Joint Institute for Nuclear Research

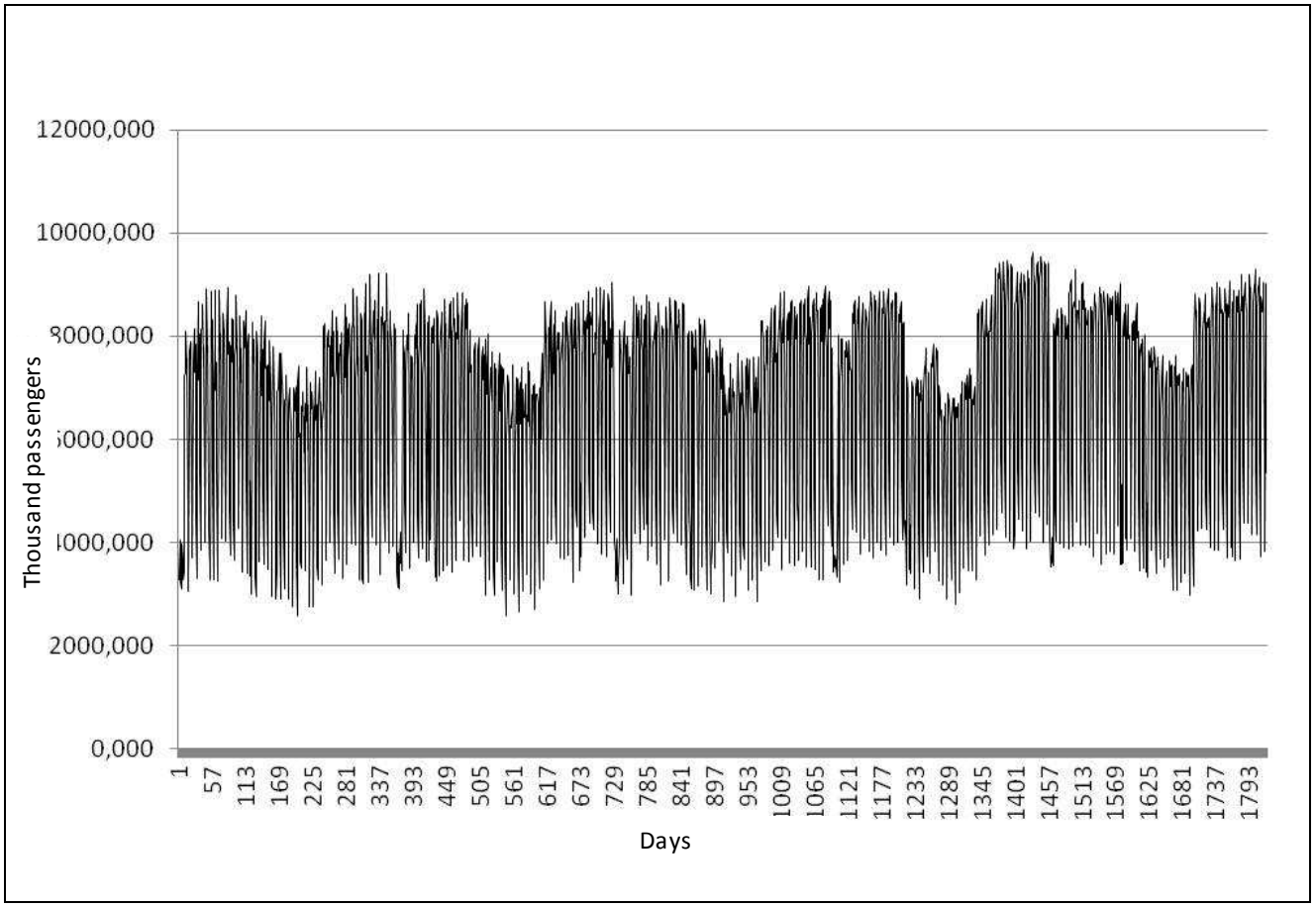
Application of artificial neural networks and singular-spectral analysis in forecasting the daily passenger's traffic in the Moscow metro

E.S. Osetrov

V.V. Ivanov

Dubna, 2017

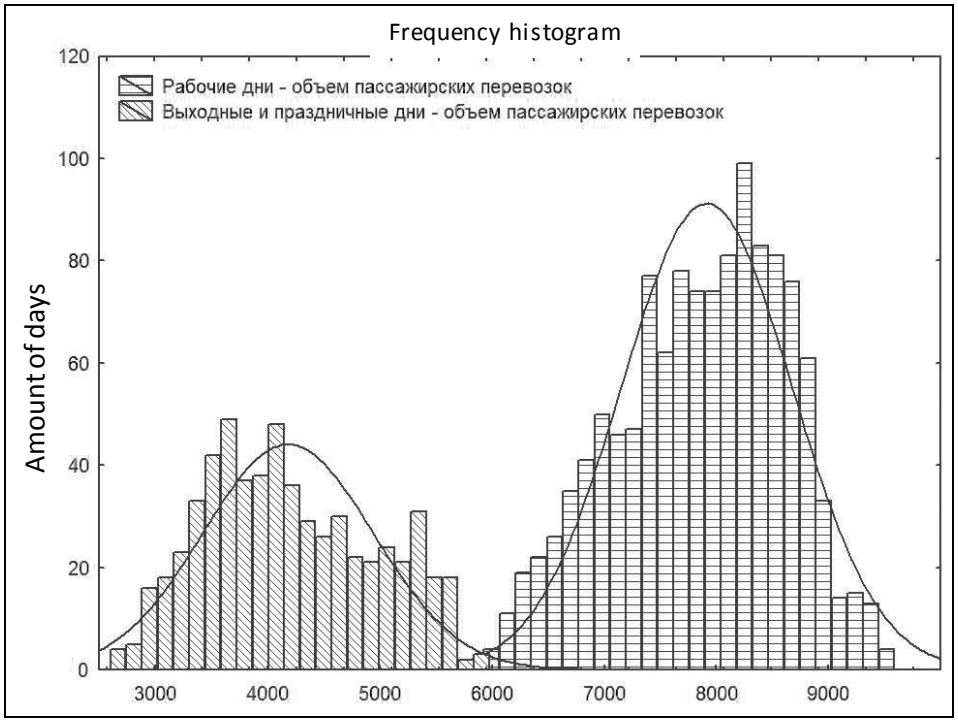
1. Analysis of the time series describing passenger traffic



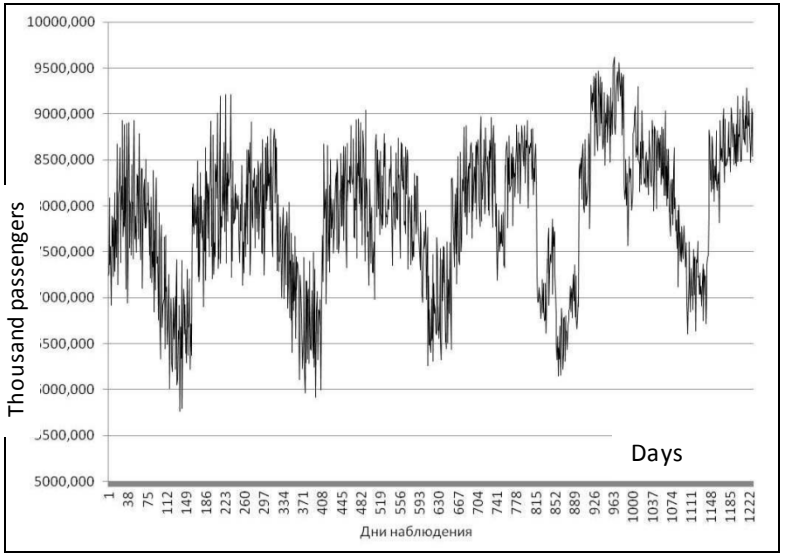
The main reason for the spread of passenger traffic - The number of passengers differ on workdays (large values) and weekend-holidays (lower values).

Time series, reflecting the dynamics of the daily volume of the passenger traffic by the Moscow subway over the past 5 years.

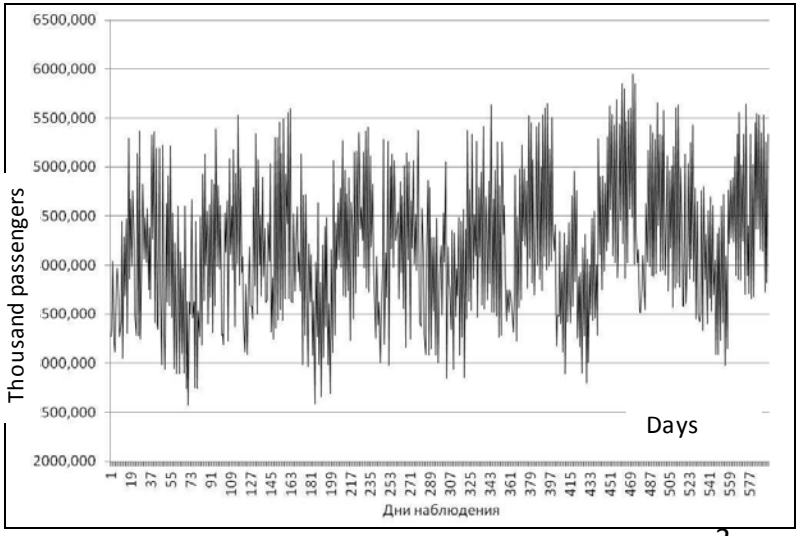
1. Analysis of the time series describing the traffic of passengers



Distributions of the daily volume of passenger traffic to working days (right) and weekends - holidays (left)

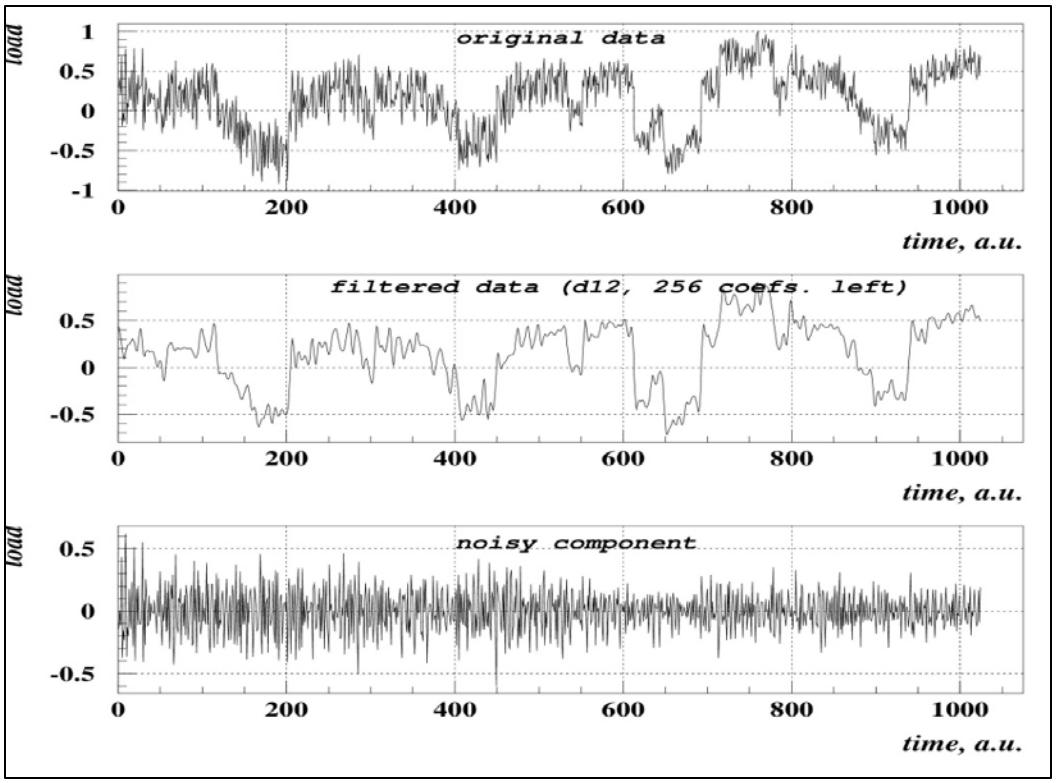


Dynamics of the daily volume of passenger traffic on working days and days off



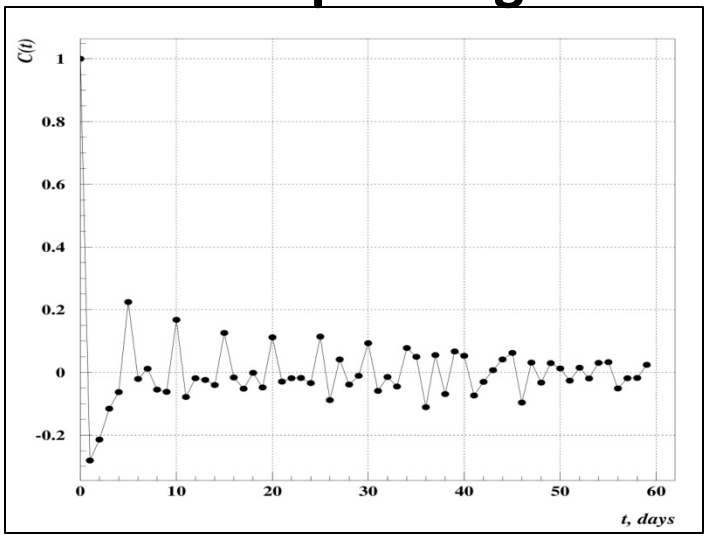
The dynamics of volumes on working days is more regular and not so much noisy as on weekends and holidays

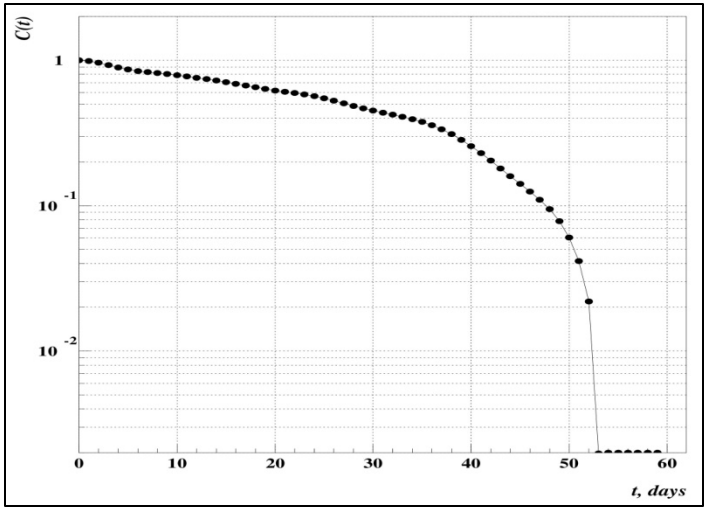
1. Analysis of the time series describing the traffic of passengers



Time series: initial, after application of wavelet filtration and noise component

The filtered series provides good opportunities for a medium-term forecast





The behavior of the autocorrelation function as a function of the correlation interval for the noise and regular components

2. Forecasting of passenger traffic with Artificial Neural Network (ANN)

Baseline data and factors

To construct the forecast model (time series) from the total of 15 variables, after the correlation analysis, the following variables were selected:

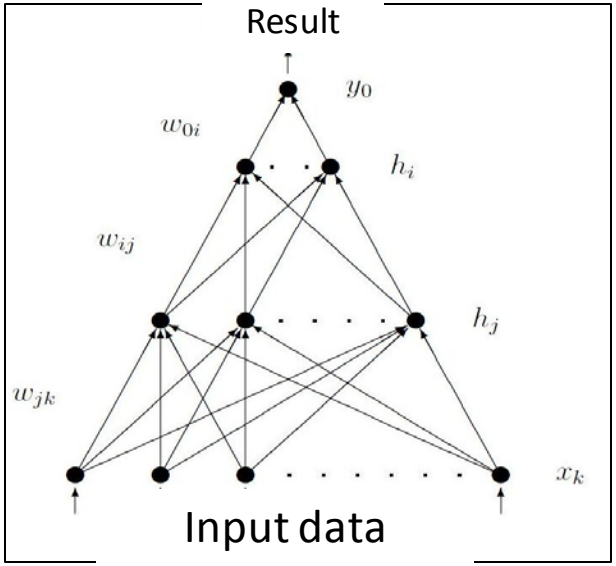
Var1 (Year of observation),
Var2 (Month),
Var3 (Day of the week),
Var4 (Type of day),
Var5 (Deviation of daytime temperature in the region from the norm)
Var6 (Daily electricity consumption in the Moscow region)
Var7 (Daily volume of passenger traffic in Moscow Metro).

The sample of initial data was 1024 observations on 7 variables

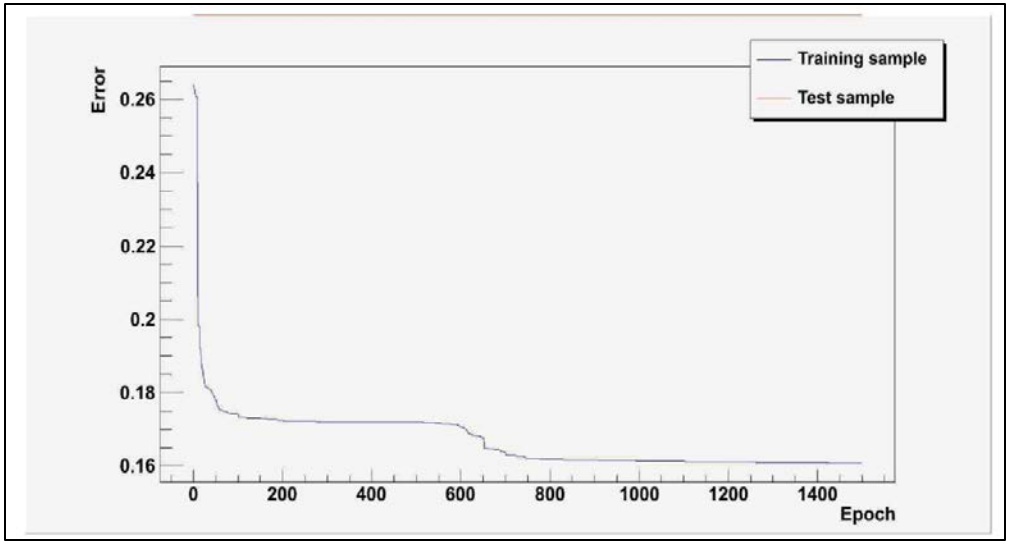
The selected variables (factors) are characterized by:

- there is no phenomenon of multicollinearity - a strong correlation (> 0.8) between different variables;
- a significant character for INS, due to the seasonality of the forecast value;
- reliable predictive values can be obtained using alternative methods of obtaining information.

2. Forecasting of passenger traffic with Artificial Neural Network (ANN)



The scheme of a multilayer perceptron with two hidden layers and one output neuron

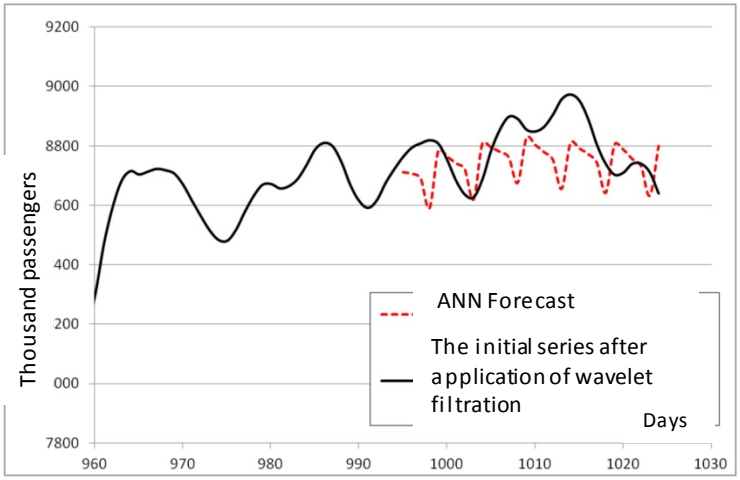


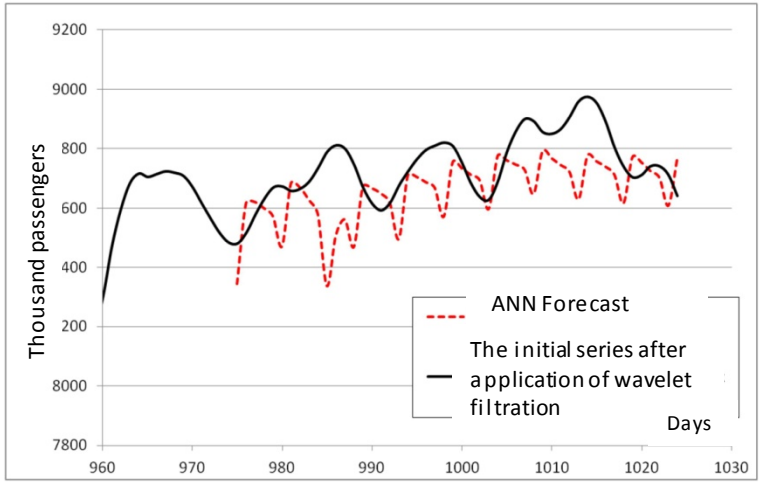
A typical schedule for changing the error when learning the ANN, depending on the epoch number: the TMVA package in the ROOT environment, the Fletcher-Reeves training method

ANN input parameters and setting::

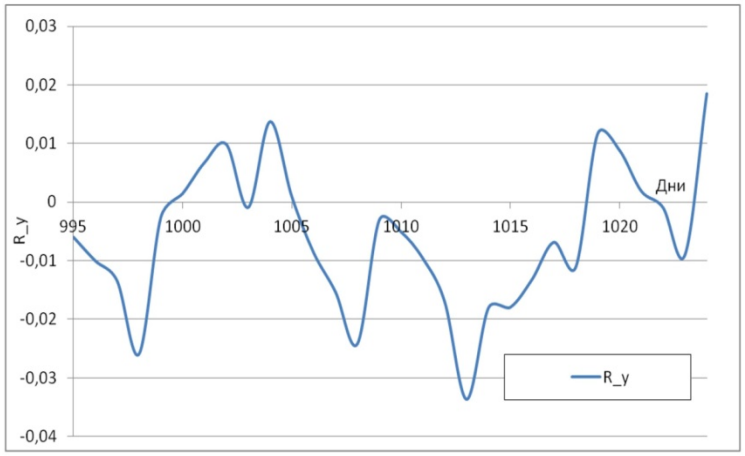
1) 7 neurons at the input, the first hidden layer contained 16 neurons, the second - 8 neurons, one output neuron.
2) Network training on samples of 994 or 974 observations.
3) The forecast horizon is 30 and 50 days.
4) The procedure of teaching 1,500 eras.
5) The training method is determined after a series of experiments - Fletcher-Reeves (showed the best results).
6) Input variables Var1-Var7 were given to the range [-1; +1].
7) After the INS training, testing was conducted.
8) When testing the network, the predicted value was fed to its input

2. Forecasting of passenger traffic with Artificial Neural Network (ANN)

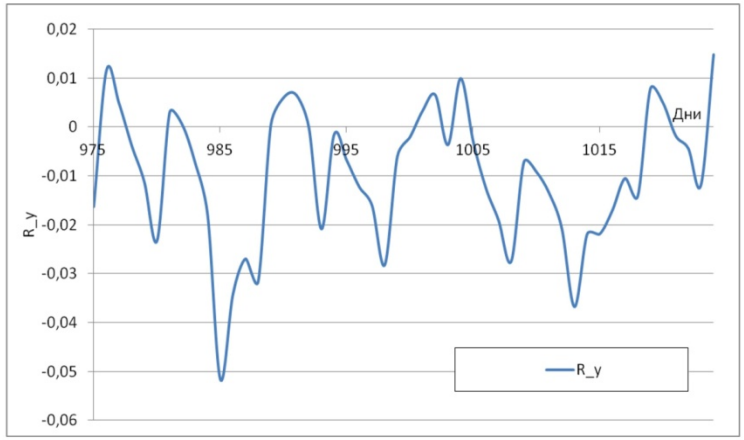




The results of forecasting ANN at 30 (left) and 50 (right) observation days in comparison with actual data

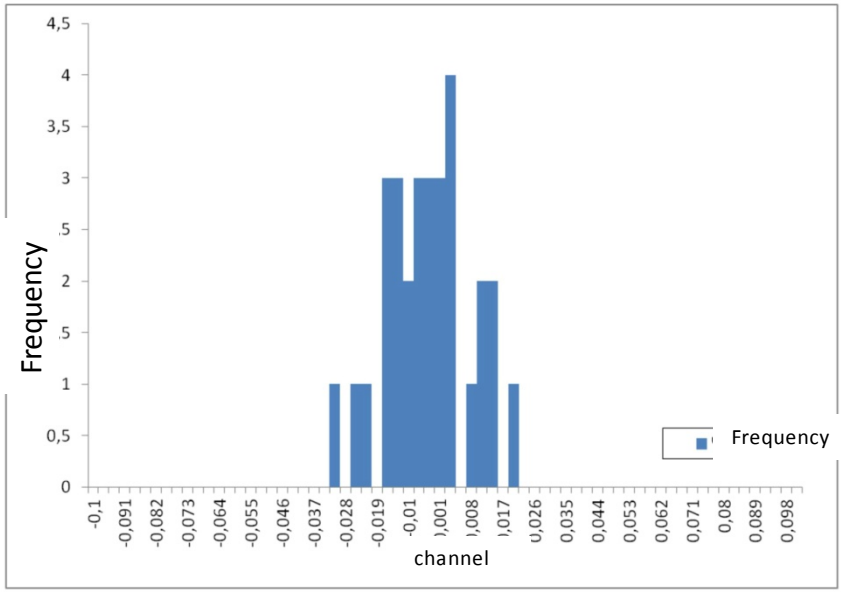


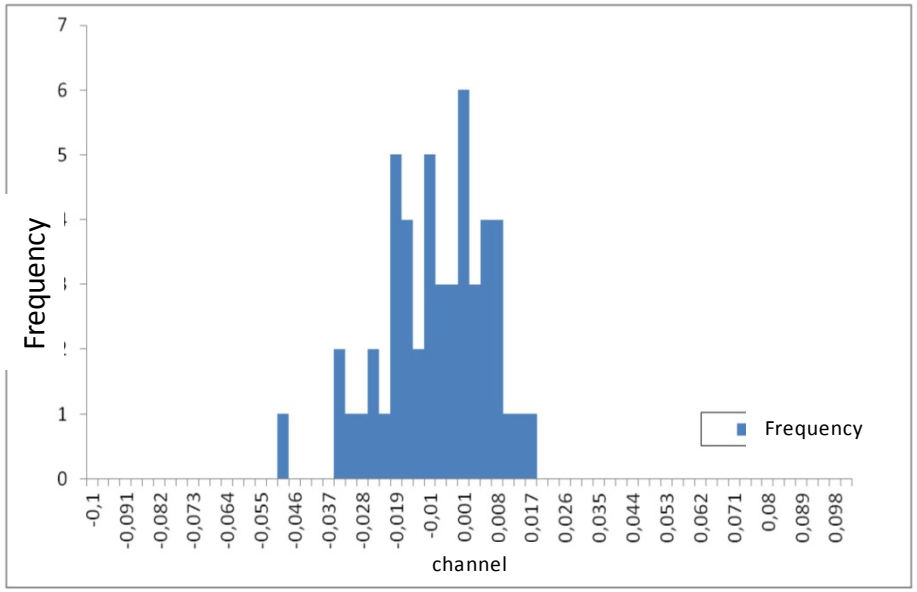
$$R_y = \frac{y_{real} - y_{prog}}{y_{real}},$$



Dynamics of relative error R_y in depending on the ordinal number of the observation day when predicting with the help of SMEs at 30 (left) and 50 (right) days

2. Forecasting of passenger traffic with Artificial Neural Network (ANN)





The distributions of the variable R_y at the prediction at 30 (left) and 50 (right) days

The results of ANN forecasting:

1. **The relative error R_y is distributed symmetrically with respect to zero and over the entire prediction interval it is just over 3% with a forecast of 30 days and does not exceed 5% with a forecast of 50 days.**

2. **Usning the ANN is possible to carry out a medium-term forecast of the passenger traffic volumes in the Moscow Metro**

3. Prediction based on Singular-Spectral Analysis (SSA)

The main stages of the "Caterpillar" –SSA method

1. transformation of a one-dimensional time series to a multidimensional form;
2. decomposition of a multidimensional series at singular points;
3. Analysis of this decomposition using the principal component method and selection of the main components;
4. reconstruction of a one-dimensional time series based on the selected components.

The transformation of a one-dimensional time series to a multidimensional form is realized by representing it in the form of a trajectory matrix:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{k,L} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_L \\ x_2 & x_3 & x_4 & \dots & x_{L+1} \\ x_3 & x_4 & x_5 & \dots & x_{L+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \dots & x_K \end{pmatrix}$$

где $L < K$ называется длиной гусеницы, а $k = K - L + 1$.

Затем находятся собственные значения λ_i, $i = 1, 2, \dots, L$ и собственные вектора $\mathbf{V} = (v_1, v_2, ..., v_l)$ ковариационной матрицы $C = \frac{1}{k} X X^T$. Матрицу $\mathbf{V}$ можно рассматривать как матрицу перехода к главным компонентам

$$\mathbf{Y} = \mathbf{V}^{\mathbf{T}}\mathbf{X} = (y_1, y_2, \dots, y_L), \qquad (9)$$

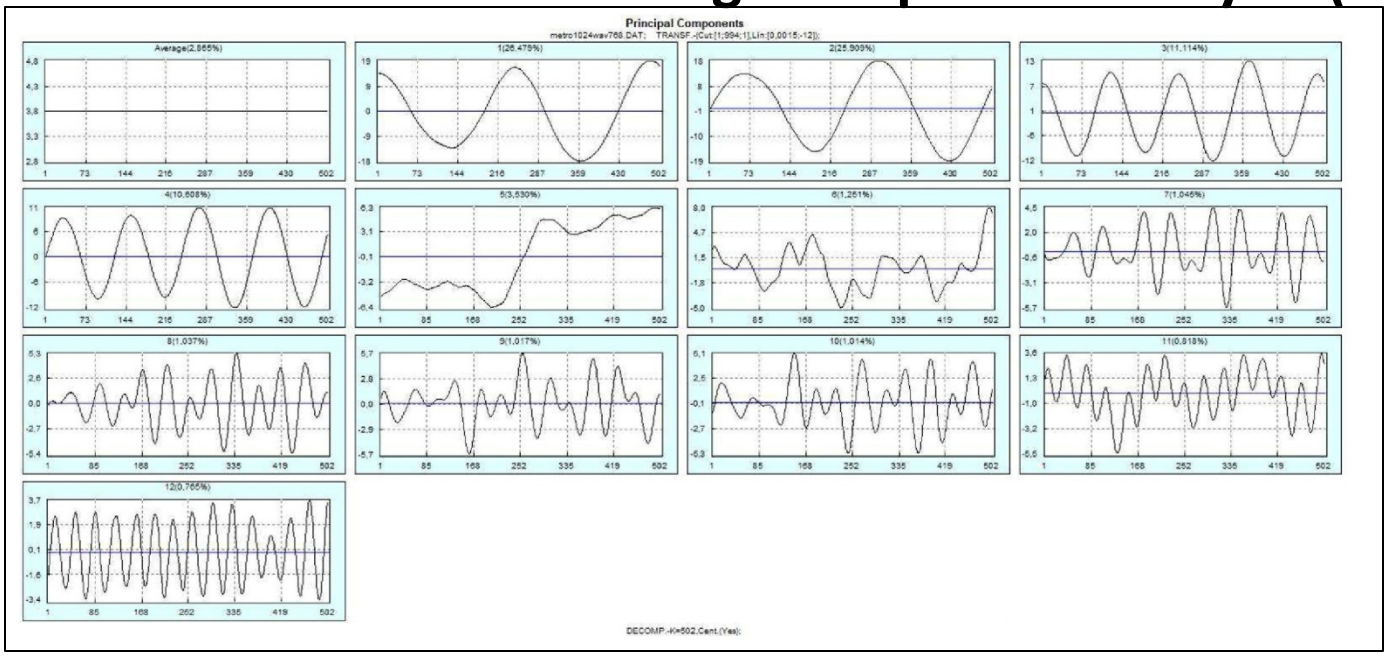
где y_i $(i = 1, 2, \dots, L)$ – столбцы матрицы, состоящие из k элементов.

Relation

$$\sum_{i=1}^{L} \frac{\lambda_i}{L} = \sum_{i=1}^{L} \alpha_i = 1$$

allows us to estimate the contribution of a_i (in ascending order) of the i-th component in the analyzed series..

3. Prediction based on Singular-Spectral Analysis (SSA)



The graphs of the first 13 components when predicted for 30 and 50 days

1. Input information for the CaterpillarSSA program is a time-series filtered by the wavelet filtering.

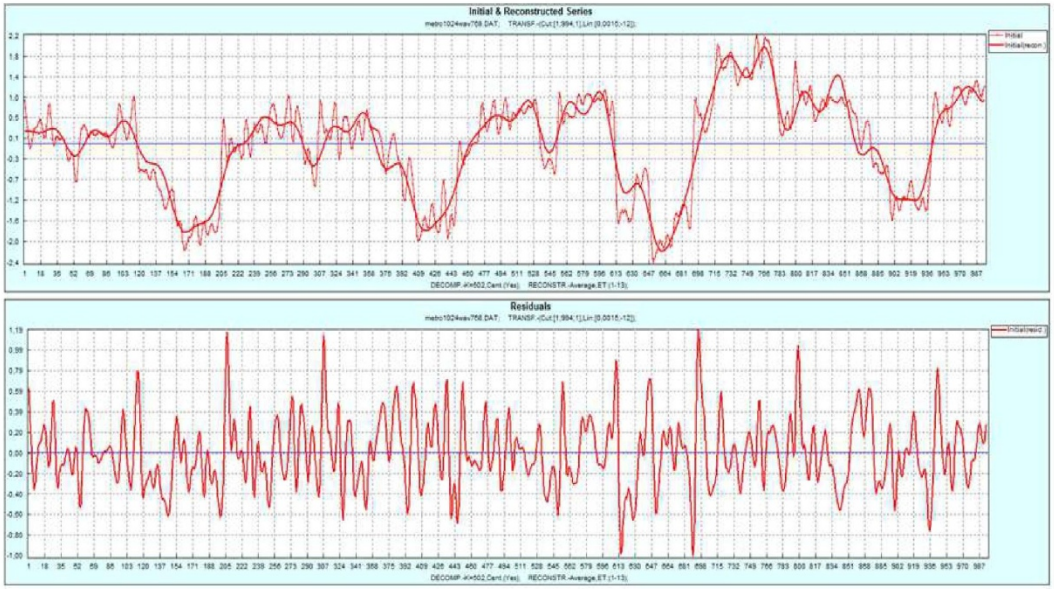
2. The time series analyzed was standardized by means of CaterpillarSSA.

3. With the transition to the multidimensional form, the length of the caterpillar was taken equal to 492, which is approximately half of the analyzed series and makes it possible to isolate all its characteristic features.

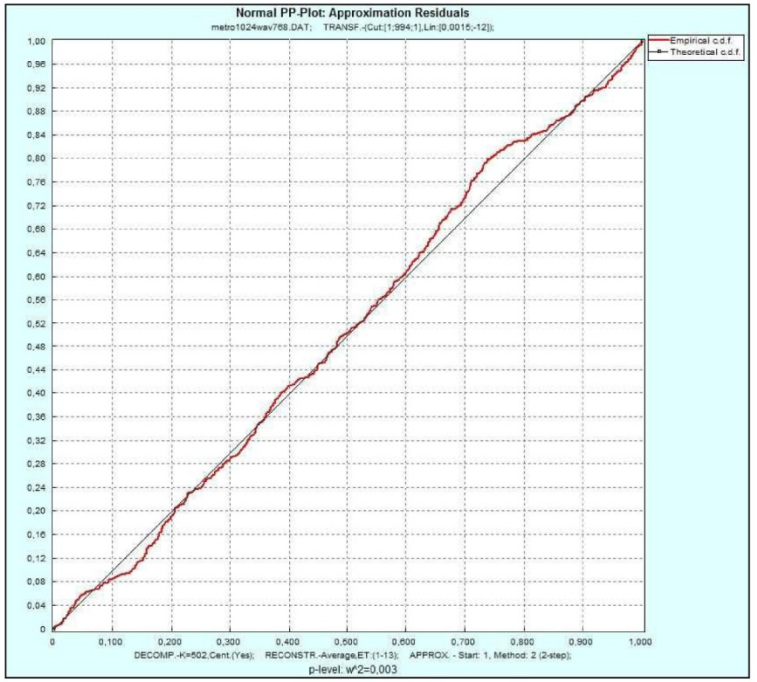
4. In the reconstruction of the one-dimensional series, 13 main components were used, their total contribution was 99.91%.

5. At the forecasting stage, the confidence interval was set to 0.25.

3. Prediction based on Singular-Spectral Analysis (SSA)



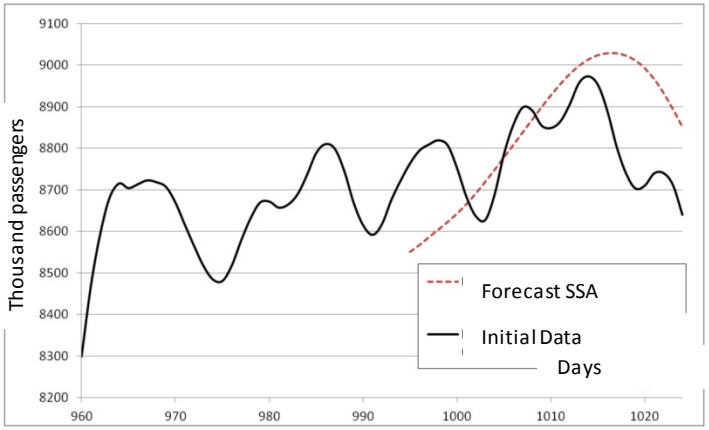
Above - the original series and its approximation.
At the bottom - a series corresponding to the discarded components

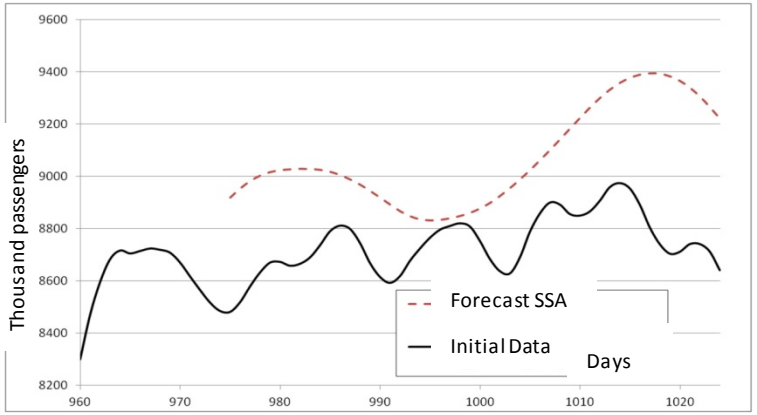


Normal pp-plot distribution corresponding to discarded components

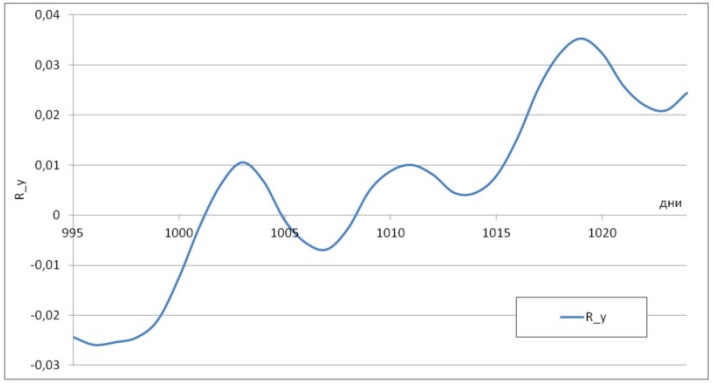
The discarded data are in good agreement with the normal distribution, from which it can be assumed that the process responding to them is close to Gaussian noise in its behavior.

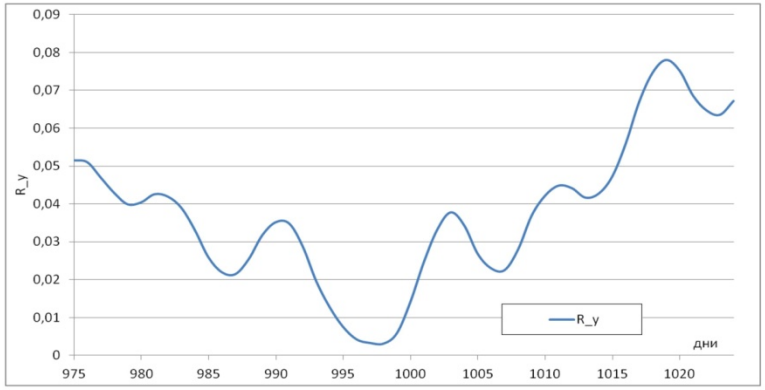
3. Prediction based on Singular-Spectral Analysis (SSA)





The results of the forecast in the "Caterpillar" -SSA approach at 30 (left) and 50 (right) observation days in comparison with actual data



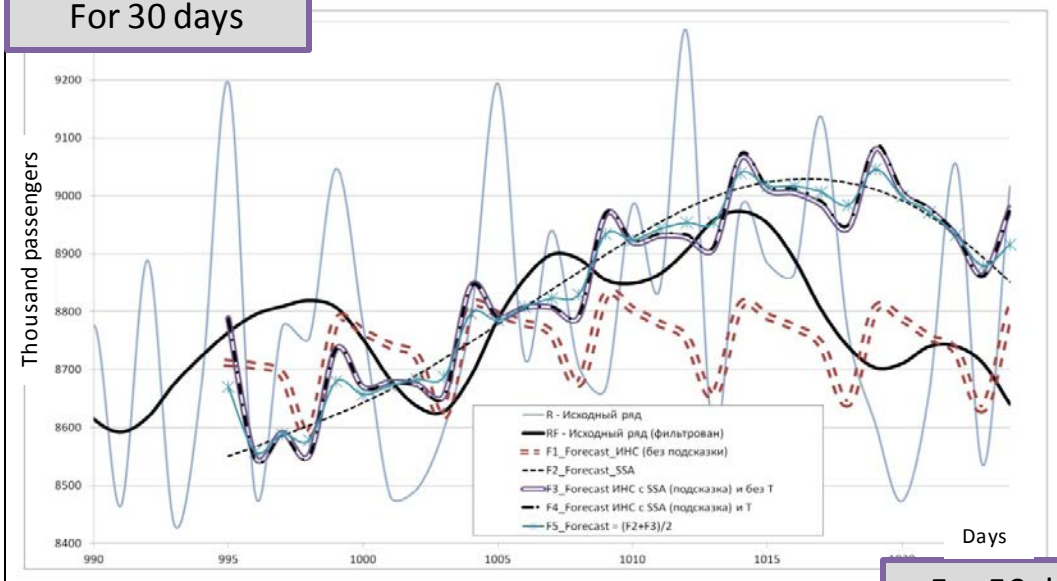


Dynamics of the Ry value depending on the ordinal number of the observation day when predicted based on the method implemented in the "Caterpillar" -SSA approach at 30 (left) and 50 (right) observations

The SSA method allows the medium-term forecast of the passenger traffic volumes in the Moscow metro with acceptable accuracy.

4. Comparison of results obtained by different methods

For 30 days



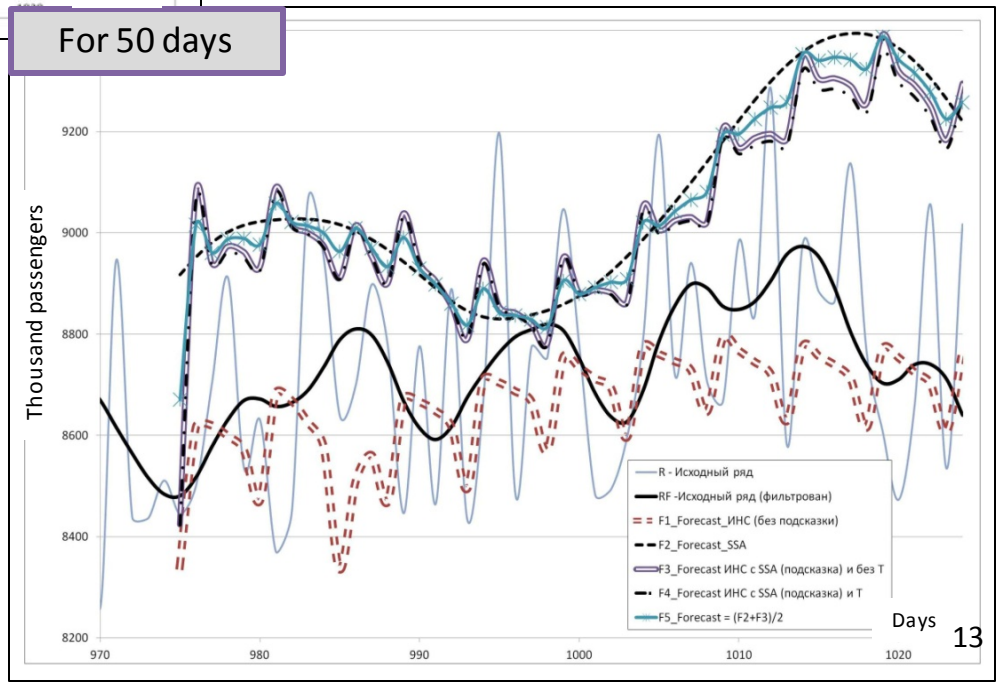
Graphs of comparative forecasting results for 30 and 50 days of observation, using the approaches developed in this work

After conducting a series of experiments, it was analyzed:

1. How does the quality of the forecast affect the exclusion of Var5 variables (deviation of the daily temperature from the norm) and Var6 (daily energy consumption).

2. How will the accuracy of the forecast change when using the ANN and SSA methods together (the SSA forecast serves as a "hint" for INS in the form of an additional neuron at the input of the network).

For 50 days



5. Results

1. The best variant of the forecast was achieved with the help of an INS rectilinear type - a multilayer perceptron, on the input of which a set of factors influencing the daily volumes of passenger traffic in Moscow Metro was submitted.

2. Our analysis has shown that one of the factors can be excluded from this set, namely, the deviation of the daily temperature from the average statistical norm. This factor has no significant effect on the results of the forecast.

3. The forecast based on the singular-spectral analysis implemented in the package "Caterpillar" - SSA, which used only the data of daily passenger traffic in the MM, turned out to be shifted relative to the forecasted data in the region of large passenger traffic figures. The same was the reason for the displacement of the forecast values in the region of large amounts of passenger traffic in the joint use of ANN and the "Caterpillar" -SSA approach.

4. The method of medium-term forecasting developed by us with the accuracy achieved in this work will help to increase the efficiency and speed of taking managerial decisions depending on the situation both in the metro and in its external environment, including climatic conditions and the dynamics of energy consumption as an indicator of industrial, business and social activity of the population of the region.

Thank you for your attention!