

Preparation of the forecasted fuel and energy balance of a boiler in an inadequate data condition

S V Guzhov¹, E G Gasho² and V A Shepel¹

¹Department of Heat and Mass Transfer Processes and Installations, National Research University MPEI, Krasnokazarmennaya St., 14, Moscow, 111250, Russia

²Department of Industrial Heat Energy Systems, National Research University MPEI, Krasnokazarmennaya St., 14, Moscow, 111250, Russia

E-mail: GuzhovSV@yandex.ru

Abstract. The task of fuel and energy balance preparation for heat energy sources in inadequate data conditions is an urgent issue, especially for the outdated boilers. It is difficult to use formulas for physically worn-out systems description, mainly because thermotechnical characteristics of the insulating components have significant deviations from the standard values. Boiler's equipment physical depreciation increases greenhouse gas emissions into the atmosphere. Therefore, dealing with modernization funding feasibility of the heat energy sources, and also for periodic adjustment works that help to reduce harmful substance emissions into the atmosphere, requires the use of a fairly complete mathematical model of a boiler and a heat network. However, the lack of data makes it very difficult to formulate a reliable model. Status analysis of every element in the system is a labor-intensive and expensive task. Another extremely labor-intensive issue is a determination of the forecasted thermal energy demand of the municipal consumers and sector of housing and public utilities. Physically worn-out buildings have low thermal resistance, which increases heat demand. Calculation of the necessary amount of heat generation by the worn-out boilers for the city districts with worn-out buildings is difficult due to lack of a wide range of data about the boiler's condition, network condition, and heated buildings condition. The paper presents the predictive energy balance preparation method using the statistical method of information processing only.

Keywords: modernization, flue gas, heat supply system, forecasted heat energy demand, mathematical modeling, regression analysis, multicollinearity, heat energy consumption forecasting.

1. Introduction

Implementation of the modernization policy for the outdated boilers requires a preliminary assessment of technical and economic effects from modernization, but this is not enough. The complexity of such calculations is that existing outdated boilers very often work in modes that are far from the designed values. Let's take a look at how closely the operation of these boilers on the fuel oil influence the current ecological situation and also effect from heat source transition to modern technologies. Additional immediacy of the problem is in the environmental impact reduction. Boiler's flue gas releases into the atmosphere large amount of hard and gaseous pollutants, such as fuel ash, oxides of carbon, sulfur, and nitrogen. Aside from that, a large amount of carbon dioxide and water vapor get into the atmosphere. The forecast with an engineering accuracy of technical and ecological effects from such heat



sources is a very difficult problem, especially in conditions with substantially inadequate and partially unreliable initial information.

Reliable calculation of the modernization feasibility will allow increasing heat supply quality, system reliability, heat energy tariff growth rate reduction, improvement of the existing ecological situation.

The objective of this paper is the calculation with given accuracy of the technical-and-economic effect from modernization of the existing obsolete and worn-out heat supply system of a city district.

2. Current ecological situation

Implementation of the modernization policy for the outdated boilers requires a preliminary assessment of technical and economic effects from modernization and not just that. Environmental impact reduction is the top priority. The complexity of such calculations is that the existing outdated boilers very often work in modes that are far from the designed values. Let's take a look at how closely the operation of these boilers on fuel oil influences the current ecological situation and also effect from heat source transition to modern technologies. Boiler's flue gas releases into the atmosphere large amount of hard and gaseous pollutants, such as fuel ash, oxides of carbon, sulfur, and nitrogen. Aside from that, a large amount of carbon dioxide and water vapor get into the atmosphere.

Sulfur dioxide or sulfurous anhydrite (sulfurous gas) leads to a tree dwarfism and leaf chlorosis (yellowing or discoloration). Humans suffer from upper air passages exasperation caused by this gas.

Nitrogen oxide (NO) and dioxide (NO₂) take part in photochemical smog formation in the surface layer of the atmosphere. Nitrogen oxide dissolution in the atmosphere vapors leads to the acid formation, which is the main cause of the "acid" rains, that are very harmful to flora and fauna. Humans develop susceptibility to acute and chronic respiratory diseases. In the upper atmosphere, nitrogen oxide triggers a lot of reactions, that make the ozone layer thinner, which can cause skin cancer in humans [1].

Solid waste, such as ash and slag, is partially responsible for environmental damage [2]. In spite of filters use, which can be in non-serviceable conditions on the worn-out equipment, every year several million tons of fine aerosol is being released into the atmosphere from the powerplant emissions, which are condensation nucleus for the water vapor and an additional cause of a precipitation formation. Its elevated concentration creates different respiratory diseases in humans.

Soot particles don't interact with the atmosphere's oxygen, but participate in a smog formation and promote skin cancer.

Carbon oxides (CO and CO₂ – carbon dioxide) practically don't interact with the other substances in the atmosphere and its lifetime is almost unlimited. Its overload in the air decreases the ability of the hemoglobin in red blood cells to carry oxygen, resulting in slow human responses, weak perception, headache appearance, drowsiness, nausea. High concentration can lead to fainting and even coma.

Product of the fuel oil combustion is a fuel oil ash, which consists of vanadium oxide (VO, V₂O₅), and also metal oxides (SiO₂, MgO, Al₂O₃, Fe₂O₃, Ni₂O₃, and other oxides). The most toxic agent - vanadium pentoxide was taken as a controlling factor. Air quality standards define the allowable limits of harmful substances content for the industrial area, and also for the city's residential area.

This paper's objective is the problem statement of the technical-and-economic effect evaluation of the outdated heat supply system modernization. Modernization will allow improvement of quality and reliability of the heat supply system, keep existing tariffs on heat energy, and also substantially improve the existing environmental situation. The main prerequisites for modernization:

- heavy wear out of the heat supply system, which is in use since 1970;
- high degree of physical wear out of the existing heat energy equipment (more than 70%);
- designed service life expiration and low operating efficiency of the main equipment;
- connection hourly heat load of the consumers is less than 20% of the designed equipment power, which has a negative impact on the effectiveness and safe operation of the boiler's equipment;

- heavy harmful substances emission into the atmosphere due to fuel oil use as a fuel.

3. The existing heat supply system description

Administrative center's square footage within a municipal district is 835.7 thousand ha (43% of the MD "Tersky District" territory), where lands of the inhabited localities – 1.47 thousand ha, including being studied residential location – 1.41 thousand ha. The number of the residential population as of January 1, 2018, is 4578 residents and has a gradual decline tendency (Table 1).

Table 1. Dynamic of the population change

Year	1970	1979	1989	2002	2009	2013	2016	2017	2018
Population, ppl.	7 268	7699	8 309	6 497	5 442	5 170	4 772	4 658	4578

The considered territory is in the subarctic area, because its climate is being formed by the influence of the White Sea and the continental part of the Kola Peninsula, and characterized as near sea climate. The warmest month – July, with a daily mean temperature of 14.3°C. Absolut maximum – +32°C. Absolut minimum – minus 40°C. The average duration of the frost-free season is 85-90 days. Territory climatic conditions fall into the construction climatic zone II-A. Air temperature of the coldest five-day period, with the reliability of 0.92, is assumed as minus 29°C, and air temperature with the reliability of 0.94 – minus 17°C. Heating season duration, on average, is 263 days. The average annual air temperature is 0.5°C. The coldest month is February, with a daily mean temperature of 11.9°C (Table 2).

Table 2. Monthly average temperature values

Month	Jan.	Feb.	March	April	May	Sept.	Oct.	Nov.	Dec.
Monthly average temperature, °C	-11.5	-11.9	-6.4	-1.5	4.3	7.5	1.4	-4.4	-8.3

Heat supply network length is 13.2 km (considering double lines). For the moment, the residential area has five boilers in use with the total power output of 31.5 MWt.

Authorized temperature charts of the heat sources and heat supply networks operation have the following values – 95/70°C and 130/70°C. However, mathematical modeling has shown that the boiler's actual temperature charts differ from the authorized temperature chart of the heat supply network 95/70°C (Figure 1) due to several reasons, including the following:

- fuel economy;
- break up possibility of the equipment of the heat supply source and networks, due to substantial wear;
- manual control of the heat energy production process.

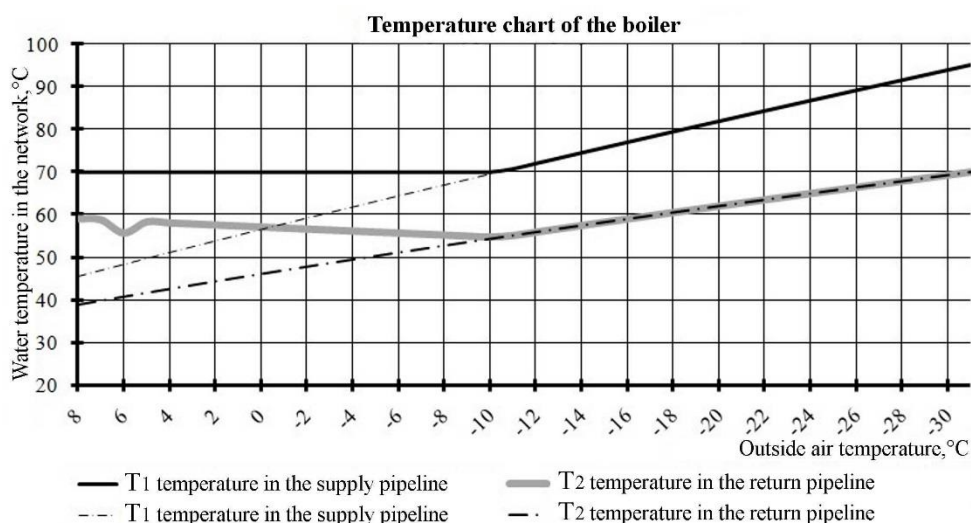


Figure 1. Calculated actual chart of heat energy delivery by the boiler

Heat supply system – closed double pipeline. Boiler's outlet pipes – D219 mm (delivery), D219 mm (return) provide heating load and DHW 2.011 Gcal/hr. Pipelines are under and above ground. Available head is 24 m, return pipeline pressure – 30 m. Thermal power flow rate for auxiliary needs 0.599 Gcal/hr. Thermal power losses in the networks 1.593 Gcal/hr. During the 2018/19 heating season load received 22.956 thousand Gcal. The fuel oil was used as fuel (Table 3). The estimated amount of emissions [3] into the atmosphere is around 5.8 million tons per year.

Table 3. Schedule of the fuel oil supply to the boiler in the 2018-19 heating season, MT/mo

July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Marc.	Apr.	May	June
34	46.5	78	125	146	172	191	170	158	140	129	64

The examination discovered the following drawbacks of the technology:

1. Fuel oil unload requires the use of steam, which accompanied by the emission of the fuel oil light fractions into the atmosphere. Substantial fuel oil quality decrease creates a smell problem, and, therefore, light hydrocarbon emissions during fuel oil unloading. When fuel oil is being heated using direct steam, the water content in the fuel oil rises sharply, which leads to the loss of steam and condensate, but also fuel oil quality degrades. As a result, the boiler's efficiency and performance reliability decrease. Water in the fuel oil, during heating, induces foaming.

2. The boiler has an inefficient flue gas cleaning system. Several high toxic components are being released into the city's atmosphere – benzopyrene, fuel oil ash, nitrogen oxides. The gas cleaning system is not an option due to the highly corrosive activity of the exhaust gas.

3. The fuel oil with a sulfur content of more than 2% is being used. The fuel oil contains sulfur in active and passive forms. Active sulfur causes corrosion in pipes, heating units installed in tanks, heat-exchange units and back-end surfaces at the temperature of a steel wall lower than the dew point.

Complexity of the stated problem is in the following aspects:

I. impossibility of using formula-based models for forecast, because the 3D model of the heat supply network was composed at the point of technical-and-economic feasibility study [4];

II. lack of the reliable information about wear out degree of each element in the main and supporting equipment of heat supply source, main pipelines, CHP units and consumers' network [5-7];

III. characteristics of the heated building enclosing structures and degree of wear out are unknown, which doesn't allow to estimate the demand in different climatic conditions with acceptable engineering precision [8, 9];

IV. impossible to rely fully on the source's heat delivery alone, due to the increased self-consumption and losses, and also due to the boiler's manual control errors [10, 11];

V. complexity of the heat supply source analysis, because its operational schedule of heat energy delivery differs from the recommended schedule [12], etc.

The described above factors undoubtedly disallow using formula-based models. Therefore, it is necessary to accept the considered boiler system as a «black box» and examine it using statistical methods of information analysis.

4. Selection of the data analysis method for construction of the heat consumption regression model

Let's examine several main methods and models, which at present are being used in a statistical approach for loads planning in heat and power systems:

- extrapolation method – statistical method, which uses the trend identification and description for forecasting;
- correlation-and-regression model – statistical model, which conducts the forecast using the mathematical description of the existing dependencies between variables;
- neural network model – nonlinear model, forecast based on it results in the model of heat energy consumption dependence on external factors, and also on retrospective data [13].

Energy consumption forecast of the considered city is done by the correlation-and-regression analysis in connection with the following aspects:

- a) method is based on the consideration of the correlation (statistical) relationships between loads on separate parts, caused by the presence of several factors [14];
- b) inadequacy of the extrapolation method of boiler's operation modes due to lack of the annual and monthly similarity of climatic factors [15];
- c) forecast accuracy far exceeds the forecast accuracy of the other models;
- d) neural network model-based forecast is a complex problem that requires BigData analysis, which is absent in this case.

The chosen correlation and regression analysis are based on the test of the hypothesis of the extent of the correlation ratio between input and output parameters. For the checking purpose most frequently used Pearson's linear correlation coefficient or Spearman's and Kendall's rank correlation coefficients. Pearson's linear coefficient – a measure of the linear correlation ratio between two quantitative attributes x and y . It is based on assumption that when characteristics of x and y completely independent, value deviations of factorial feature from the average $(\bar{x}-x_i)$ have random nature and should randomly match with different deviations $(\bar{y}-y_i)$. In the case of the substantial advantage of matching or mismatching of these deviations, it is assumed that there is a relationship between x and y .

Rank correlation coefficients are based not on the correlation of the correlated characteristic values itself but on their ranks – numerical orders, which are assigned separately to each individual value x and y in the ranked series. Calculation formula of Kendall's rank correlation coefficient (τ) for connected ranks:

$$\tau = \frac{S}{\sqrt{\left(\frac{n \cdot (n-1)}{2} - U_x\right) \cdot \left(\frac{n \cdot (n-1)}{2} - U_y\right)}} \quad (1)$$

where S – actual total score sum when estimator +1 of each pairs of ranks with the same change order and when estimator -1 of each pairs of ranks with the opposite change order;

$U_x = U_y = 0.5 \cdot \sum[t \cdot (t-1)]$ – number of scores, which correct (decrease) maximum score sum at the expense of repetitions (grouping) of t ranks in each series.

Using the examined coefficients, it is possible to prove the presence of connection, establish its direction and relationship degree, but impossible to determine the significance of the being examined factor among the other factors.

It is worthwhile to evaluate the degree of conformity of the forecast model and the object's actual behavior by using the determination coefficient. When $R^2=1$, it is possible to conclude that the func-

tional relationship between input and output variables exist, when $R^2=0$ – that it is non-existent. If $0 < R^2 < 1$, then the closer R^2 to 1, the better quality of the function approximation with initial data. Value R^2 increases when one more regressor is added, therefore, when choosing between several regression equations one shouldn't rely on determination coefficient R^2 alone. To eliminate this, the following formula can be used:

$$R^2 = 1 - \frac{ESS/(N-k-1)}{NSS/(N-1)}, \quad (2)$$

where numerator – unbiased estimator of errors dispersion, denominator – unbiased estimator of dispersion Y; TSS (Total Sum of Squares) - all dispersion or variation Y, which characterizes the degree of random value spread of regression function around mean value Y; ESS (Error Sum of Squares) - sum of squares of regression residuals.

The singularity of the suggested method is in the introduction of the «idealized heat consumption» factor into an initial data array, which is determined by using classical formula-based mathematical methods. This step allows both increase accuracy of the statistical calculation and clearly determine deviations in the object's actual operational mode from the recommended.

5. Heat energy consumption forecast

As a test set for model construction was used heat energy consumption data for every 3 hours during the period from 01.06.2018 to 31.05.2019. The being composed regression model doesn't consider the summer period when boilers deliver heat in the DHW mode. Goal seek and best model selection were done in the EXCEL environment. The process of actual heat consumption forecast includes the following stages:

- analysis of the initial information and filtering off of the unreliable data such as single unexplainable discharge or drop of heat consumption; dimension errors; lack of metering and monitoring readings, etc.;
- determination of the actual and «idealized» production of heat energy for the heat supply system under study during the selected period of time;
- determination of the set of factors, that is, based on the expert's opinion, having an impact on heat consumption values and filtering off some of them using collinearity and multicollinearity criteria;
- formulation of the forecast mathematical function or group of predictive rules;
- determination of the values of all input factors at the forecast time period, or range of their values using principle: «optimistic», «neutral», «pessimistic» forecasts;
- determination of the forecast heat consumption values considering all selected factors and periods.

As a factorial features were taken: outside air temperature (X_1); atmospheric pressure at the station level (X_2); relative humidity at the altitude of 2m, % (X_3); wind direction, rhumb (X_4); wind speed, m/s (X_5); total cloud cover (X_6); sum of all clouds in view (X_7); dew point temperature at the altitude of 2m (X_8); accumulated precipitation quantity, mm (X_9). Total heat consumption is taken as an effective feature.

Formula based calculation of the «idealized» heat energy delivery can be done using formula:

$$Q_{\text{form}} = G_{\text{suppl}} \cdot C_p \cdot (t_{\text{suppl}} - t_{\text{ret}}) \quad (3)$$

where $C_p=1,001\text{kcal/kg}\cdot\text{R}$, G_{suppl} - heat transfer medium consumption in the supply pipeline kg/hr; t_{suppl} and t_{ret} – heat transfer medium temperature in the supply and return pipeline.

Calculation of the paired correlation coefficients can be done by the Student's t-criterion:

$$t_{obs} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}. \quad (4)$$

where r – paired correlation coefficient, calculation based on the selected for research data, n - number of observations.

To find t_{criter} using Student's table the value of significance level was taken as $\alpha=5\%$ (confidence probability 95). To test on statistical significance, let's compare $t_{criter} = -1.646$ and paired correlation coefficients, calculated by EXCEL. Factors X_6, X_9 – statistically are not significant and can be eliminated from further analysis.

Test of the input factors multicollinearity assumption using the Farrar-Glauber method on first type statistical criteria:

$$FG = - \left[n - 1 - \frac{2k+5}{6} \right] \ln(\det[R]), \quad (5)$$

where $n=1976$ – number of observations; $k=9$ – number of factors.

At $v=k/2(k-1)=1$ degrees of freedom and significance level $\alpha=0.05$, the number of degrees of freedom $d=9 \cdot (9-1)/2=36$. Tabulated criterion value equals to 50.998.

Calculated value:

$$FG = - \left[1976 - 1 - \frac{18+5}{6} \right] \ln(0.00148623) = 6975.$$

Considering that the calculated value is bigger than the tabulated then with the assumed reliability one can say that in the array of factors exists multicollinearity with the probability of 95%.

Let's identify a multicollinearity of initial factors using statistical criteria of the second type (Fisher's criterion):

$$F_k = (d_{kk} - 1) \frac{n-k}{k-1}, \quad (6)$$

where d_{kk} -diagonal elements of the matrix $n=1976$ – number of observations; $k=9$ – number of factors.

At $v_1=n-k$ and $v_2=k-1$ degrees of freedom and significance level α : $d_1=1976-9=1967$; $d_2=9-1=8$; then F_{table} corresponds to value 2.071.

If $F_{calcul} < F_{table}$ then there is no reason to decline the hypothesis of multicollinearity absence. Based on the conducted analysis a relationship between factors X_1, X_8 was identified and followed by the elimination of X_8 from the calculation.

Test of the variables on multicollinearity on statistical criteria of the third type (Student's criterion) demonstrated the increase of residual dispersion from 0.70 for $Q=F(X_1)$ to $R^2=0.762$ $Q=F(X_1; X_4; X_5; X_2; X_3; X_7)$. To test the validity of the factor's selection, let's introduce into calculation factors X_6, X_8, X_9 that have been excluded before. Considering that the determination coefficient value will not change $R^2=0.762$, then predictors selection was adequate.

Carry out of the linear regression coefficients calculation determines the type of the function:

$$Q = 3.11 - 0.0936 \cdot X_1 - 0.0010 \cdot X_4 + 0.0203 \cdot X_5 - 0.0010 \cdot X_2 + 0.0042 \cdot X_3 + 0.0038 \cdot X_7.$$

Based on the obtained mathematical dependence and calculated the actual schedule of heat energy delivery the data about heat energy delivery for the next heating season was obtained. It was assumed that the climatic factors for future and present calendar years are similar.

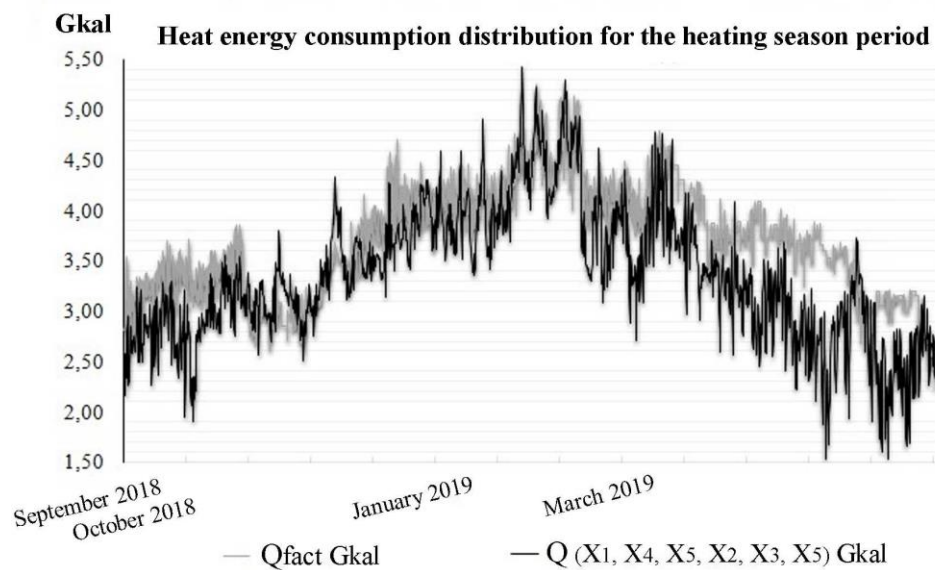


Figure 2. Forecasted value of heat energy consumption by boiler's load during heating season with sampling period 3 hr.

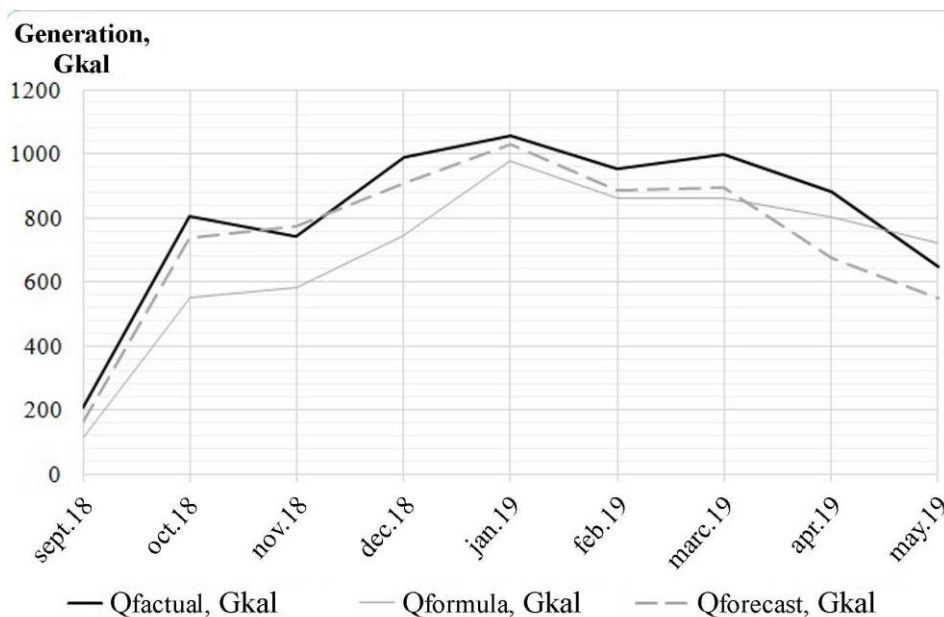


Figure 3. Forecasted value of heat energy consumption by boiler's load during heating season with monthly sampling period

A decrease in the forecasted delivery of heat energy was observed during the whole heating season. September and November are the exceptions because the heat source intentionally delivers more heat to warm up buildings.

6. Conclusion

The total decrease in the forecast with confidence probability 95% and significance level 5% is 529 Gcal, which is 13.3% and equivalent to the emissions decrease by 0.77 million tons per year. This effect can be achieved by the exclusive use of the heat source operational modes optimization.

The paper presents the construction method of the forecast function of heat consumption by the heat supply system in a substantial lack of information about the network's structure and condition. Affecting factors set justification has been conducted. The formulation of the forecast mathematical function of heat consumption by the city's district has been conducted.

It has been shown that formula-based models, which can be used with the existing level of data, deliver big differences from the actual and cannot be used.

The results of the process analysis of different factors influence on the heat energy consumption amount for the object of research have been presented. It has been discovered, during research, that the amount of heat consumption decreased in comparison to the previous period due to several parametric factors, which confirms the necessity of both additional modes tuning up and heat source modernization. The presented mathematical methods allow us to determine the degree of climatic factors influence on the total heat consumption amount. It should be noted, that discovered dependence has strictly individual character, which is unique to the system under analysis only. Test for collinearity of different system with similar characteristics, and as a result, the general form of the function $Q=f(X_i)$, can be different from the obtained in the paper.

The results of the model calculation can also be used for calculation of the expected productive delivery during the next periods, economical estimation of the worn-out heat supply system further modernization, pricing policy formulation, estimation of the budget and borrowed fund investment efficiency, etc.

Acknowledgments

This paper was prepared using the results of the project, which has been conducted with the support of the Russian Science Fund (unique identifier of the grant RSF № 16-19-20568).

References

- [1] Gao J, Kovats S, Vardoulakis S, Wilkinson P and Liu Q 2018 Public health co-benefits of greenhousegas emissions reduction: A systematic review *Science of The Total Environment* **627** 388-402
- [2] Cho J, Kim J, Lee S and Koo J 2016 A bi-directional systematic design approach to energy optimization for energy-efficient buildings *Energy and Buildings* **120** 135-44
- [3] Zheng X, Streimikiene D, Balezentis T, Mardani A and Liao H 2019 A review of greenhousegasemission profiles, dynamics, and climate change mitigation efforts across the key climate change players *J. of Cleaner Production* **234** 1113-33
- [4] Huang X and Wang Y 2019 Investigating the effects of 3D urban morphology on the surface urban heat island effect in urban functional zones by using high-resolution remote sensing data: A case study of Wuhan, Central China *ISPRS J. of Photogrammetry and Remote Sensing* **152** 119-31
- [5] Gasho E G, Guzhov S V and Krolin A A 2019 On the forecasting of trends of specific mixed loads of electric consumers of microdistricts and cities change in the conditions of inadequate data *IOP Conf. Ser.: Mater. Sci. Eng.* **552** 012013
- [6] Larsen S, Hongn M, Castro N and González S 2019 Comparison of four in-situ methods for the determination of wallsthermal resistance in free-running buildings with alternating heat flux in different seasons *Construction and Building Materials* **224** 455-73
- [7] Rasooli A and Itard L 2019 In-situ rapid determination of walls' thermal conductivity, volumetric heat capacity, and thermal resistance, using response factors *Applied Energy* **253** 113539
- [8] Deconinck A and Roels S 2015 A Maximum Likelihood Estimation of the Thermal Resistance of a Cavity Wall from On-site Measurements *Energy Procedia* **78** 3276-81
- [9] Rasooli A, Itard L and Ferreira C 2016. A response factor-based method for the rapid in-situ determination of wall'sthermal resistance in existing buildings *Energy and Buildings* **119** 51-61

- [10] Al-Sanea S and Zedan M 2012 Effect of thermal bridges on transmission loads and thermal resistance of building walls under dynamic conditions *Applied Energy* **98** 584-93
- [11] Ostrowski P, Szelejowski F and Zymelka P 2019 Assessment of improvement in the energy and exergy efficiency of the gas heat plant after the exhaust gas cooled down below the dew point and the use of recovered heat *Energy* 116179
- [12] Brögger M, Bacher P and Wittchen K 2019 A hybrid modelling method for improving estimates of the average energy-saving potential of a building stock *Energy and Buildings* **199** 287-96
- [13] Huang S, Tao B, Li J and Yin Z 2019 On-line heat flux estimation of a nonlinear heat conduction system with complex geometry using a sequential inverse method and artificial neural network *Int. J. of Heat and Mass Transfer* **143** 118491
- [14] Nouvel R, Mastrucci A, Leopold U, Baume O and Eicker U 2015 Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support *Energy and Buildings* **107** 204-12
- [15] Amasyali K and El-Gohary N 2018 A review of data-driven building energy consumption *Renewable and Sustainable Energy Reviews* **81(1)** 1192-205