Extraction and Forecasting Time Series Of Production Processes

Anton Romanov, Evgeny Egov, Irina Moshkina, Dyakov Ivan
Ulyanovsk State Technical University, Ulyanovsk, Russia

Abstract. In this paper, the basic principles of building information support for an automated system for balancing the production capacities of large industrial enterprises are formulated. In addition, the model of forecasting time series in the task of power balancing is given.

1. Introduction

The technological preparation of complex production at large enterprise requires the analysis of production capacities. The aim is to increase the efficiency of the use of material, technical and human resources [1]. Achieving the goal requires several tasks: input data definition, the creation of models reflecting the state of production processes, development of balancing algorithms.

The solution of the set tasks implies the possibility of forming a unified information environment for technological support of production.

2. Time series of enterprise

We analyse production processes of aviation factory. The task is to balance the production capacity. The current approach of management is based on using common methodology. Methodology contains algorithms and coefficients, accumulated from statistic of production. The main disadvantage of this approach is a strong discrepancy between the real production indicators and the collected statistical data [2]. Denote the limitations of applying methodology:

- long extraction time of statistical coefficients from production indicators;
- impossibility of dynamic adaptation of calculations into separate periods shorter than the forecast horizon;
- the methodology does not provide for adaptation to a specific production.

Summarizing, we can note a significant averaging in the calculations, which reduces the accuracy.

By analyzing these techniques it is easy to see, that the coefficients are aggregated and averaged information from the indicators of production processes. Such processes are easily represented by discrete time series. When analyzing production processes, it was found that this discrete interval is the month - the minimum forecast horizon, and the time point at which the indicators are unchanged.

We extract the following types of time series: employee work time fund, tool work time fund, performance ratio, area usage, tool wear.

For all types of process we identify monthly indicator values. Now we can identify models of processes by using time series. Very important to find the following characteristics of time series: seasonality, local and global tendencies.

3. F-transform

We use F-transform to smooth production time series. Generally, the F-transform of a function $f : P \longrightarrow \mathbb{R}$ is a vector whose components can be considered as weighted local mean values of f. Throughout this paper we will assume that $\mathbb{R}$ is the set of real numbers, $[a, b] \subseteq \mathbb{R}$, and $P = \{p_1, \dots, p_l\}$, $n < l$, is a finite set of points such that $P \subseteq [a, b]$. Function $f : P \longrightarrow \mathbb{R}$ defined on the set P is called *discrete*.

Below, we will remind basic facts about the F-transform as they were presented in [3].

The first step in the definition of the F-transform of f is a selection of a *fuzzy partition* of the interval $[a, b]$ by a finite number $n \geq 3$ of fuzzy sets $A_1, \dots, A_n$. According to the original definition, there are five axioms which characterize a fuzzy partition: *normality, locality, continuity, unimodality and orthogonality* (the Ruspini condition) [3].

A fuzzy partition is called *uniform* if the fuzzy sets $A_2, \dots, A_{n-1}$ are shifted copies of the symmetrized A_1. The membership functions $A_1, \dots, A_n$ in the fuzzy partition are called *basic functions*. We say that the basic function A_k *covers* a point p_j if $A_k(p_j) > 0$.

Figure 1 shows a uniform fuzzy partition of an interval [a, b] by fuzzy sets $A_1, \dots, A_n$, $n \geq 3$, with triangular membership functions. The formal expressions for an uniform fuzzy partition of an interval $[a, b]$ by fuzzy sets $A_1, \dots, A_n$, $n \geq 3$, with triangular membership functions are given below where $h = \frac{b-a}{n-1}$.

$$A_1(x) = \begin{cases} 1 - \frac{(x-a)}{h}, & x \in [a, x_2], \\ 0, & \text{otherwise}, \end{cases}$$

$$A_k(x) = \begin{cases} \frac{|x-x_k|}{h}, & x \in [x_{k-1}, x_{k+1}], \\ 0, & \text{otherwise}, \end{cases}$$

$$A_n(x) = \begin{cases} \frac{(x-x_{n-1})}{h}, & x \in [x_{n-1}, b], \\ 0, & \text{otherwise}. \end{cases}$$

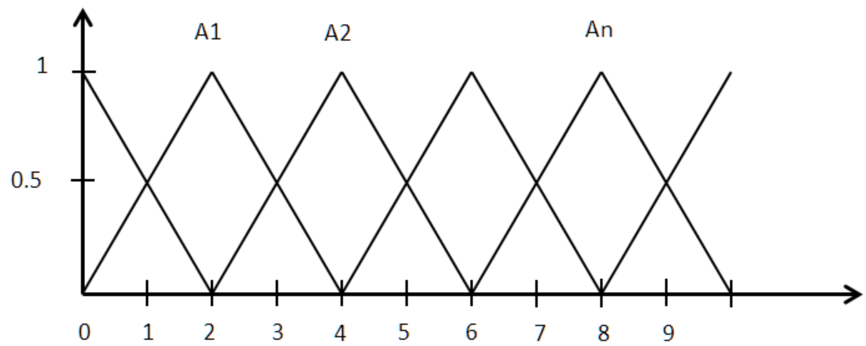

Figure 1. An example of a uniform fuzzy partition by triangular membership functions

In the subsequent text we will fix the interval $[a, b]$, a finite set of points $P \subseteq [a, b]$ and relaxed fuzzy partition $A_1, \dots, A_n$ of $[a, b]$. Denote $a_{kj} = A_k(p_j)$ and consider $n \times l$ matrix A with elements a_{kj}. We will say that A is a *partition matrix* of P. Below, a matrix of a special uniform partition is presented.

Assume that the points $p_1, \dots, p_l \in [a, b]$ are equidistant so that $a = p_1$, $b = p_l$, $p_{i+1} = p_i + h$, $i = 1, \dots, l-1$, and $h > 0$ is a real number. Let $A_1, \dots, A_n$ be a uniform partition $[a, b]$ such that each basic function A_k has a triangular shape and covers fixed number of points, say N. Moreover, let nodes $x_0, x_1, \dots, x_n, x_{n+1}$ be among the points $p_1, \dots, p_l$ so that $x_0 = p_1$, $x_{n+1} = p_l$. If N is an odd number, say $N = 2r-1$, then $l = (n+1)r - 1$. In this particular case, the basic function A_k covers the points $p_{(k-1)r+1}, \ldots, p_{(k+1)r-1}$, so that

$$A_k(p_{(k-1)r+1}) = \frac{1}{r}, \ldots, A_k(p_{kr-1}) = \frac{r-1}{r}, A_k(p_{kr}) = 1,$$
$$A_k(p_{kr+1}) = \frac{r-1}{r}, \ldots, A_k(p_{(k+1)r-1}) = \frac{1}{r}.$$

3.1. Discrete F-transform

Once the basic functions $A_1, \ldots, A_n$ are selected, we define (see [4]) the (direct) *F-transform* of a discrete function $f : P \longrightarrow \mathbb{R}$ as a vector $(F_1, \ldots, F_n)$ where the k-th *component* F_k is equal to

$$F_k = \frac{\sum_{j=1}^{l} f(p_j) \cdot A_k(p_j)}{\sum_{j=1}^{l} A_k(p_j)}, \quad k = 1, \ldots, n. \tag{1}$$

In order to stress that the *F-transform components* $F_1, \ldots, F_n$ depend on $A_1, \ldots, A_n$ we say that the F-transform is taken with respect to $A_1, \ldots, A_n$.

4. Forecasting TS based on fuzzy trends

The fuzzy elementary trend modeling method [7], [8] is used to predict numerical values and fuzzy trends in the state of an organization project in a given product. The forecast uses hypothesis testing:

(i) *Hypothesis 1.* The hypothesis of conservation of trend. The Forecast is constructed on base the previous period. The formula for the predicted value

$$\tau_{t+1} = \tau_t + \tau_p,$$

where τ_{t+1} – forecast for the next period of time; τ_t – the real value at time t; τ_p – the value of the trend over the previous period of time.

(ii) *Hypothesis 2.* The hypothesis of stability of the trend. The moving average is used to predict

$$\tau_{t+1} = \tau_t + G\tau_p,$$

where $G\tau_p$ – the importance of a dominant fuzzy trend. Consider the trend of the previous selected period. We select the predominant cluster of trends. The forecast for the above formula is calculated. The trend is built. Optimistic forecast for the some number of occurrences of trends used. The highest average trend is selected.

(iii) *Hypothesis 3.* Forecasting for a given period on the basis of fuzzy elementary trends. Stages of the prediction algorithm for the period based on trends:

The expert sets the number of considered trends for the previous period. For example, for half a year - a set of trends **A**. Either he sets the pattern set of trends. The presumed trend following this set is known.

$$\tau_{t_{n-m}}, \ldots, \tau_{t_{n-1}}, \tau_{t_n}$$

Search for a set of trends **A** in all other previous periods.

$$\{\tau'_{t_{n-l-k}}, \ldots, \tau'_{t_{n-l-k-1}}, \tau'_{t_{n-l}}\}$$

If such a set of **B** is found in which the **C** trend is located after this found set **B** then trend **C** is considered into account. The forecast equal to the trend **C** is constructed.

$$\tau_{t+1} = \tau_t + \tau'_{t_{n-l+1}}$$

If the set **B**, which would coincide with the set **A**, was not found then the search for the set is repeated, but it is already not looking for its complete coincidence. We select new pattern **A** is shorter into one trend. This is repeated until a suitable set of trends B [9]. To select the best hypothesis, an entropy time series is additionally introduced [10].

5. Forecasting TS using the adaptation algorithm

To solve the prediction problem for the time series $Y = \{t_i, x_i\}, (i = 1, 2, ..., n)$ with the help of fuzzy similarity, provided that the hypothesis of the expert hypothesis exists that the fuzzy tendency of the time series $Z = \{t_i, z_i\}, (i = 1, 2, ..., k)$ is a predictor of the time series Y, the forecast hypothesis correction algorithm [9] is used. Algorithm 1 includes three phases. In the first phase, fuzzy elementary trends in the time series Y are predicted:

$$\tau_{t+1}^{Y} = f(\tau_t^{Y}),$$

here τ_{t+1}^{Y} is the predictive fuzzy elementary trend of the time series Y;

τ_{t+1}^{Y} – the current fuzzy elementary trend of the time series Y;

F – dependence in the fuzzy elementary trends of the time series Y.

In the second phase, the forecast fuzzy elementary trend of the time series Y is corrected taking into account the components of the main trends of the time series $G\tau_Y$ and the time series of the predictor $G\tau_Z$, respectively:

$$\hat{\tau}_{t+1}^{Y} = r(\tau_{t+1}^{Y}, G\tau_Y, G\tau_Z),$$

where τ_{t+1}^{Y} – predictive fuzzy elementary trend of time series Y;

$\hat{\tau}_{t+1}^{Y}$ – predictive fuzzy elementary trend of the time series Y after the adjustment;

$G\tau_Y$ – basic fuzzy trend of time series Y;

$G\tau_Z$ – the main fuzzy trend of time series Z;

r – correction rules.

The third phase is used to obtain the estimate of the predicted value of the numerical time series Y. On this basis, the following algorithm for predicting short-term fuzzy trends in time series is proposed.

Algorithm 2. Step 1. Conversion of the numerical TS $Y = \{t_i, x_i\}, (i = 1, 2, ..., n)$ to fuzzy TS $\tilde{Y} = \tilde{x}_t, \tilde{x} \in \tilde{X}, t = 1, 2, ..., n$:

$$\tilde{x_t} = Fuzzy(x_i), x_i \in X, \tilde{x}_t \in \tilde{X}$$

In this case, the intervals at which fuzzy sets are defined, their form and name are specified by the user based on the features of the domain.

Step 2. Conversion of fuzzy TS $\tilde{Y} = \tilde{x}_t, \tilde{x} \in \tilde{X}, t = 1, 2, ..., n$ into an indistinct time series of fuzzy elementary trends:

$$\tau_t^{Y} = \langle \tilde{v_t}, \tilde{a}_t, \mu_t \rangle,$$

$$\tilde{v_t} = TTend(\tilde{x}_t, \tilde{x_{t+1}}),$$

$$\tilde{a_t} = RTend(\tilde{x}_t, \tilde{x_{t+1}}),$$

$$\mu_t = min(\mu(\tilde{x_t}), \mu(\tilde{x}_{t+1})).$$

We first define the set of types of FT types $\tilde{V} = \{Fall, Growth, Stability\}$, and a set of names of intensities of FTs $\tilde{A} = \{Strong, Medium, Weak\}$.

Step 3. Construction of a model for changing the components of fuzzy elementary trends of time series Y and their prediction for one period:

$$\tilde{v}_{t+1} = \tilde{v}_t \times \tilde{v}_{t-1} \times ... \times \tilde{v}_{t-p} \circ R_{\tilde{v}}(t, t-p),$$

$$\tilde{a}_{t+1} = \tilde{a}_t \times \tilde{a}_{t-1} \times ... \times \tilde{a}_{t-p} \circ R_{\tilde{a}}(t, t-q).$$

Step 4. Forecasting the numerical time series Y with the preliminary defuzzification the component of the fuzzy trend $\tau_{t+1}^{Y} = \langle \tilde{v}_{t+1}, \tilde{a}_{t+1}, \mu_{t+1} \rangle$:

$$x_{t+1} = x_t + v_{t+1} \cdot a_{t+1}.$$

Step 5. Application of the algorithm for identifying the main trend (see phase 3. Algorithm 1) for the time series Y and calculating its components $G\tau_Y = \langle \tilde{v}^Y_{G\tau}, \tilde{a}^Y_{G\tau}, \mu^Y_{G\tau} \rangle$. Defuzzification of the components of the main fuzzy trend of the time series Y.

Step 6. Application of the algorithm for identifying the main trend (see phase 3. Algorithm 1) for the time series Z and calculating its components $G\tau_Z = \langle \tilde{v}^Z_{G\tau}, \tilde{a}^Z_{G\tau}, \mu^Z_{G\tau} \rangle$. Defuzzification of the components of the main fuzzy trend of the time series Z.

Step 7. Correction of the predictive fuzzy elementary trend of the time series $Y\hat{\tau}^Y_{t+1} = r(\tau^Y_{t+1}, G\tau_Y, G\tau_Z)$:

$$\hat{\tau}^Y_{t+1} = v_{t+1} \cdot a_{t+1} + v^Y_{G\tau} \cdot a^Y_{G\tau} + v^Z_{G\tau} \cdot a^Z_{G\tau}.$$

Step 8. Calculation of the corrected forecast value of the numerical time series Y for one period $x'_{t+1} = x_t + \hat{\tau}^Y_{t+1}$.

6. Forecasting time series using entropy

To predict new values, the phase plane method of the fuzzy trend is used. The method uses a change in the fuzzy trend in the phase plane. For this, each fuzzy trend is given a weight. To construct the phase plane by a fuzzy trend, each trend is given a value, weight. The rate of increment of the trend is defined as the difference between the weight of the trend in the previous point and the weight of the trend at the current point.

Table 1. Scheme for coding points on the phase plane

Code	Weight trends	Trend dynamics	Point in the plane	Interpretation
0	= 0	= 0	origin	A point at the origin of coordinates means the system is in a stable state
1	> 0	> 0	first quarter	In the system there is growth or stabilization after the fall
2	< 0	> 0	second quarter	A smaller drop after a big fall
3	< 0	< 0	third quarter	In the system there is a fall or stabilization after growth
4	> 0	< 0	fourth quarter	Smaller growth after large growth
5	= 0	> 0	of axes X is 0	Stabilization after falling (appears only after 3 states)
6	= 0	< 0	of axes X aligns to 0	Stabilization after growth (appears only after 1 state)
7	> 0	= 0	of axis Y from 0	Growth after growth
8	< 0	= 0	of axes Y lower than 0	Fall after fall

It is necessary to unify the information on the relationship between the weight of the trend and its speed. The encoding of possible locations of points on the coordinate plane is introduced.

The next step is to get the predicted value. Determination of the correspondence between the displacement of points on the phase plane and trends. The trend values for each code are stored. The obtained rules shown in table 2.

Table 2. Extracted rules

Stability	**Growth**	**A fall**
0 or 5 or 6 $\rightarrow$ 0	0 or 2 or 3 or 4 or 5 or 6 or 8 $\rightarrow$ 4	0 or 1 or 2 or 3 or 4 or 5 or 6 or 7$\rightarrow$ 2
1 or 4 or 7 $\rightarrow$ 5	1 or 4 or 7 $\rightarrow$ 7	2 or 8 $\rightarrow$ 8
2 or 3 or 8 $\rightarrow$ 6	4 $\rightarrow$ 1	2 $\rightarrow$ 3

From these rules, a number of conclusions are drawn:

(i) If the point of the phase plane for the predicted point lies in the center of coordinates (code 0) or on the X axis (codes 5 and 6), the value of the fuzzy trend will correspond to the Stability.

(ii) If the point is in the first or fourth quarter, or lie on the Y axis above 0, then the value of the fuzzy trend will match Growth. The intensity can be taken into account. In this case, the Growth trend is broken down into 3 growth trends, taking into account the intensity: Weak, Medium or Strong.

(iii) If the point is in the second or third quarters, or on the Y axis below 0, then the value of the fuzzy trend will match the Fall. The intensity can be taken into account. In this case, the tendency of the Fall is divided into 3 tendencies of the Fall, taking into account the intensity: Weak, Medium or Strong.

When analyzing the time series, the schemes of the displacement of a point on the phase plane are determined. The frequency of the transition of a point from one area (labeled code) to another is determined. On the basis of frequency, a frequently encountered displacement pattern is determined. In this scheme, the code of the starting point is equal to the location code of the point on the phase plane for the last known point of time series. By the code value for the end point, the scheme determines possible trends (there may be more than one). The predicted values are then calculated on the basis of these trends. The drawbacks of the method include the acquisition of several equally probable values from which you want to choose one.

7. Balance system

We have developed information system, that implements next functions:

- performs calculation of production capacities;
- reveals a deficit and forms recommendations for balancing capacities by determining the possibility of redistribution of the volumes of the same type of work;
- identifies the need to enter additional production areas and equipment;
- identifies the need for recruitment and redeployment of staff.

The basic input data is the production program. It sets the list of products and the scope of work for their creation, distributed by period. Based on current indicators of production processes, their dynamics, an enterprise can redistribute the amount of work between time periods.

As we above define, are three types of resources exist: human, material and production area. For balancing, the following steps are required:

(i) Identify the units for which we are balancing.

(ii) For each unit, we calculate the current capacity for each of the three types of resources.

(iii) For each unit, we define free capacity for each of the three types of resources.

So, next steps depend on resource type. For human resources, we need to set the following possibilities for balancing: transfer between units and hiring new workers. Limiting factors in this are the skills of specific employees in the translation and the delayed start of the work of the employee in hiring.

We need to append balancing algorithm by next steps:

- If there are free human resources and a transfer of workers is possible, then we fulfill it.
- Otherwise, we hire workers.

This steps show the priority, which on enterprise is defined.

Material resources, such as equipment and machines, are difficult to transfer between departments. Therefore, when balancing, the possibilities for redistribution of planned work are shown. If there are no available resources, then the only option is to purchase new equipment.

One more situation possible with the production areas. Their redistribution is also impossible. Enterprise can start using new production areas. Note, that the use of production areas directly depends on the equipment occupying these areas. Therefore, in the planning phase, it is also possible to select the optimal accommodation options.

Current implementation of the information system is based on average values of indicators throughout the year. We propose to analyze the time series of indicators at more frequent intervals. To do this, an important role will be determined by the accumulated information in the enterprise information systems.

8. Conclusion

The analysis of existing algorithms, data and information systems has shown a strong accumulation of errors in calculations. It was shown the great impact of operational monitoring of indicators.

These principles allow improving the quality of technological preparation of complex industries [5] [6]. Proposed methods of prediction of time series are improve the quality of management decisions.

9. References

[1] Yarushkina N, Negoda V, Egorov Yu, Moshkin V, Shishkin V, Romanov A and Egov E 2017 Modeling the process of technological preparation of production based on ontological engineering *Automation of management processes* **4** pp 94-100

[2] Yarushkina N, Afanasyeva T, Negoda V, Samokhvalov M, Viceroy A, Guskov G and Romanov A 2017 Integration of design diagrams and ontologies in the objective of the balancing of the capacity of the aviation-building enterprise *Automation of management processes* **4** pp 85-93

[3] Perfilieva I, Yarushkina N, Afanasieva T and Romanov A 2013 Time series analysis using soft computing methods *International Journal of General Systems* **42:6** pp 687-705

[4] Perfilieva I 2006 Fuzzy transforms: Theory and applications *Fuzzy Sets and Systems* **157** pp 993-1023.

[5] Yarushkina N 2009 Modern intellectual analysis of fuzzy time series *Integrated models and soft computations in artificial intelligence: proceedings of fifth scientific-practical conference* Vol. 1 pp 19–29

[6] Yarushkina N 2004 Fuzzy neural networks with genetic adjustment *Scientific session of MEPhI-2004 "Neuroinformatics-2004". VI-th All-Russian Scientific and Technical Conference. Lectures on neuroinformatics. Part 1. By materials School-seminar "Modern Problems of Neuroinformatics"* p 173

[7] Herbst G and Bocklish S 2009 Online Recognition of fuzzy time series patterns *International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy*

[8] Kacprzyk J and Wilbik J 2009 Using Fuzzy Linguistic summaries for the comparison of time series *International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic*

[9] Yarushkina N and Timina I 2016 Automated design based on the model for predicting fuzzy trends in metrics *Fifteenth National Conference on Artificial Intelligence with International Participation (CAI 2016)* pp 156-163

[10] Yarushkina N, Timina I and Egov E 2017 Measure of entropy in software quality assessment *VII Russian scientific-practical conference "Fuzzy systems, soft computing and intelligent technology"* pp 201-208

Acknowledgements

The authors acknowledge that the work was supported by the framework of the state task of the Ministry of Education and Science of the Russian Federation No. 2.1182.2017 / 4.6 "Development of methods and means for automating the production and technological preparation of aggregate-assembly aircraft production in the conditions of a multi-product production program", RFFI-16-47-732070. The reported study was funded by RFBR and the government of Ulyanovsk region according to the research project № 18-47-730022 and № 18-47-732016