ingredients, such as salt and pepper, the mass fraction is constant in every formulation. For others, the mass fraction varies. For example, equation 4 assumes the content of meat ingredients is in the range of 30.00 to 60.00%. The mass fraction within specified ranges is determined using a pseudorandom number generator (PRNG). The second subtype of the variable mechanism is when the mass fraction of an ingredient is defined by another ingredient. For example, the type of plant ingredient defines the amount of added water by a specific coefficient. We then established rules for selecting ingredients from different groups or ingredients. It was assumed that one or two meat ingredients and one or two plant ingredients could be selected in each recipe. After this, the process of assigning mass fractions to the ingredients starts. First of all, obligatory ingredients with constant mass fraction are assigned. Then, the mass fractions of ingredients determined by the first subtype of the second mechanism (meat ingredients, fat ingredients). The unoccupied reminder of one (first line in equation 4) is distributed by the second subtype of the second mechanism between the plant ingredients and water. The chemical composition of the recipes is calculated by matrix multiplication of the established mass fractions and chemical composition of the ingredients. Formulations falling outside the acceptable range, as well as those with an amino acid score below 100%, are excluded from further consideration.

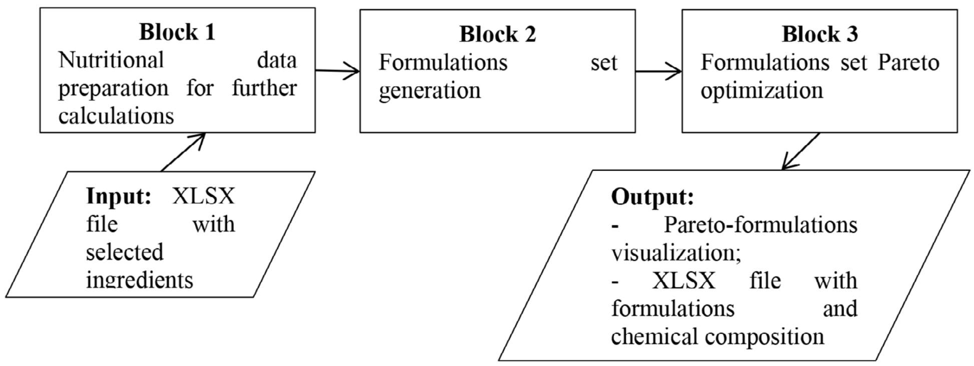

Figure 1. Algorithm for generation and Pareto optimization of meat-and-plant canned food formulations

Block 3 performs Pareto optimization of the generated set of formulations (equation 3) using three minimization criteria. The Pareto-optimal subset is then visualized in a 3D plot using the ggplot2 (Wickham, 2016) and plotly (*Sievert*, 2020) packages. Figure 2 displays 1,000 generated formulations, of which 44 were Pareto-optimal. The block outputs an XLSX file with four sheets: complete set of generated formulations; chemical composition of each formulation; Pareto-optimal formulations; chemical composition of Pareto-optimal formulations. Two examples of Pareto-optimal formulations are presented in Table 1, while Table 2 shows the optimization criteria values.

Table 1. Examples of generated Pareto-optimal formulations.

Ingredient	Mass fraction, %	
	Formulation 1	Formulation 2
Pork heart	31.65	–
Pork liver	–	15.93
Turkey liver	–	14.08
Millet groats	23.25	20.92
Buckwheat groats	–	10.05
Onion	2.87	2.36
Pork fat	4.97	8.16
Vegetable oil	3.28	–
Salt	1.40	1.40
Black pepper	0.04	0.04
Water	32.54	27.06
TOTAL	100.00	100.00

Table 2. Values of optimization criteria for two generated Pareto-optimal recipes

Criterion	Formulation 1	Formulation 2
Rationality coefficient of amino acids (RAAC)	30.25 %	29.72 %
Euclidean distance of fatty acids (EDFA)	13.04	10.51
Formulation cost per 100 g	6.81 RUB	8.73 RUB

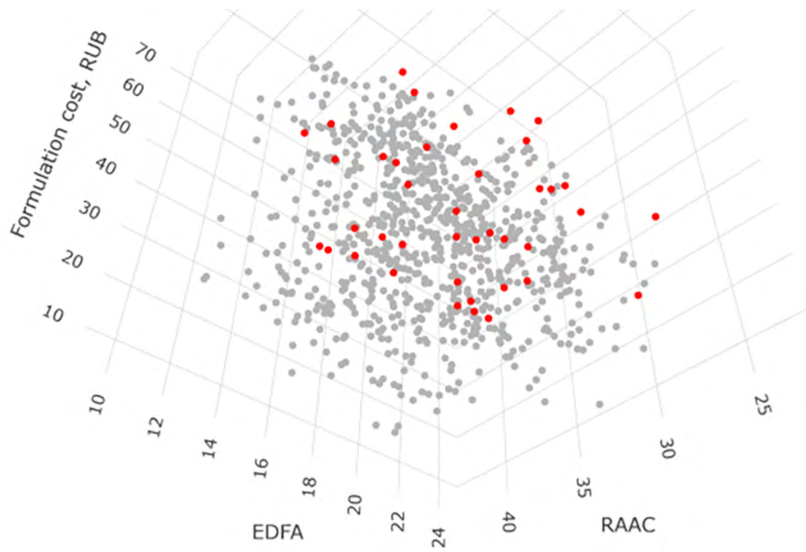

Figure 2. 3D visualization of all generated (gray points, N = 1,000) and Pareto-optimal (red points, N = 44) formulations across three optimization criteria, i.e., the cost (in rubles), the rationality coefficient of amino acids (RAAC), and the Euclidean distance of fatty acids (EDFA).

4. Conclusion

Food formulation development involves various costs, including time, financial resources, and other expenditures. Modern computational technologies help mitigate these costs by enabling rapid processing of large datasets with specified precision. Depending on the objectives, practitioners can utilize either specialized software or broader programming tools. The latter approach allows for personalization of the formulation process and incorporation of custom quality metrics during multi-criteria optimization of the final product.

The R-based algorithm presented in this study generates multiple meat-and-plant canned food formulations compliant with regulatory standards for this product category. Through multi-criteria optimization, it identifies a Pareto-optimal subset of formulations based on selected criteria.

References

Becker, D., Schmitt, C., Bovetto, L., Rauh, C., McHardy, C., & Hartmann, C. (2023). Optimization of complex food formulations using robotics and active learning. *Innovative Food Science & Emerging Technologies*, 83, 103232. https://doi.org/10.1016/j.ifset.2022.103232.

Droficheva, N. V. (2023). The use of digital technologies in the food and processing industry]. *Plodovodstvo i vinogradarstvo Yuga Rossii,* 84, 209–235. https://doi.org/10.30679/2219-5335-2023-6-84-209-235

Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing, (2021). Metodicheskie rekomendatsii MR 2.3.1.0253-21 [Methodological guidelines MR 2.3.1.0253-21 «Norms of physiological needs in energy and nutrients for various population groups of the Russian Federation»].

Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing. (2013). (GOST 32245-2013) [Meat-containing canned food. General specifications (GOST 32245-2013)]. Standartinform.

FAO—Food and Agriculture Organization of the United Nations. (2013). *Dietary protein quality evaluation in human nutrition: report of an FAO expert consultation, 31 March-2 April, 2011, Auckland, New Zealand.*

FoodSolver. (n. d.). Your virtual food lab. Accessed May 17, 2025, from https://www.foodsolver.ai/

Foodworks.online. (n. d.). The gold standard for nutrition analysis. Accessed May 17, 2025, from https://foodworks.online/

Hackenberger B. K. (2020). R software: unfriendly but probably the best. *Croatian Medical Journa*l, 61(1), 66–68. https://doi.org/10.3325/cmj.2020.61.66

Khabibullin, R. E., Zhackslykova, S. A., & Reshetnik, O. A. (2014). Sovremennye programmy avtomatizirovannogo rascheta retseptur kombinirovannykh pishchevykh produktov [Modern software for automated calculation of combined food product formulations]. *Vestnik Kazanskogo Tekhnologicheskogo Universiteta,* 21, 257–260.

Lipatov, N. N., & Rogov, I. A. (1987). Metodologiya proektirovaniya produktov pitaniya s trebuemym kompleksom pokazateley pishchevoy tsennosti [Methodology for designing food products with required nutritional value indicators]. *Izvestiya vuzov. Pishchevaya tekhnologiya,* 2, 9–15.

Lipatov, N. N. (1990). [Principles and methods of designing food product formulations that balance diets]. *Izvestiya vysshikh uchebnykh zavedeniy. Pishchevaya tekhnologiya*, 199(6), 5–10.

Lipatov, N. N., Sazhinov, G. Yu., & Bashkirov, O. N. (2001). Formalized analysis of amino and fatty acid balance of raw materials for designing baby food products with specified nutritional adequacy]. *Khranenie i pererabotka selkhozsyr'ya,* 8, 11–14.

Lisin, P. A., Moliboga, E. A., Kanushina, Yu. A., & Smirnova, N. A. (2012). Evaluation of amino acid composition of food product formulations]. *Agrarnyy vestnik Urala*, 3(95), 26–28.

Lisitsyn, A. B., Chernukha, I. M., & Nikitina, M. A. (2021). Design of multicomponent food products: Textbook]. ROSBIOTEKH. https://e.lanbook.com/book/277133 (Access limited to authorized users)

Musina, O., Putnik, P., Koubaa, M., Barba, F. J., Greiner, R., Granato, D., & Roohinejad, S. (2017). Application of modern computer algebra systems in food formulations and development: A case study. *Trends in Food Science & Technology,* 64, 48–59. https://doi.org/10.1016/j.tifs.2017.03.011

Nikitina, M. A., & Chernukha, I. M. (2018). Multicriteria optimization of product formulation composition]. *Teoriya i praktika pererabotki myasa,* 3(3), 89–98. https://doi.org/10.21323/2414-438X-2018-3-3-89-98.

Nikitina, M. A., & Sus, E. B. (2016). Konstruktor retseptur – IT-reshenie ot VNIIMP [Recipe constructor – IT solution from VNIIMP]. *Vse o Myase*, 6, 53–55.

NIQ. (n. d.). BASES AI Product Developer: Sensory & Product Development. Retrieved May 17, 2025, from https://nielseniq.com/global/en/solutions/bases-ai-product-developer/

Ooms, J. (2025). _writexl: Export Data Frames to Excel 'xlsx' Format_. R package version 1.5.3. Retrieved May 10, 2025, from https://CRAN.R-project.org/package=writexl

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved May 10, 2025, from https://www.R-project.org/

Rodriguez-Martinez, A., Ayala, R., Balchunas, M., Gutierrez-Del-Rio, D., Chen, T., Debbabi, Y., & Selles Vidal, L. (2023). _NutrienTrackeR: Food Composition Information and Dietary Assessment_. https://doi.org/10.32614/CRAN.package.NutrienTrackeR

Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC.

Ushey, K., Allaire, J., Wickham, H., Ritchie, G. (2024). _rstudioapi: Safely Access the RStudio API_. R package version 0.17.1. Retrieved May 10, 2025, from https://CRAN.R-project.org/package=rstudioapi

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.

Wickham, H., Bryan, J. (2025). _readxl: Read Excel Files_. R package version 1.4.5. Retrieved May 10, 2025, from https://CRAN.R-project.org/package=readxl

Authors info

Ivan Yurov, https://orcid.org/0009-0004-4404-6489
Boris Kapovsky, https://orcid.org/0000-0003-0256-8655
Marina Nikitina, https://orcid.org/0000-0002-8313-4105