Generation of a rational training sample when predicting power consumption for train traction

V.E. Osipova¹, D.A. Yakovlev¹

¹ Trans-Baikal Institute of Railway Transport (Irkutsk State University of Railway Transport), 11 Magistralnaya Str., Chita, 672040, Russia

E-mail: Valeri3307@mail.ru

Abstract. The paper considers peculiarities of forecasting in the field of railroad facilities based on approximation of time series and neural networks. The objects of traction power supply of the Transbaikal Railway are considered as the object of the study. Forecasting on the basis of approximation for consumers with maximum power consumption increases the accuracy of forecasting the electric energy consumption by 4...7%. The application of neural networks in forecasting of power consumption allows reducing the value of error to 2%, thus leading to significant reduction of costs for fuel and energy balance of a structural division or an enterprise as a whole.

1. Introduction
At present, the existing system of JSC Russian Railways is increasingly aimed at the formation of a unified information space for freight transport and logistics in order to ensure sustainable competitiveness of Russian Railways by increasing the attractiveness of transport and logistics services provided to clients through digital technologies [1, 2].

Digitalization of railways is primarily aimed at reducing the impact of the human factor on the condition of railway facilities and is quite attractive by the speed of its services provided to the consumer.

One of the first stages of digitalization is the digitalization of economy, namely the application of digital intelligent technologies in management in order to cut expenses, increase security and reduce costs.

One of the most effective ways to reduce costs in the railway sector is proper formation of the fuel and energy balance of a structural unit, which mainly depends on the accuracy of load forecasting.

Load forecasting is an important component of the energy management system within the power system. Accurate load forecasting helps power supply enterprises to make decisions on mandatory installation, reduces power margin during rotation, and correctly designs the maintenance plan of a device [3].

2. Analysis of power consumption data for hauling operations
Forecasting in the field of railway facilities represents a multilevel of the structure and uncertainty of factors describing the system status. Let us consider the peculiarities of power forecasting using the example of objects of the traction power supply of the Transbaikal Railway [4].

Figure 1 shows the flow chart of fuel and energy balance formation of the Transbaikal Railway.
According to the analysis of statistical data, the average annual productivity of the Transbaikal Railway makes 1641118.1 thousand t/km, annual linear range – 14187344 km, rail traffic – 1019.30 million t/km, and energy consumption – 62195720 kWh. Figure 2 shows the dynamics of power consumption of the Transbaikal Railway section.

The analysis of statistical data on power consumption dynamics allows building the interdependence between the main parameters involved in the formation of fuel and energy balance and determining the main factors affecting the consumption of fuel and energy resources of a structural division of the Transbaikal Railway.

All factors affecting power consumption on train traction can be described by the following function

\[ W_i(t) = S \cup T, \]

where \( W_i(t) \) – vector function characterizing a set of influencing factors;
S – vector characterizing the value of freight transport;
T – vector characterizing the state of the traction power supply system of the section.

In turn, factors affecting power consumption on train traction can be divided into constant and time-variable factors. Constant factors can be described by the following function

\[ T_{c}(t) = [TT, L(t), F(t)] \]  

where \( T_{c}(t) \) – vector function characterizing the set of influencing factors that do not change over time.

The factors that vary over time can be described by the following function

\[ S_{s}(t) = [m(t), t(t), Q(t), L(t)] \]  

where \( S_{s}(t) \) – vector function characterizing the set of influencing time-variable factors.

If \( T \to \min \) and \( S \to \min \), then \( W(t) \to \min \)

\[ W_j(t) = \begin{cases} L_{j}(t) \in \left[ L_{r_{\min}} \to L_{r_{\max}}, m(t) \in \left[ m_{\max} \to m_{\min}, Q(t) \in \left[ Q_{\min} \to Q_{\max} \right] \right) \end{cases} \]

where \( m(t) \) – vector function characterizing the weight of a train set

\( t \) – vector function characterizing the size of a train-to-train interval;
\( Q \) – vector function characterizing cargo turnover size;
\( Lr \) – vector function characterizing the size of a linear run – type of traction power supply;
\( L \) – vector function characterizing section way profile;
\( F \) – vector function characterizing distance between traction substations.

3. Time series approximation

The analysis of information describing the state of the studied system was carried out on the basis of the time series approximation \( W = W_j(t) \) [5].

To approximate the time series \( W = W_j(t) \) the following function was used

\[ W_j(t) = A + t \cdot [B + C \cdot \sin(\alpha t + \phi)]. \]

Parameters \( A, B, C, \alpha, \phi \) were determined by minimizing the error function using the modification of a gradient method described below.

To shorten the entry, let us denote the required vector of parameters through \( \mathbf{P} = [p_1, p_2, p_3, p_4, p_5] \). Let some initial value of this vector be given \( \mathbf{P}^0 = [p_1^0, p_2^0, p_3^0, p_4^0, p_5^0] \). The selection of this initial value determines which local minimum will be found as a result of calculations using the proposed algorithm. Let us also give the initial value of the vector step length (the meaning of which will be disclosed below) \( \mathbf{1}^0 = [\rho_1^0, \rho_2^0, \rho_3^0, \rho_4^0, \rho_5^0] \) [6].

A sufficiently small increment \( d > 0 \) is required to calculate the components of the gradient vector. The value \( d \) is selected as small as possible, in compliance with the requirement, so that the difference in values of the function to be minimized \( s(p_1^n, ..., p_j^n + d, ..., p_5^n) - s(p_1^n, ..., p_j^n, ..., p_5^n) \) can be calculated with the accuracy up to 2-3 significant figures for any \( j \).

The components of the gradient vector at point \( \mathbf{P}^n \)

\[ \mathbf{G}^{(\mathbf{P}^n)} = \begin{bmatrix} g_1^{(\mathbf{P}^n)} & g_2^{(\mathbf{P}^n)} & \ldots & g_5^{(\mathbf{P}^n)} \end{bmatrix} \]

represent values of partial derivatives

\[ g_j^{(\mathbf{P}^n)} = \frac{\partial s^{(\mathbf{P}^n)}}{\partial P_j}. \]

and with sufficient accuracy can be calculated numerically using the following expression
Next, the normalized gradient vector is calculated
\begin{equation}
\mathbf{T}(\mathbf{p}^n) = \begin{bmatrix} t_1(\mathbf{p}^n) & t_2(\mathbf{p}^n) & \ldots & t_s(\mathbf{p}^n) \end{bmatrix}^T,
\end{equation}
representing a unit vector in the direction of the fastest growth \( s(\mathbf{P}) \), wherein the components of vector \( \mathbf{T} \) are calculated through the components of vector \( \mathbf{G} \) using the following expression:
\begin{equation}
t_i(\mathbf{p}^n) = \frac{g_i(\mathbf{p}^n)}{\sqrt{\sum_{k=1}^{s} g_k(\mathbf{p}^n)^2}}.
\end{equation}

The predicted value \( \mathbf{P} \), for which the function \( s \) is expected to be smaller than the previous value, is calculated based on equality
\begin{equation}
\mathbf{P}^{n+1} = \mathbf{P}^n - L^n \mathbf{T}^n,
\end{equation}
where \( L^n \mathbf{T}^n \) – vector “composed” of component products of vectors \( \mathbf{L}^n \) and \( \mathbf{T}^n \), i.e.
\begin{equation}
L^n \mathbf{T}^n = \begin{bmatrix} \ell_1^n \ell_1^n & \ell_2^n \ell_2^n & \ldots & \ell_s^n \ell_s^n \end{bmatrix}^T.
\end{equation}

After obtaining a new value of vector \( \mathbf{P} \), the function \( s(\mathbf{P}) \) is calculated. In case all components of the vector are multiplied by \( q>1 \), i.e. \( L^{n+1} = q L^n \), the vector length of a pitch \( \mathbf{L} \) is increased in order to speed up the process of finding the minimum \( s(\mathbf{P}) \).

Otherwise, i.e. if \( s(\mathbf{P}^{n+1}) \geq s(\mathbf{P}^n) \), \( \mathbf{L} \) is corrected according to the following algorithm:
1. \( \mathbf{T}^{n+1} \) is calculated.
2. \( \mathbf{T}^{n+1} \) is compared by each component of vector \( \mathbf{T}^n \) and components of vector \( \mathbf{T} \), which signs have changed, are determined.
3. The components of vector \( \mathbf{L} \) corresponding to the numbers of \( \mathbf{T} \) components that changed the sign are multiplied by \( b, 0 < b < 1 \), i.e. if \( \text{sign}(t_i^{n+1}) \neq \text{sign}(t_i^n) \), then \( \ell_i^{n+1} = b \ell_i^n \), otherwise \( \ell_i^{n+1} = \ell_i^n \).
4. The actions according to the described algorithm are repeated with a reduced vector step length \( L^{n+1} \), and \( \mathbf{P}^n \) is taken again as a starting point, i.e. \( \mathbf{P}^{n+1} = \mathbf{P}^n \).

Computational experiments showed that the best values of \( q \) and \( b \) are \( q=1.1 \) and \( b=0.1 \). The initial vector step length does not significantly affect the calculation process, as usually 5..10 steps are sufficient for dynamic adjustment of \( \mathbf{L} \).

The most convenient criterion for completing a computational process is inequality \( |L^{n+1}| < d \), which corresponds to the reduction in a vector step length by finding the minimum \( s(\mathbf{P}) \) to values, at which the numerical definition of the gradient \( \mathbf{G}(\mathbf{P}) \) loses its meaning.

Figure 3 shows the forecasting results.
Calculations showed that this variation of the algorithm increases the accuracy of electric energy consumption forecast by 4.7%.

4. Neural networks in forecasting power consumption

Today, hybrid systems based on complex mathematical machines of fuzzy logic [7, 8] and neural networks [9, 10] are becoming ever more popular along with traditional methods of formation of fuel and energy balance of an enterprise and forecasting that provide for more accurate description of processes capable of self-learning in the process of operation and producing more accurate results.

The advantages of such systems are caused by a number of factors:

1) complexity of mathematical models of real control systems and processes related to the desire to improve their adequacy and to take into account a growing number of different factors influencing the decision-making process;

2) focus on models that take into account incomplete and inaccurate source data;

3) possibility of application in the description of technical systems, which have uncertainty and accuracy of initial data, makes it difficult or even impossible to use accurate quantitative methods and approaches.

Next, let us consider possible options for generating a learning sample in order to build a learning curve that is most adequately solved. The essence of neural networks learning process is to perform the following multi-step procedure [7].

Step 1. The learning set (set of problems) is specified

\{(X_1, D_1); (X_2, D_2); ...; (X_L, D_L)\},

The elements of which represent learning pairs \((X_i, D_i)\), \(i=1, 2, ..., L\). In this case \(X_i = \left[ x_{i1}^{(1)} x_{i2}^{(1)} ... x_{in}^{(1)} \right]^T\) is the 1st input vector (or the 1st input image) proposed by neural networks; \(D_i = \left[ d_{i1}^{(1)} d_{i2}^{(1)} ... d_{im}^{(1)} \right]^T\) – vector of reference (required) neural networks reactions in response to the 1st input vector \(X_i\); \(L\) – number of different learning pairs of the sampling.

Step 2. The initial state of neural networks is set by assigning some random small values to all its weights \(w_{ij}^k\). Here \(W_{ij}^k\) – link weight connecting the output of i neuron of k layer with the input of j neuron of \(k+1\) layer.
Step 3. The input vector \( \mathbf{X}_1 \) is sent to network input and reactions \( y^{(i)} \) (i=1,2,..., m) of the output layer neurons are determined.

Step 4. The difference \( \mathbf{e}^{(1)} \) between desirable reaction of network \( \mathbf{D}_1 \) and its actual exit \( \mathbf{Y}_1 = \begin{bmatrix} y_1^{(0)} & y_2^{(0)} & \cdots & y_m^{(0)} \end{bmatrix} \), i.e. \( \mathbf{e}^{(1)} = \mathbf{D}_1 - \mathbf{Y}_1 \) and a total squared error \( E^{(1)} = 0.5 \sum_{j=1}^{m} (d_j - y_j^{(0)})^2 \) is calculated.

Step 5. The weights \( w^k_{ij} \) of neural networks are corrected to reduce error \( E^{(1)} \).

Step 6. Steps 3...5 are repeated for each pair of the learning set \( (\mathbf{X}_1, \mathbf{D}_1) \) until the error in the whole set does not reach small predetermined value \( E^* \).

The learning result is such adjustment of synaptic link weights in neural networks when each input vector the network generates the required (or close to it) output. The obtained forecast values are shown in Table 2.

| Month        | Actual electrical energy consumption, kWh | Forecast value of the consumer with maximum power consumption kWh | Error, % | Forecast power consumption of the neural network model kWh | Error, % |
|--------------|------------------------------------------|---------------------------------------------------------------|--------|----------------------------------------------------------|--------|
| January      | 7199.5                                   | 6784.5                                                          | 4.15   | 7098.591                                                 | 1.00909 |
| February     | 6007.5                                   | 6591.2                                                          | -5.837 | 6157.888                                                 | -1.50388|
| March        | 6885.3                                   | 7055.5                                                          | -1.702 | 7005.913                                                 | -1.20613|
| April        | 6496.1                                   | 6918.2                                                          | -4.221 | 6601.133                                                 | -1.05033|
| May          | 6471.8                                   | 6796.2                                                          | -3.244 | 6601.492                                                 | -1.29692|
| June         | 6098.6                                   | 5887.4                                                          | 2.112  | 6192.307                                                 | -0.93707|
| July         | 6890.5                                   | 6450.1                                                          | 4.404  | 6997.888                                                 | -1.07388|
| August       | 6685.3                                   | 7171.2                                                          | -4.859 | 6770.667                                                 | -0.85367|
| September    | 6781.1                                   | 7054.4                                                          | -2.733 | 6884.785                                                 | -1.03685|
| October      | 6917.1                                   | 6717.9                                                          | 1.992  | 7057.237                                                 | -1.40137|
| November     | 7708.8                                   | 7250.1                                                          | 4.587  | 7802.307                                                 | -0.93507|
| December     | 8066.6                                   | 7712.1                                                          | 3.545  | 8196.959                                                 | -1.30359|

Analyzing the early obtained data (Table 1), let us carry out a comparative analysis on the error level for two forecast options. The result is presented as a comparative diagram in Figure 4.
5. Conclusion
Based on the above diagram it can be concluded that the error of forecast based on approximation for the 1st level electric consumer made from 4 to 6%. The application of neural networks in forecasting of power consumption allows reducing the value of error to 2%, thus leading to significant reduction of costs for fuel and energy balance of a structural division or an enterprise as a whole.

References
[1] Rawlinson S, Pike C, Starmans M 2017 Digital railway Digital Rail UK.
[2] Timur Bek 2018 Digital Road Economy Gudok Issue 84.
[3] Zaria Mohammed Shamseidin 2015 Short Term Electrical Load Forecasting Using Fuzzy Logic, A Dissertation Submitted In Partial Fulfillment for the Requirements of the Degree of M.Sc. In Electrical Engineering (Microprocessor and Control).
[4] Q. Ren, L. Baron, and M. Balazinski 2006 Type-2 takagi-sugeno-kang fuzzy logic modeling using subtractive clustering in Fuzzy Information Processing Society Annual meeting of the North American 120-125.
[5] Zakaryukin V.P., Krukov A.V., Raevsky N.V., Yakovlev D.A. 2017 Modeling and forecasting of power consumption processes of railway transport.
[6] Hianikova T 2008 Application of Fuzzy logic in financial time series analysis.
[7] K. Bataine, M. Naji and M. Saqer 2011 A comparison study between various fuzzy clustering algorithms Editorial board 335.
[8] Osipova V.E, Yakovlev D.A. 2019 The Methodology of the Department of Fuel and Energy Resources of the Enterprise Based on Fuzzy Logic Journal of Siberian Federal University. Engineering &amp; Technologies.
[9] Saha S 2013 Comparison of Performance Analysis using Different Neural Network and Fuzzy
Logic Models for Prediction of Stock Price.

[10] Osipova V.E., Komogortsev M.G. 2016 Automated calculation of train-to-train intervals using the apparatus of artificial neural networks Transsib bulletin 4 (28).