# STUDYING IDENTIFICATION ALGORITHMS IN FOOD TECHNOLOGY

**Oksana Takhumova[1], Tatiana Bondarenko[2], Sergei Shlykov[3], Ruslan Omarov[3]**
[1]Kuban State Agrarian University named after I.T. Trubilin, Russia;
[2]Kuban State Technological University, Russia; [3]Stavropol State Agrarian University, Russia
scopush@gmail.com

**Abstract.** Food production specificity requires improvement of systems of automatic regulation of the processes in aggregates, apparatus, and facility. In creating an adaptive process control system for food production, based on the model of a control object, an additional analysis should be done to choose the identification algorithm of a real, fairly representative sample of input data and output measurement results. By now a lot of recurrent identification algorithms are development and in one or another extent investigated. The little part of these algorithms is implemented into software, such as Matlab, Unity Pro, and other similar ones. But the problem is that among all of these algorithms the best one for real conditions of identification is absent. And a practicing engineer, who takes up the implementation of an adaptive control system with an identifier, is faced with the problem of choosing a specific identification algorithm. In this research, by using simulation modelling, over 53 recurrent identification algorithms were analyzed, plus the main modifications of these algorithms, and a total of 47 possible estimation criteria for non-stationary multidimensional dynamic objects. Based on this analysis of the object class in question, several algorithms were recommended to engineering practice. Possibilities of a software package that has the most complete set of parametric identification algorithms today are discussed. For these specific conditions of the identification algorithms comparison into the package by identifying the stationary coefficients into object equation, the most effective were: Aizerman-1, Kachmazh, Nagumo-Noda, Rastrigin, Kalman Filter, Forgetting Factor, Tsypkin. With non-stationary coefficients of the object - Kachmazh, Nagumo-Noda, Kalman Filter. The best result was shown by the Nagumo-Noda identification algorithm.

**Keywords:** algorithms, parameter estimation, parametric models, discrete signals.

## Introduction

When dealing with meaningful representations of food systems, several important issues have to be considered: data can be plagued by uncertainty, particularly, when chemical, physical, and biological phenomena concur to define the process; heterogeneity of available information is also likely, as a vegetable involved in a process can be characterized by more than 40,000 genes, whereas the quality of the final product can be assessed using just a few sensory features; qualitative and quantitative information, from expert perception of food quality, to nano-properties of ingredients, may also coexist in the same process. Consequently, when applying machine learning to agri-food data, the user has to carefully account for variance, manage heterogeneous data and be able to include both qualitative and quantitative values in the final model.

As gathering data in food science is an expensive and time-consuming process, available datasets are often sparse and incomplete, which poses a challenge to both human modelling practitioners and machine learning algorithms. This issue has been long acknowledged by the community, and ongoing projects have been approved to tackle it, by defining roadmaps to achieve an e-infrastructure for open science 1, and by fostering cooperation between food scientists and modelling experts 2. In order to obtain reliable models, it thus becomes necessary to acquire additional information from external sources. Experts in a specific domain can provide invaluable insight into products and processes, but this precious knowledge is often available only in the form of intuition and non-coded expertise. Including expert insight in a model is not a straightforward process, but it can effectively be tackled by having humans interacting with a machine learning process, through visualization, or via specialists in encoding implicit domain knowledge [1-3].

One successful approach in modelling complex processes is to stack smaller models, such that predictions are propagated between multiple layers formed by these sub-components [4;5]. In such cases, typically, rich datasets and vast amounts of knowledge are available to describe the stacked components and their interactions. When little data are available, and prior knowledge is limited, mathematical regression techniques can be used to model these complex systems [6;7]. However, a multitude of candidate models can be obtained through these techniques. Deciding which of these models is the best with respect to the study domain and problem at hand, may be carried out automatically based on the fitness criteria, or delegated to domain experts [8;9]. While the former is efficient, but can result in models that do not capture the reality of the underlying system, the former

may be grounded albeit time-consuming. Similarly to Turkay et al. [10;11], our approach uses mathematical regression to generate candidate solutions. However, we combine automatic evaluation of candidate models with expert evaluations to ensure both model robustness and validity [12].

Food production specificity requires systems improvement for automatic regulation of the processes in devices, units, and installations. This technological equipment composes the complex control objects with a significant time lag, clearly marked critical boundaries, near which it is necessary to run a process with variable parameters [13].

One of the most effective types of object management systems with variable parameters is adaptive systems with an identifier. The identification block in these systems continuously or periodically refines the parameters of the control object model. It means that the identification block realizes hardware or software of the recurrent identification algorithms.

However, identification algorithms that are designed and used to estimate unknowns, but unchanged in time parameters in terms of modelling for independent input signals and other explicit and implicit prior conditions, in practice, show far from the best results. Non-stationarity of parameters, their interrelation, changing probabilistic characteristics of input signals can lead to theoretically difficult predictable results of building models, such as production facilities.

Today a large number of recurrent algorithms for identification are developed and in varying degrees investigated. Identification algorithms are scattered across numerous monographs, special periodicals, and technical reports. A far from complete bibliography is given in this work [14]. The modest part of these algorithms set is realized in the form of the software tools, such as Matlab, Unity Pro, and other similar systems. The problem is that among all of these algorithms there is not the best one for real identification conditions.

A practicing engineer realizing an adaptive control system with an identifier faces the problem of choosing a specific identification algorithm. In such control systems the identifier allows obtaining models of a given quality for control of a model, when by incoming data, measurement results the model parameters are refined.

## Materials and methods

In this research, based on simulation modelling, over 53 algorithms for recurrent identification were analyzed, plus the main modifications of these algorithms with a capacity to evaluate the operation of an identifier by a large number of criteria, in total, up to 47 criteria for non-stationary multidimensional simply connected dynamic objects.

Under uncertainty, iterative or recurrent identification algorithms got wide application, possessing a number of advantages: easy to implement, significant speed, the possibility to obtain consistent estimates. Some design features in the iterative algorithms are introduced under the process of their direct implementation in control systems, which is associated with the features and properties of the process or object under study.

As mentioned above, a number of recurrent identification methods are implemented in the well-known Matlab software package. In the Matlab package it is possible to obtain estimates of the parameters of an object model by direct programming, i.e. using the coefficient polynomial calculation function of ARMAX or by the Simulink subsystem and System Identification Toolbox. However, neither the package as a whole, nor its individual parts, even the algorithms, are inappropriate to use on industrial controllers in adaptive control systems. Moreover, the size and packet cost are not the main reasons to refuse using it in industrial regulators. All Matlab tools work with one-dimensional single-connected objects. In general, the package is not focused on obtaining current estimates of parameters and their analysis in real time. In the manuscripts [15, 16] it was proposed to expand the capabilities of Matlab by including the Recursive Identification Algorithms Library package in the Simulink subsystem. However, only the release of the Matlab package R2013b in the System Identification Toolbox presented three algorithms of recurrent identification.

In addition, the identification algorithms used in Matlab: the forgetting factor and the Levenberg-Marquard algorithm [17, 18], on different samples do not always show good results for non-stationary objects in comparison with other algorithms. This means that there is no universal best algorithm for all identification conditions.

Lacking clear recommendations for the use identification algorithms is apparently one of the reasons that the leading manufacturers of industrial controllers and software for them prefer to use in their systems classic PID controllers or self-tuning PID controllers [19]. However, for control systems of non-stationary objects, such controllers lose the accuracy of settings and often require a heuristic selection of settings, i.e. the operator presence [20].

Almost all considered identification algorithms constitute a nonlinear function, so obtaining analytical estimates for most algorithms causes considerable difficulties. In addition, for practical applications important parameters are high order of the differential or difference equation and an arbitrary number of inputs. Therefore, the comparison of the algorithms was performed by numerical simulation. To obtain reliable convincing results based on numerical modelling, it is necessary, on the one hand, to be able to obtain quite diverse input signals and the ability to consider various types of parametric nonstationarity. For this purpose, the software package that allows performing the task was created [21].

The software package provides the ability to either simulate input signals for probability distributions: uniform, normal, gamma, beta, exponential, Laplace, Cauchy, or using data from a real control object. Implemented algorithms for identifying coefficients in the autoregressive equation [22]: algorithms based on P- and SP-approaches, Aizerman, Kacmage, Nagumo-Noda, MNC, RMNK, Avedian, Rastrigin algorithms, Zipkin algorithms, Kalman filter, forgetting factor algorithm, Levenberg-Marquard algorithm, and others. A total of 53 identification algorithms were analyzed, plus the main modifications of these algorithms.

To assess the quality of identification using various adaptation criteria: by output and parameters, point and integral, convergence rate estimates, dispersion error ratio, error value, rms estimates — it is only possible to use 47 criteria.

In a general case, the object for comparison of identification algorithms was understood as a technological process model, which is described as a multidimensional simply connected object obtained by decomposition from a multidimensional multi-connected object with observed inputs $x_r r = \overline{1, L}$ and output $y$ [22]:

$$y(n) = \sum_i C_j(n) X(n-i) + \sum_j A_j(n) y(n-j) + v(n), \tag{1}$$

where   $C_j^T(n) = (C_1(n), C_2(n), \ldots, C_L(n))$  – vector of unknown parameters under input variables;

$A^T(n) = (a_1(n), a_2(n), \ldots, a_K(n))$  – vector of unknown parameters with output variables;

$X(n) = \sum_{r=1}^{L} x_r(n), v(n)$ – interference.

The model for an object given by equation (1) is usually written in a more compact form:

$$y_n = \varphi_n^T \theta_{n-1}, \tag{2}$$

where

$$\varphi_n^T = \begin{pmatrix} x_1(n), x_2(n), \ldots, x_L(n), x_1(n-1), x_2(n-1), \\ \ldots, x_L(n-1), \ldots, x_1(n-I), x_2(n-I), -\ldots, x_L(n-I) \end{pmatrix}$$
$$y(n-1), y(n-2), \ldots, y(n-K))$$

is generalized vector of inputs;

$$\theta_{n-1}^T = (\hat{c}_1(n-1), \hat{c}_2(n-1),\ldots,\hat{c}_L(n-1), \hat{c}_1(n-2), \hat{c}_2(n-2),\ldots,$$

$$\hat{c}_L(n-2), \hat{c}_1(0), \hat{c}_2(0)\ldots,\hat{c}_L(0), \hat{a}_2(n-1),\ldots \hat{a}_K(n-1))$$

is vector of estimated parameters at the time step (n-1), respectively;

$\hat{C}_j(n-i)$ – estimates of the coefficients by $L$ inputs $x_j(n)$ for the time lag to $I$;

$\hat{a}_i$ – estimates of the coefficients by the outputs at the corresponding moments of time.

**Results and discussion**

The paper analyzes the convergence of the object model coefficients by assessing the identification quality in the case of stationarity or non-stationarity of the multidimensional object with ten inputs, using various identification algorithms that were implemented in a software package with a stable object. At the output of the object additive interference is added.

Figure 1 shows the object and model coefficients in the case of periodic non-stationarity, when identified by the Nagumo-Noda algorithm, in comparison to Figure 2 – the identification algorithm based on the Kalman filter.
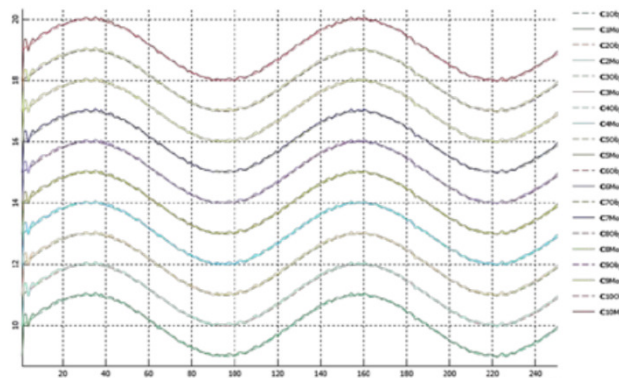


Fig. 1. **Coefficients of object $C_{oi}$ and model $C_{mi}$:** Nagumo-Noda Identification Algorithm
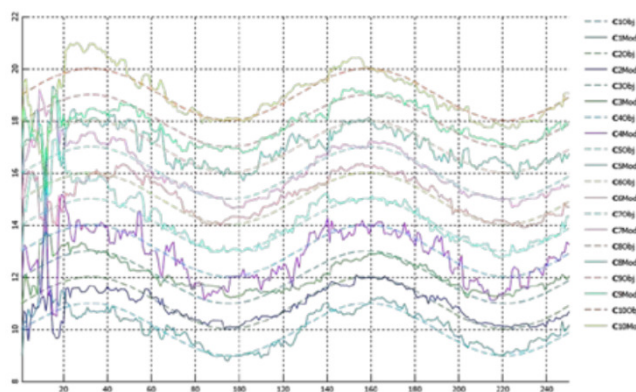


Fig. 2. **Coefficients of object $C_{oi}$ and model $C_{mi}$:** identification based on Kalman filter

Considering this model example, we would like to draw attention to the obvious conclusion, which follows from the qualitative analysis of the tracking results for the parameters. The identification algorithm based on the Kalman filtering, specially designed for the recurrent estimation the non-stationary systems, under certain conditions loses accuracy tracking to Nadum-Noda identification algorithm.

For the methodology of building adaptive systems with an identifier this conclusion makes it necessary to use under design more than one unique estimation algorithm, no matter how good it is according to the algorithm developer's conclusion, and the whole set of algorithms (53 algorithms [6]

are collected in our package) for real identification conditions. Choosing the best algorithm can proceed the hardware-software implementation of the control system.

## Conclusions

1. For our specific conditions of identification algorithms in the package by identifying the stationary coefficients in the equation of the object (1), the most effective were: Yzerman-1, Kachmazh, Nagumo-Noda, Rastrigin, Kalman Filter, Forgetting Factor, Tsypkin.
2. At non-stationary coefficients of the object - Kachmazh, Nagumo-Noda, Kalman Filter.
3. The best result was shown by the Nagumo-Noda identification algorithm.

## References

[1] Tonda A. et al. Interactive Machine Learning for Applications in Food Science. In: Zhou J., Chen F. (eds) Human and Machine Learning. Human-Computer Interaction Series. Springer, Cham. 2018.

[2] Lutton E., Tonda A., Boukhelifa N., Perrot N. Complex systems in food science: Human factor issues. In Jan Van Impe, editor, FoodSIM. EUROSIS-ETI, 2016.

[3] Koza J.R. Genetic programming: on the programming of computers by means of natural selection, volume 1. MIT press, 1992.

[4] Cros M.-J., Duru M., Garcia F., Martin-Clouaire R. A biophysical dairy farm model to evaluate rotational grazing management strategies. Agronomie, 2003. vol. 23. no. 2. pp. 105-122.

[5] Baudrit C., Sicard M., Wuillemin P.-H., and Perrot N. Towards a global modelling of the camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic bayesian networks. Journal of Food Engineering, 2010. vol. 98. no. 3. pp. 283-293.

[6] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., et al. Scikit-learn: Machine learning in python. Journal of Machine Learning Research, 2011. vol. 12(Oct). pp. 2825-2830.

[7] Chabin T., Barnabe M., Boukhelifa N., Tonda A., Velly H., Lemaitre B., Perrot N., Lutton E. Lideogram: An interactive evolutionary modelling tool. In Proceedings of the International Conference on Artificial Evolution (Evolution Artificielle), 2017.

[8] Krause J., Perer A., Bertini E. Infuse: interactive feature selection for predictive modelling of high dimensional data. IEEE transactions on visualization and computer graphics, 2014. vol. 20. no. 12. pp. 1614-1623.

[9] Cheng J., Bell D.A., Liu W. An algorithm for bayesian belief network construction from data. In proceedings of AI & STAT97, 1997. pp. 83-90.

[10] Turkay C., Slingsby A., Lahtinen K., Butt S., Dykes J. Supporting theoretically-grounded model building in the social sciences through interactive visualisation. Neurocomputing, 2017.

[11] De Jong K.A. Evolutionary computation: a unified approach. MIT press, 2006.

[12] Krause J., Perer A., Bertini E. Infuse: interactive feature selection for predictive modelling of high dimensional data. IEEE transactions on visualization and computer graphics, 2014. vol. 20. no. 12. pp. 1614-1623.

[13] Isermann R., Münchhof M. Identification of Dynamic Systems. An Introduction with Applications. Berlin, Springer-Verlag, 2011. 705 p.

[14] Bobál V., Böhm J., Fessl J., Macháèek J. Digital Self-tuning Controllers: Algorithms, Implementation and Applications. London, SpringerVerlag, 2005. 318 p.

[15] Ljung L. System Identification Toolbox: For use with MATLAB. Prentice-Hall, Englewood Cliffs, N.J., 2004. 830 p.

[16] Navrátil P., Bobál V. Recursive identification algorithms library. Proceedings17th International Conference on Process Control, 2009. pp. 516-523.

[17] Благовещенская М.М., Злобин Л.А. Информационные технологии систем управления технологическими процессами. (Information technologies for Process Control Systems). Moscow, Vysshaia shkola, 2005. 768 p. (In Russian).

[18] Александров А.Г., Паленов М.В. Самонастраивающийся ПИД/И-регулятор. (Automation and teleautomatics). 2011, vol. 72, no. 11, pp. 2010. (In Russian).

[19] MATLAB and Simulink for Technical Computing. Documentation on the program MATLAB 2011. MathWorks. 2011. [online][16.02.2019] Available at: http://www.mathworks.com/.

[20] Unity PRO. Control Block Library. Schneider Electric, 2009, no. 4. [online][16.02.2019] Available at: http://www.schneider-electric.com/

[21] Макаров В.В., Михеев А.С. Имитационная система моделирования адаптивной системы управления с идентификатором. Свидетельство о государственной регистрации программ для ЭВМ №2013610594. (Simulation system of adaptive control system with identification). Certificate of state registration computer program №2013610594. (In Russian).

[22] Жиров М.В., Макаров В.В., Солдатов В.В. Идентификация и адаптивное управление технологическими процессами с нестационарными параметрами (Identification and adaptive Control for Time-varying Technological Process). Bauman State Technical University, 2011. 203 p. (In Russian).