

Issues of Improving the Accuracy of Demand and Sales Forecasting Using Decomposition of Components and Fuzzy Error Estimation

Leonid Mylnikov

*Department of Information
Technologies and Automation Systems
Perm National Research Polytechnic
University
Perm, Russia
leonid.mylnikov@pstu.ru*

Dmitrii Vershinin

*Microprocessor Automation Means
Department
Perm National Research Polytechnic
University
Perm, Russia
nsenz@yandex.ru*

Artur Mikhailov

*Microprocessor Automation Means
Department
Perm National Research Polytechnic
University
Perm, Russia
mihailovarthur@rambler.ru*

Abstract—The efficiency of production and sales systems focused on public markets and the enlargement of variable parts of the orders depends on the accuracy of demand and forecasting and planning of production volumes. To tackle the problem of improving the forecasting accuracy of time series in this paper we have tested the hypothesis that the parameters associated with the flow of orders contain several components that can be described separately using existing approaches. Hence, the forecasting error can be represented as a set of fuzzy numbers. Hence, the forecasting error can be represented as a set of time series' fuzzy numbers. As a result of the hypothesis investigation, we obtained forecasting data in its fuzzy form, which already contains results of possible deviations and their probability. Moreover, this method of using fuzzy numbers may improve the accuracy of forecasting. The use of fuzzy forecasts allows us to solve planning and management problems in its fuzzy formulation and thereby to obtain results containing assessments of the range of possible deviations, risks and possible strategies of behaviour without additional research.

Keywords—*planning, forecast, regression, deviations, fuzzy number, time series decomposition, decision making support.*

I. INTRODUCTION

During the process of solving different problems of production management and economic systems tasks, there are problems with obtaining information about the behaviour of the external environment can appear. It is important since we cannot control its behaviour, but having known the characteristics of the behaviour we can find the optimal state of the system performance of which we want to increase (adjust the state of the ruling object to the external environment behaviour).

Working in the area of this paradigm, the quality of decisions made depends on the accuracy of the forecasts used. For production and sales systems, the parameters characterizing the external environment are sales and changes in product prices. To solve the planning problems, we need to obtain the values of these parameters for upcoming periods.

Despite the simple idea related to the use of forecasts, the solution of the prediction problem has lots of difficulties in preparing the necessary data and selecting the required data

volume, the behaviour distinction of the predicted parameters, in the dependence on the factors which cannot be observed by decision makers, the influence of the competitors behaviour on the parameters of the market environment and so on.

As a result, these days we can see a variety of methods and approaches developed for making predictions of time series. To solve the problems of predicting the parameters values represented as time series, well-known patterns that take into account dependencies of the parameters under consideration [1] or S-curves), classical regression methods (methods, based on proximity measures), recurrent methods (such as autoregression (ARIMA) based on moving averages, weighted average, exponential smoothing, Bayesian analysis and so so), fractal forecasting methods [2], methods of machine learning [3] are frequently used. There are also well-known approaches based on the use of retrospective data periods (windows) (called reinforcement learning for time series) which allow to give models the property of adaptability, methods based on the data decomposition into separate components (for example, decomposition of time series into harmonic components (Fourier series, wavelet analysis [4]), decomposition into trends and periodic components – seasonal and trend decomposition using Loess (STL) [5] or into trends and oscillations – singular spectrum analysis (SSA) [6]), the use of patterns with well-studied processes (for example, using dynamic time wrapping method [7]), adaptive approach for values periods [8]. To forecast a group of independent parameters, approaches such as partial least squares (PLS) [9], lasso [10], neural networks, tensor train networks are frequently used [11].

The other task is the selection of criteria for evaluating forecasting methods (for example, R², root-mean-square error (RMSE), mean absolute error (MAE)) and determination, based on them, methods that show the best performance in certain situations. In the paper [3] it is shown that for problems of forecasting crime rates, forest fire, drop of interest rates the best results show such methods as generalized linear regression, projection methods, lasso; for air quality, cars prices, crime rates, daily products demand predictions --- generalized linear regression, lasso, Bayesian models; to predict news ratings, air pollution rates we can

use regression rules, random forests, bagging. Thus, the choice of a particular forecasting method directly depends on the data and the method for evaluating the results. Hence, this kind of forecasting methods is pretty much versatile. For example, we can make predictions of values, characterized by instant values and average estimations of statistical data. As a result, we have the task of developing approaches instead of methods to be able to work with the data that allow us to obtain accurate forecasts for management and planning tasks.

One of the approaches to formalizing both subjective component and the uncertain nature of predictions is the method that uses membership function to describe the stochastic nature of the data.

The development of forecasting methods is currently on the way of obtaining hybrid methods (methods based on the algorithms of sequential or joint use of existing approaches to adjust the predictions obtained).

II. METHODOLOGY

The value of the predicted parameter $y(t)$ is in form $y(t) = g(t) + \varrho$, where $g(t)$ - the function that will be replacing the statistical data and will be used to predict their values in future periods, ϱ - error value.

The forecast function will be represented as $g(t) = g_{\text{periodic}}(t) + g_{\text{trend}}(t) + g_{\text{rest}}(t, t_G) + g_{\text{interference}}(t)$, where $g_{\text{periodic}}(t)$ - periodic component, $g_{\text{trend}}(t)$ - trend component, $g_{\text{rest}}(t)$ - residual component of the forecast, $g_{\text{interference}}(t)$ - component that allows us to take into account the mutual influence of the parameters while considering control objects with several parameters, t_G - time frames for making predictions.

Each of the components can consist of several components (for example, it is possible that

$g_{\text{periodic}}(t) = \bigoplus_{i=1}^n g_{\text{periodic},i}(t)$, where n number of components) as well as being absent depending on the dataset.

In order to describe the value of ϱ_j , we check its compliance with one of the distributions. As a result, we determine the distribution to which error (random component) fits best and which will be used to build the membership function during the translation into a fuzzy form.

If the error fits ϱ_j the Gaussian distribution, we can use Gaussian function to translate forecast data $g_j(t)$ into fuzzy form (fuzzification): where $m(g_j(t))$ - membership function, $g_j(t)$ - the forecast value for j parameter, x_g - the predicted value deviations range.

Errors transformation into its fuzzy form allows us to translate each predicted value into the fuzzy form by taking

as a distribution function the membership function for ϱ_j and as a result receiving a table-defined fuzzy function.

The presence of periodic components is determined by life cycles (see fig.1) which are obtained using the retrospective data and represented either as table-defined functions or if there are gaps, as regression models.

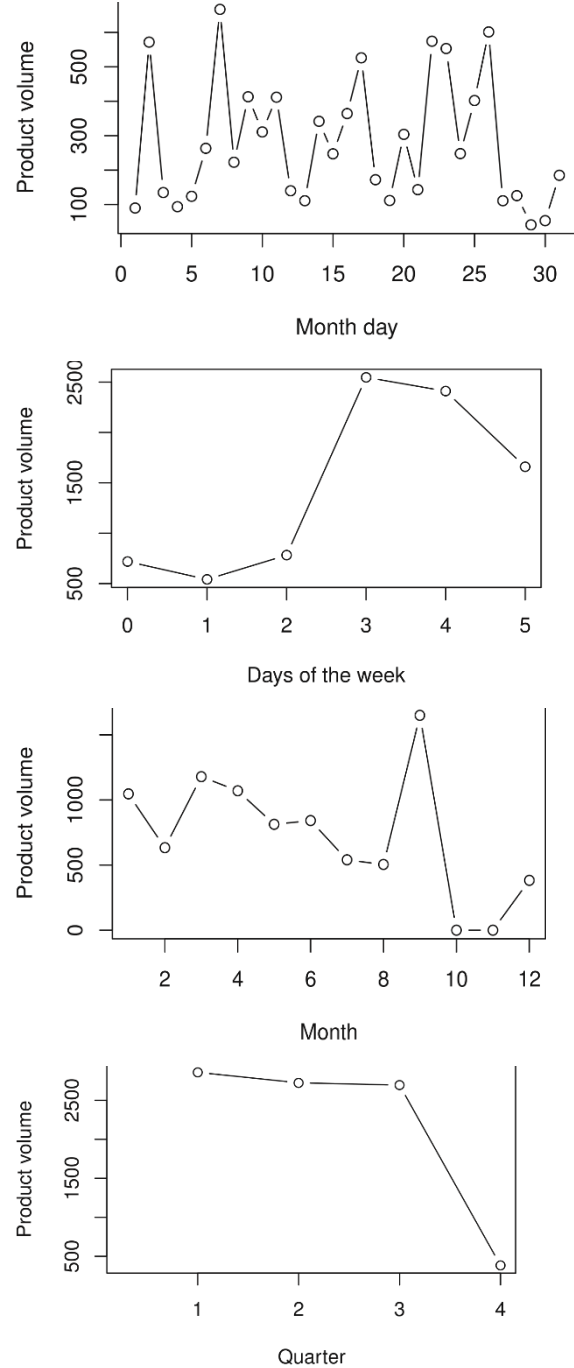


Fig. 1. Instances of extracted periodic components using Dolly Girl Lunch Box sales volume as an example from a 20XX year Data-Mining-Cup task, a) depending on the number of days, b) depending on a weekday, c) depending on a month, d) depending on a quarter

In order to identify these cycles, we will summarize all the values corresponding to the numbers of days (in a month), months, weeks, quarters, etc. and divide the number of such values. As a result, we will determine the value of

the component $g_{periodic}(t)$ which afterwards will be subtracted from the values of the available statistics and move on to the search of the remaining components.

The parameters that describe production systems and their economic parameters contain the trend components ($g_{trend}(t)$) and can be described using innovation [1] and S curves. These curves are usually used to describe the parameters in which trends (days, months, quarters) cannot be distinguished (but we can observe them on a big-time interval). It should be noted that these intervals can vary depending on the parameter considered. Innovation curves describe economic parameters (revenue, sales volume, the market share held by the company or the product), number of competing firms or products, number of people involved in the project, number of products and so on). S-curves describe technical and technological parameters (development or integration of new technology costs, performance index, phases of a technology or a process and so on), which shows the stage of the technology development and prospects for its modification (sometimes innovation and S-curves cannot be distinguished. In this case the latter is considered to be a special case of the former).

Registration and identification of these trends imply consideration of statistical data for a vast period of time (such data may not be), the use of only one state of the innovation curve development can be used for a group of products, information about similar products is not taken into account if we do not have enough statistical data.

As a result of these operations, we obtain the residual component of the time series ($g_{res}(t, t_G)$). To predict its values, we will use one of the known methods for time series (see fig.2 and fig.3). At the same time, to impart an adaptability property, we will use only a limited number of previous values/time frame that will shift as new values arrive. Forecasting in this way predicts one new time step value. This method raises the problem of choosing the value of t_G that leads to the best result:

$$\arg \min_{1 \leq t_G \leq t_N} \int_t^{t+t_F} e(t) dt \rightarrow \min,$$

Where t_F — planning horizon, t_N — amount of retrospective data in the training set.

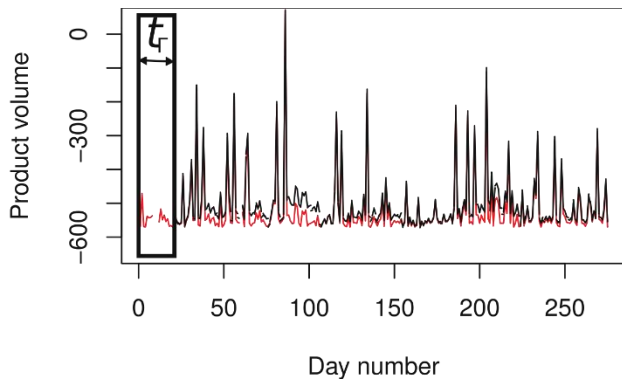


Fig. 2. Prediction of the residual component

Considering this problem, the dependence on the method for which the value of t_G and the value of the planning horizon t_G will be selected is obvious.

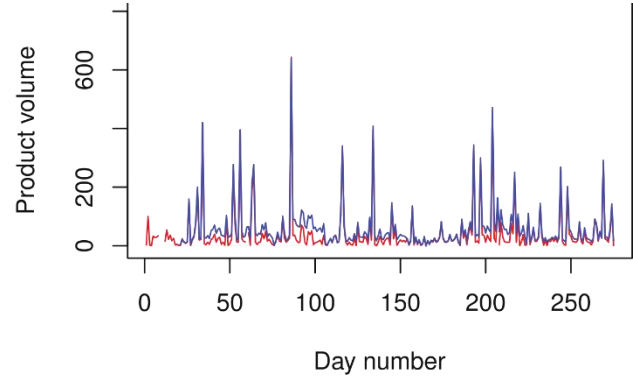


Fig. 3. Recovery of the time series values

The value of the planning horizon can be both set as the desired characteristic and determined by checking the prediction accuracy on the test set of retrospective data as described in [12].

Consideration of interactions ($g_{interference}(t)$) is possible if we consider a set of parameters the values of which must be predicted jointly. In the paper [13] it is proposed to use cognitive maps with variable values of coupling coefficients to take into account these components.

Based on the results of the forecast, we recover our data. Then, on the test sample, we determine the type of distribution to which ϱ belongs. Next, we add this data to the training sample and build a forecast. In general, the forecast algorithm will look as show (see fig.4).

- Step 1.** Split the set of retrospective data into training and test samples (usually the ratio is 2/3 and 1/3 respectively);
- Step 2.** For all the retrospective values determine the corresponding day of the month, day of the week, month number, quarter number;
- Step 3.** Calculate the periodic components for the month, day of the week, month and quarter;
- Step 4.** Subtract the periodic components from the test values and analyze the data obtained for trends (based on the match of an analyzed parameter to the group of trends and the sufficiency of data for observing trends based on visual analysis). If there are no trends that can be identified in this way, move on to step 6;
- Step 5.** Using the least-squares method, determine the coefficients for the function describing the trend and subtract this component from the training sample;
- Step 6.** Choose a prediction method, time frame size and planning horizon. Using the selected method, we describe the residual component of the predictive function;
- Step 7.** The same approach for the month-day transformation (a vector of days for every month will be created at this step);
- Step 8.** Considering the parameter which is included in the group of the parameters characterizing the subject area or system, we use the method described in [12] for the component that takes into account their mutual influence;
- Step 9.** Restore the forecast values based on the components and determine the distribution which corresponds to the nature of the error. Use the membership function to describe the error based on the distribution;
- Step 10.** Add the values from the test sample to the training sample and use the resulting data to build the final forecast;

Fig. 4. Algorithm for obtaining the values of the parameter specified by the time series

III. RESULTS

Using the described approach, we can increase forecasting accuracy. For testing purposes, we will compare the prediction results $g(t)$ with retrospective data of the test sample $g^*(t)$. In this case, the use of acceptance criteria will not be so efficient, as they give negative results because of the large amount of data. Therefore, it is necessary to use parametric criteria that give a numerical assessment of the hypothesis's significance. Let's have a look at hypothesis comparison based on the Fisher criterion. Taking into

account the value and number of degrees of freedom, we can determine the value of the hypothesis probability (p) and conclude the significance of the hypothesis. To do this we need that $p > 0.5$. Comparing the results received for different planning horizons (t_G) and using multiple prediction methods for the entire time series and component $g_{rest}(t, t_G)$ of our method, we can observe the growth of the hypothesis significance see table 1. Despite the increase of importance, there are errors in the forecasts which we will check for affiliation to one or another distribution (see table 2).

TABLE I. TEST OF HYPOTHESIS SIGNIFICANCE BASED ON FISHER CRITERION

Prediction method	Sales volume prediction			
	Product 1	Product 2	Product 1	Product 2
	$t_G = 15$		$t_G = 90$	
kNN	4.212e-09	0.0001319	1.151e-09	2.2e-16
RVM	0.8415	1.428e-10	0.09669	2.2e-16
kNN for g_{rest}	6.259e-08	0.03222	0.5648	0.003682
RVM for g_{rest}	0.9184	0.9563	0.8564	0.8488

TABLE II. THE VALUE OF FAILURE BY TESTING FORECAST MEMBERSHIP DEVIATION FROM DISTRIBUTIONS BY USING SIGNIFICANT HYPOTHESIS (THE FORECASTING OF THE COMPONENT g_{rest} WITH HELP OF RVM)

Disctribution	Sales volume prediction			
	Product 1	Product 2	Product 1	Product 2
	$t_G = 15$		$t_G = 90$	
Norm	0.05328509	0.009346162	0.012756422	0.008422482
Cauchy	0.01278417	0.010889446	0.009643750	0.008389444
Weibull	0.28947612	0.900763835	0.13274692	0.222961854
Gamma	0.9816906	2.615226	0.609907	0.8248954
Exp	0.8605149	2.062905	0.4822252	0.5070099
Beta	0.1916015	2.413735	0.1611847	0.6752295

As we can see from the table 2 the smallest error corresponds to the normal and Cauchy distribution. For these distributions, their functional description of probability density is known, which will be used to get the membership function. As a result, we obtain a fuzzy forecast function.

IV. DISCUSSION

The improvement of the prediction accuracy gives us opportunities for improving the results of planning and predictive management problems. Additionally, forecasts conversions into their distinct forms extend opportunities of results analysis.

We can consider problems of decision supports systems [14] and expert systems where we can use the data in its fuzzy form. In these tasks the translations allow us to improve the obtained estimates objectivity by eliminating the use of expert assessments or their representation in the formal form.

As an example, let us consider a dynamic volume-scheduling problem of finding the optimal output volumes for a given time ($t+1$) based on the product portfolio for time t .

$$\max_{k=1, N} \left(\sum_{i=1}^K ((C_i - X_i^k(t+1) - Y_i - X_i^k(t+1)) \rightarrow \max, \right.$$

$$X_i^k(t+1) \in U_i^k(t+1), " i, k,$$

$$W_i^k \in W^k(t), " i, k,$$

$$X_i^k(t+1) \in U_i^k(t+1), " i, k,$$

$$\bigwedge_{i=1}^N X_i^k(t) \in S_i(t+1), " i,$$

$$X_i^k(t+1) \geq 0, " i, k,$$

Where i — product number, k — production line number, C_i — the product price of i type, X — product quantity, U^k — available production capacity of k line, S — product demand limits i (based on the forecast).

As a result of the transition to the fuzzy form, the solution of the problem can be carried out in its fuzzy form [15] and then the solutions of the problem can be illustrated as shown (see fig.5 and fig.6).

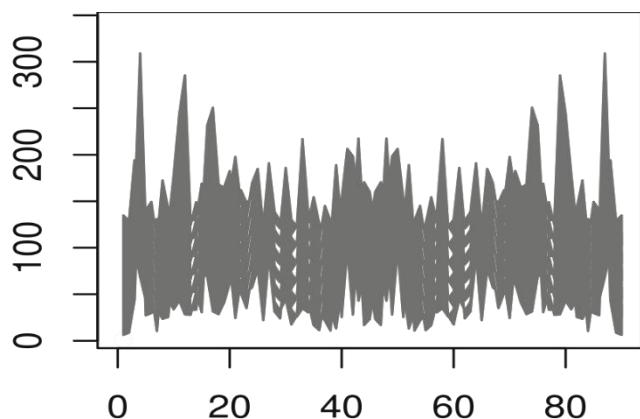


Fig. 5. The changes of the criterial function values in the fuzzy form, depending on the degree of expert confidence (instant values)

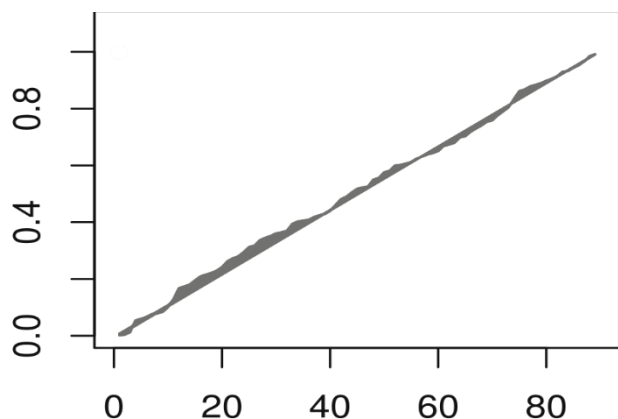


Fig. 6. The changes of the criterial function values in the fuzzy form, depending on the degree of expert confidence (cumulative values)

The presented result allows us to draw conclusions about the contribution of each product to the profit of the production system, analyze the level of its diversification and trends to the unpredictability of market behavior in the future. At the same time, we can observe separations of increased probability areas number of which exceeds product planned quantity. That allows us to talk about the possible development trajectories for the considered production system with the analyzed product portfolio. Moreover, we can talk about the choice of any behavior strategy (expert's degree of confidence --- for example, there are well-known approaches as the principle of guaranteed result (minimax or wald criteria).

V. CONCLUSION

The paper discusses a method of increasing the accuracy of predictions and the translation of their values in a fuzzy form. Despite the possibility of improving the accuracy, the approach does not answer the questions related to the automatic choice of the prediction method for the random component (g_{rest}) and the creation of the membership

function. Taking into account that the nature of the data describing the parameters, which characterize the external environment, may change, it may be necessary to change the value of the time frame (if the mobility of the external environment changes), membership function and forecasting method for g_{rest} , which can be the subject of further research.

REFERENCES

- [1] Amberg M., Mylnikov L. "Innovation project lifecycle prolongation method: Innovation and Knowledge Management". In: Twin Track Economies: Challenges & Solutions, (IBIMA'2009), Cairo, Egypt, 2009, pp.491-495.
- [2] Mandelbrot B., Hudson R. "The Misbehavior of Markets: A Fractal View of Financial Turbulence". Basic Books, Cambridge, 2006.
- [3] Fernández-Delgado, M., Sirsat, M. S., Cernadas, E. "An extensive experimental survey of regression methods". In: Neural networks: the official journal of the International Neural Network Society, 2019, pp.11-34.
- [4] Daniel T., Lee A. "Wavelet Analysis". In: Theory and Applications, Hewlett-Packard Journal, 1994.
- [5] Cleveland R., Cleveland W., McRae J., Terpenning I. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess". In: Journal of official statistics, 1990.
- [6] Fenghua W., Jihong X., Zhifang H. "Stock Price Prediction based on SSA and SVM". In: Procedia Computer Science, vol. 31, 2014, pp.625-631.
- [7] Luzianin I., Krause B. "Similarity Measurement of Biological Signals Using Dynamic Time Warping Algorithm". In: Proceedings of International Conference on Applied Innovation in IT, (ICAII'2016), Koethen, Germany, 2016.
- [8] Lin J., Keogh E., Lonardi S. "A symbolic representation of time series, with implications for streaming algorithms". In: Proceedings of the 8th ACM SIGMOD workshop on mining and knowledge discovery. ACM Press, New York, 2003, pp.1-10.
- [9] Wold S., Sjöström M., Eriksson, L. "PLS-regression: a basic tool of chemometrics". In: Chemometrics and Intelligent Laboratory Systems, vol.58, 2001, pp.109-130.
- [10] Santosa F., Symes W. "Linear inversion of band-limited reflection seismograms". In: SIAM Journal on Scientific and Statistical Computing, vol.7, 1986, pp.1307-1330.
- [11] Oseledets I.V. "Tensor-train decomposition". In: SIAM Journal on Scientific Computing, vol. 33, 2011 pp.2295-2317.
- [12] Mylnikov L., Kuetz M. "The risk assessment method in prognostic models of production systems management with account of the time factor". In: European Research Studies Journal, 2017, pp.291-310.
- [13] Suslova A., Mylnikov L., Krause B. "Prognostic model for estimation of innovative activity factors of regions by example of the patenting data based on cognitive map modeling". In: European Researcher, vol.61, 2013, pp.2508--2517.
- [14] Mardani A., Jusoh A., Zavadskas E. "Fuzzy multiple criteria decision-making techniques and applications - Two decades review from 1994 to 2014". In: Expert Systems with Applications, vol. 42, 2015, pp.4126-4148.
- [15] Villacorta P., Rabelo C., Pelta D. "FuzzyLP: An R Package for Solving Fuzzy Linear Programming Problems". In: Granular, Soft and Fuzzy Approaches for Intelligent Systems, vol. 344, Springer International Publishing, 2017, pp.209-230.