The use of correlation and regression analysis for assessment of the energy effectiveness of the dc electric locomotives auxiliary equipment

Stanislav Istomin

Abstract. Through additional processing of the modern movement parameter recorders data of the DC electric locomotive 2ES6 the article first presents the results of the actual consumption of electricity for own needs and the proportion of these costs from the consumption of trains traction is determined, which in terms of operational depot is difficult to implement. The estimation of influencing factors on the energy consumption for own needs of 2ES6 series electric locomotives is made. As a result it was found that the internal energy consumption is influenced by such factors as rolling stock mass, axle load and environment temperature. Statistic models were made to normalize internal electricity consumption and their quality estimation was fulfilled. It is found that the remainders of the multiple regression equation, which take the above factors into account, obey the normal distribution law, indicating the adequacy of their further use to assess the energy efficiency of the 2ES6 series DC electric locomotives auxiliary equipment. The use of regression models will allow to identify electric locomotives with auxiliary equipment with low energy efficiency and to send them to unscheduled repairs in time to restore the required technical condition.

1 Introduction

One of the priority tasks of the energy strategy of the Russian Railways holding for the period up to 2015 and for the future up to 2030 is a significant increase in the energy efficiency of train traction. Thus, in general, OAO "Russian Railways" expected reduction in specific consumption of traction and energy resources on the traction of trains to the level of 2015 should be by 2020 – 2.5-4.4%, by 2030 8.0-9.0[%1]. To implement this priority task, in particular, measures are aimed to improve the electric rolling stock (ERS) energy characteristics, when they are created at the manufacturers or modernization of operating locomotives by selecting the optimal energy consumption parameters of auxiliary electrical equipment and the organization of control and management of its energy-efficient operation.

At present the biggest part of ERS of OAO "Russian Railways doesn’t have internal electricity consumption record, which hampers the assessment of the subsidiary equipment energy efficiency and the development of recommendations for its improvement. In general

* Corresponding author: istomin_sg@mail.ru
the internal energy consumption assessment can be performed analyzing auxiliary equipment parameters, given in the ERS user manuals; however such assessment method doesn’t allow to reliably define the value and nature of energy consumption of that equipment, since its operating modes are not taken into account.

The most reliable data of ERS internal energy consumption could be received experimentally. Experimental investigations [2, 3], carried out on the trans-Siberian railway by the scientists of OSTU for DC electric locomotives of the VL10 series and the scientists of IrSTU for AC electric locomotives of the VL85 series, show that the internal energy consumption of the DC ERS is up to 10% of the total consumption for the traction of trains, and for the AC ERS up to 20%, with the main part of the energy consumption is used for the ventilation system.

Quite a high value of internal energy consumption of the ERS of the old series is primarily due to the lack of an automated system for monitoring and controlling the energy consumption by the ERS auxiliary equipment, taking into account the influence of various random factors that determine the energy-saving modes of its operation.

2 Object and methods of research

On DC electric locomotives of 2ES6 series equipped with an auxiliary converter, which allows to control asynchronous drives of auxiliary machines, there is no accounting for electric power for own needs; however, on the cartridges of the registrars of the electric locomotive 2ES6 movement parameters, being the components of the microprocessor-based control and diagnostic systems (MPSUD), there is information about the current of electric locomotive own needs of and the voltage on the current collector, which allows to obtain information about the internal energy consumption, while additional processing of the data array, and to evaluate the energy efficiency of auxiliary equipment [4, 5].

For the first time the authors tested such method of analysis in relation to the section Moskovka – Barabinsk with a length of 316 km, which is characterized mainly by a flat profile of the path. The sample was made up of trips of empty and loaded trains with 2ES6 series electric locomotives, operating at this section throughout the calendar year. As a result of data processing of registrar cartridges of the 2ES6 series electric locomotives movement parameters, it was determined that the share of the energy consumption for own needs from the general energy consumption in the even direction makes from 2,91% to 8,66% (figure 1), and in odd – from 2,72 (figure 2).

These results show that the share of energy consumption for own needs of electric locomotives 2ES6 from the total energy consumption as well as on the VL10 electric locomotive reaches 10%. However, the presence of only this information does not let conclude that the use of a static converter on 2ES6 electric locomotives, allowing to control asynchronous drives of auxiliary machines, is unreasonable, since the experiments are performed in different areas and under different conditions. Based on the obtained results it is possible to make an unambiguous conclusion that there is a considerable potential for reducing energy consumption for own needs, which is possible to realize due to improvement of the static converter operation algorithms, selection of the optimal consumption parameters of auxiliary electrical equipment through the study of the influence of various random factors that determine energy-saving modes of operation.
Fig. 1. The share of energy consumption for own needs from the total energy consumption for traction on the section Moskovka-Barabinsk (even direction).

Fig. 2. The share of energy consumption for own needs from the total energy consumption for traction on the section Moskovka-Barabinsk (odd direction).

To identify the influencing factors on the specific energy consumption (SEC) for the own needs of 2ES6 electric locomotives and to form statistical models to assess their energy efficiency, a correlation and regression method of mathematical statistics was used. The main calculations were made in the program Microsoft Excel using the package "Data analysis" based on the information of the traffic parameters recorders of the sample trips with 2ES6 electric locomotives in an even and odd direction throughout the calendar year.

3 Results

The investigation was carried out in several stages.

At the first stage such factors were chosen as the rolling stock mass (m), axle load (no), atmospheric temperature (t), average current collector voltage (Ucp), number of stops (N) and electric locomotive schedule speed (Vy).

At the second stage of the investigation for a set of factors mentioned above, the variants of statistical models (table 1) were formed in order to further assess their adequacy and statistical significance of these models coefficients.
Table 1. Variants of statistical models to assess their adequacy and statistical significance of these models coefficients.

| No model | The variant of statistical model |
|----------|---------------------------------|
| 1        | $a_{c.h.} = B_0 + B_1 \cdot m + B_2 \cdot n_o + B_3 \cdot t + B_4 \cdot U_{cp} + B_5 \cdot N + B_6 \cdot V_{mes}$ |
| 2        | $a_{c.h.} = B_0 + \frac{B_1}{m} + \frac{B_2}{n_o} + B_3 \cdot t + B_4 \cdot U_{cp} + B_5 \cdot N + B_6 \cdot V_{mes}$ |

In the third stage, the coefficients B of the generated statistical models are determined, the standard error, the t-statistic, the p-value, the lower and upper limits of the confidence interval (tables 2 and 3), and the adequacy of these models is checked by evaluating the significance level $\alpha$ (Microsoft Excel refers to this indicator as p). P-value is the error probability when the null hypothesis deviates, i.e. the hypothesis that all regression coefficients are zero. If the p-value is less than 0.05, the hypothesis of insignificance of regression is rejected and it is considered that regression is significant. In our case, for the selected statistical models, the p-value does not exceed the significance level of 0.05.

At the fourth stage, the statistical significance of regression coefficients for equations №1 and №2 from table 1 is estimated. The results of the evaluation are shown in tables 2 and 3.

Table 2. The results of estimation of the statistical significance of regression coefficients for equations No 1.

| Factors                        | Direction | Regression coefficient B | Standard mistake | t-statistics | P-value | Lower limit of the confidence interval | Upper limit of the confidence interval |
|--------------------------------|-----------|--------------------------|------------------|--------------|---------|---------------------------------------|---------------------------------------|
| Free regression term           | Even      | 20,8                     | 11,8             | 1.75         | 0.083   | -2.79                                 | 44.4                                  |
|                                | Odd       | 12.9                     | 5.06             | 2.55         | 0.012   | 2.93                                  | 22.9                                  |
| Rolling stock mass, t          | Even      | -0.00063                 | 0.00025          | 1.75         | 0.016   | -0.0011                               | -0.00012                              |
|                                | Odd       | -0.00057                 | 0.00016          | -3.64        | 0.0004  | -0.0008                               | -0.00026                              |
| Axle load, t                   | Even      | -0.206                   | 0.072            | -2.89        | 0.005   | -0.349                                | -0.064                                |
|                                | Odd       | -0.098                   | 0.047            | -2.11        | 0.036   | -0.191                                | -0.006                                |
| Temperature, °C                | Even      | -0.062                   | 0.012            | -5.28        | $9.2 \times 10^{-7}$                 | -0.086                                | -0.039                                |
|                                | Odd       | -0.064                   | 0.0063           | -10.1        | $9.9 \times 10^{-19}$                | -0.076                                | -0.051                                |
| Average voltage on the current collector, V | Even      | -0.0026                  | 0.003            | -0.77        | 0.439   | -0.009                                | 0.004                                 |
|                                | Odd       | -0.00079                 | 0.00148          | -0.53        | 0.58    | -0.004                                | 0.002                                 |
| Number of stops, pcs           | Even      | 0.0247                   | 0.303            | 0.081        | 0.94    | -0.578                                | 0.627                                 |
|                                | Odd       | 0.095                    | 0.173            | 0.55         | 0.59    | -0.245                                | 0.437                                 |
| Schedule speed, km/h           | Even      | -0.027                   | 0.04             | -0.66        | 0.506   | -0.106                                | 0.053                                 |
|                                | Odd       | -0.034                   | 0.021            | -1.63        | 0.106   | -0.076                                | 0.007                                 |
Table 3. The results of estimation of the statistical significance of regression coefficients for equations No 2.

| Factors                        | Direction | Regression coefficient B | Standard mistake | t-statistics | P-value | Lower limit of the confidence interval | Upper limit of the confidence interval |
|--------------------------------|-----------|--------------------------|------------------|--------------|---------|----------------------------------------|---------------------------------------|
| Free regression term           | Even      | 9,78                     | 8,54             | 1,15         | 0,26    | -7,19                                  | 26,7                                  |
|                                | Odd       | 3,4                      | 5,13             | 0,66         | 0,51    | -6,72                                  | 13,5                                  |
| Rolling stock mass, t          | Even      | 12055                    | 1822             | 6,62         | **2.8×10^{-9}** | 8434                                  | 15676                                 |
|                                | Odd       | 8404                     | 1495             | 5,62         | **8.7×10^{-8}** | 5451                                  | 11356                                 |
| Axle load, t                   | Even      | 11,1                     | 7,67             | 1,44         | 0,152   | -4,17                                  | 26,3                                  |
|                                | Odd       | 13,0                     | 5,43             | 2,39         | **0,018** | 2,29                                  | 23,7                                  |
| Temperature, °C                | Even      | -0,061                   | 0,0089           | -6,85        | **9.5×10^{-10}** | -0,078                                | -0,043                                |
|                                | Odd       | -0,063                   | 0,0064           | -9,83        | **6.0×10^{-18}** | -0,075                                | -0,05                                 |
| Average voltage on the current collector, V | Even | -0,0025                   | 0,0024           | -1,04        | 0,298   | -0,0073                                | 0,0023                                |
|                                | Odd       | -0,0001                   | 0,0015           | -0,077       | 0,934   | -0,003                                 | 0,0029                                |
| Number of stops, pcs          | Even      | 0,21                      | 0,23             | 0,91         | 0,36    | -0,246                                | 0,664                                 |
|                                | Odd       | 0,051                     | 0,177            | 0,289        | 0,773   | -0,298                                | 0,399                                 |
| Schedule speed, km/h           | Even      | -0,017                    | 0,03             | -0,554       | 0,581   | -0,077                                | 0,044                                 |
|                                | Odd       | -0,034                    | 0,022            | -1,57        | 0,119   | -0,077                                | 0,009                                 |

Let us consider the mechanism of estimating the statistical significance of regression coefficients on the example of table 2. In the dependent variable (specific energy consumption for the 2ES6 electric locomotive own needs), a greater contribution is made in the even and odd direction – the temperature of the air (p=9.2×10^{-7} and p=9.9×10^{-19}), and a smaller contribution in the even and odd direction – the number of stops (p=0.94 and p=0.59). A negative sign of the coefficients for the variables means that with an increase, for example, in an even direction of the rolling stock mass, the axle load and the temperature, the specific energy consumption rate falls. A positive sign of the coefficient for the variable (number of stops) means that with the increase in the number of stops, the level of specific consumption increases. The regression equation coefficients B1, B2, B3 for even and odd directions are statistically significant at the significance level p=0.05; the regression equation coefficients B4, B5, B6 are statistically insignificant, since p>0.05) [6].

As a result of the above analysis, it was found that statistically significant factors affecting the SEC to the 2ES6 electric locomotives own needs in an even and odd direction are the rolling stock mass, the axle load and the atmospheric temperature for equation No. 1, and the mass of the rolling stock and the atmospheric temperature for equation No. 2 in an even direction, and in an odd one - the rolling stock mass, the axle load and the atmospheric temperature.

At the fifth stage, on the basis of the above results, statistical models were formed, taking into account the above statistically significant factors, to assess the energy efficiency of the 2ES6 electric locomotives auxiliary equipment and its quality was assessed (table 4).
Table 4. Comparison of determination coefficients $R^2$ for different statistical models.

| Direction | No model | The variant of statistical model | Determination coefficient $R^2$ |
|-----------|----------|----------------------------------|-------------------------------|
| Even      | 1        | $a_{c.h.} = 10,3 - 0,0006 \cdot m - 0,214 \cdot n_o - 0,061 \cdot t$ | 0,587                         |
|           | 2        | $a_{c.h.} = 0,43 + \frac{14040}{m} - 0,059 \cdot t$ | 0,752                         |
| Odd       | 3        | $a_{c.h.} = 8,27 - 0,0006 \cdot m - 0,097 \cdot n_o - 0,061 \cdot t$ | 0,645                         |
|           | 4        | $a_{c.h.} = 1.05 + \frac{8390}{m} + \frac{13.3}{n_o} - 0,061 \cdot t$ | 0,637                         |

To assess the quality of forecasting, the coefficient of determination $R^2$ is used:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{a}_i - \bar{a})^2}{\sum_{i=1}^{n}(a_i - \bar{a})^2},$$

where $\sum_{i=1}^{n}(\hat{a}_i - \bar{a})^2$ - is a sum of squares of regression remainders;

$\sum_{i=1}^{n}(a_i - \bar{a})^2$ - is a total sum of squares.

This coefficient gives a numerical estimate of how the variability of the predicted value is explained by the factors included in the model.

It should be noted that in order to analyze various regression equations, the regression coefficients of these equations were determined for all values of the sample from the general population.

According to the results of the study (table 4), the following conclusions can be made:

- the values of $R^2$ in all cases exceed 0.5, which indicates a strong relationship between $a_{c.h.}$ and the selected factors;
- the use of the model № 2 in the even direction and № 3 in the odd direction will allow to achieve a much more reliable determination of the SEC standard for the 2ES6 electric locomotives own needs in comparison with the models № 1 and №4 (higher values of $R^2$ are obtained) and to perform a qualitative assessment of the energy efficiency of their auxiliary equipment. It is also important to note that these parameters (axle load, rolling stock weight and temperature) are known before the trip, which allows further to carry out operational forecasting of the SEC for the own needs of the 2ES6 electric locomotives. An important prerequisite for the correct use of the regression equation is the fact that the remainders obey the normal law (figure 3).
The regression equation remainders

Fig. 3. Distribution of the regression equation remainders according to the normal law.

Test of normality of remainders was performed using the goodness of fit Pearson

\[ \chi^2 = \sum_{i=1}^{k} \left( \frac{n_i - n_i'}{n_i} \right)^2, \]

where \( n_i \) are the observed frequencies in the I-th grouping interval;

\( n_i' \) - theoretical frequencies.

In the considered cases, at the level of significance of 0.05, there was no reason to reject the hypothesis of normality. The observed values were lower than the critical ones.

In the end, according to the obtained regression equations, the corresponding norms of SEC for own needs of the 2ES6 electric locomotives were determined and their convergence with the actual values was analyzed (figure 4 and 5).

Fig. 4. Convergence of the actual and calculated values of the specific power consumption for own needs 2ES6 electric locomotives on the Moskovka-Barabinsk section (even direction).
According to the results of the analysis, it can be concluded that the obtained regression equations № 2 and № 3 from table 4 accurately describe the nature of the change in the SEC for the own needs of electric locomotives 2ES6, so they can be used to assess the energy efficiency of the auxiliary equipment of these electric locomotives on the Moskovka – Barabinsk section.

**Fig. 5.** Convergence of the actual and calculated values of the specific power consumption for own needs 2ES6 electric locomotives on the Moskovka-Barabinsk section (odd direction).

**4 Conclusions**

The studies conducted by the authors showed the reasonability of using the correlation-regression methods of mathematical statistics to assess the energy efficiency of auxiliary equipment of 2ES6 electric locomotives on the example of one section, however, the final answer about the suitability of these methods may be given by conducting full-scale studies on sections with different types of track profile.

**References**

1. V.A. Gapanovich, S.N. Epifancev, V.A. Ovsejchuk, *Energy Strategy and the Russian railways electrification* (Eko–Press Publ., Moscow, 2012)
2. A.A. Baklanov, P.Y. Medlin, *The analysis of VL10 electric locomotive energy consumption for own needs* (Omsk, 1974)
3. A.A. Baklanov, *Influence of different factors on auxiliary equipment energy consumption of an electric locomotive VL10* (Omsk, 1977)
4. D.I. Bondarevskiy, The Trans-Siberian Bulletin, 42-46 (2016)
5. S.G. Istomin, D.I. Bondarevskiy, The Trans-Siberian Bulletin, 151-156 (2017)
6. A.A. Halafyan, *STATISTICA 6. Statistical data analysis* (OOO «Binom-Press», Moscow, 2007)