Nonlinear Phenomena in Complex Systems, vol. **26**, no. 2 (2023), pp. 166 - 173

On the Recovery of Lost Data During the Formation of Digital Twins of Heat Supply Systems

Alexei Sednin*
Research, Development and Innovation Centre for Automation and Control Systems in Heat and Power Engineering and Industry, Belarusian National Technical University, 65 Nezalezhnasti Ave., 220013 Minsk, BELARUS

Anatoly Zherelo†
Center of Information Technologies, Belarusian State University, 4 Nezalezhnasti Ave., 220030 Minsk, BELARUS
(Received 27 March, 2023)

This paper proposes general approaches to the process of collecting information with a focus on use of BigData and IoT technologies. As a basis for the proposed approaches, the implemented heat supply management systems developed by the Scientific Research Center for ACS TEP of the Belarusian National Technical University are used. Further directions of research of the developed systems are indicated.

PACS numbers: 88.05.Sv, 88.40.fc
Keywords: district heating systems, data mining, machine learning
DOI: https://doi.org/10.33581/1561-4085-2023-26-2-166-173

Introduction

The global trend of universal digitalization requires the need for a wide usage of information technologies in district heating systems. Nowadays, intelligent process control systems are increasingly being used [1–4]. The features of such systems are the use of multi-parameter industrial computers and controllers that provide process control based on big data analysis Numerous sensors and devices that control technological processes serve as a source of data [5].

The significantly increased volume of collected data leads to the need to form new requirements for data processing, namely, the tasks of validation and forecasting, require the creation of a fundamentally new platform that ensures the harmonization of requirements for all participants in the processes of production, transport and energy consumption. The main goal of collecting and processing data is to form a digital twin of the object of study, since it is the formation of a twin that will subsequently allow us to proceed to the construction of detailed mathematical models, and, as a result, improve the controllability and predictability of the behavior of the system as a whole.

The requirement for the completeness of the information being used is a natural requirement for a model building, since it is impossible to build an adequate model when completeness is unfulfilled. In practice, gaps in the collected data occur quite often. As a result, the problem of restoring gaps in the collected data is relevant. This paper proposes an approach to data recovery for the case when the data has periodicity.

1. Data sources

Currently, a project is being implemented in Minsk to create an automated process control system for a heat supply system [6, 7]. The main goals of the project are:

- ensuring centralized functional-group

*E-mail: sednin@bntu.by
†E-mail: zherelo@bsu.by

control of thermal and hydraulic regimes of heat sources, main heating networks and pumping stations, taking into account daily and seasonal changes in circulation costs with feedback according to actual thermal and hydraulic regimes in distribution heating networks of the city;

- creation of an effective system for protecting equipment of heat sources and heating networks in emergency situations;
- ensuring the collection and archiving of data on the thermal and hydraulic modes of operation of heat sources, main heating networks, pumping stations and distribution heating networks of the city for monitoring, operational management and analysis of the effectiveness of their functioning;
- creation of an information base for solving optimization problems that arise during the operation and modernization of district heating facilities, including the implementation of the method of dynamic central control of heat supply with optimization of coolant temperatures in the supply and return pipelines of heating mains, forecasting the heat load using neural network programming and other intellectual algorithms, etc.

The consistent development of this system will lead to the construction of a full-scale corporate computer network that combines the computing and information resources of the central office and structural divisions of the enterprise, and will also provide access to information from all interested city services and higher authorities.

The development and implementation of systems for operational dispatching and control is a necessary condition for more efficient management of the operation of heat networks and heat supply systems, and a better supply of consumers with thermal energy.

The task of introducing a data collection and analysis system, including the use of automatic control systems, is to achieve the energy efficiency of heat networks by reducing heat losses in the production and transport of heat energy.

The information obtained during the collection should be accumulated in a single system and then used to solve the following tasks [1, 8]:

- accumulation of information on the state and modes of operation of heating network facilities;
- construction of mathematical models of operating modes of heat generating equipment, conversion processes, transport and consumption of thermal energy;
- forecasting the possibility of creating emergency situations, assessing the risks of their occurrence;
- optimization of operating modes of technological equipment;
- full-scale accounting of production and consumption of thermal energy.

2. Restoring of the lost data

The widespread introduction of smart sensors, energy metering and control devices, the accumulation of archived data in control systems and the requirement to take into account many heterogeneous indicators in the synthesis of predictive models to support decisions have led to problems associated with the processing of large arrays of unstructured data [9–11].

The widespread introduction of automated process control systems makes it possible to ensure uninterrupted data collection from technological facilities, resulting in a huge information base for analytics and forecasting [12].

Usually, to recover missing data, the relationship between the readings of various types of sensors can be used. However, situations are possible when, for a number of reasons, omissions cover all the collected parameters at once and, accordingly, it is impossible to use previously restored dependencies. In what follows, we consider just such a case.

Let us further consider the approach to data recovery on the data obtained as a result of the operation of the automated control system of the central heating station (CHS).

On figure 1 shows the main frame of the automated control system, which is designed for operational control and management of the technological equipment of the central heating station. The main functions of the control system are to regulate the temperature in the hot water circuit according to a given daily and weekly schedule; maintaining the required pressure in the water supply circuit (if any); temperature and pressure control in the heating circuit depending on the outdoor temperature; control of boundary values of technological parameters, implementation of necessary technological interlocks. In addition to the control functions of the automated control system, the central heating substation provides the connection of the required number of cold and hot water and electric energy meters in order to organize commercial or technical metering of energy resources.

The initial data for some processed parameter is a sequence of vectors of the form (t_i, x_i, f_i), where t_i is the time of measurement, x_i is a parameter value and f_i is a some service flag for the current measurement. In the considered case, the flag value f_i equal to 5 is significant for us, since this value indicates a break in the data transfer, and as a result, these cases must be filtered out.

After the initial study of the data, it was found that 2 main types of gaps in the data can be distinguished:

- Short-term, as a rule, it is no more than one or two unidentifiable indicator values in a row where flag value $f_i = 5$.
- Long-term, usually long gaps in the observed data series without any data at all.

The missing information is either erroneous measurement points (marked as an error by the smart energy meter, or containing NaN values, or out of realistic range for this DH system) or missing measurements (no measurement has been logged for more than an hour, which creates a gap in the time series) [13, 14].

The initial data (on the example of the daily change in the consumption of cold water in the hot water supply system) are presented on the figure 2. Here the axis t is the time of measurement (in seconds from 1970-01-01 00:00:00), and the axis $value$ is the value indicator.

Since the observed process has a certain degree of continuity in terms of its physical properties, and also has a pronounced periodicity, we will use the Fourier transform to restore gaps.

At the first stage, since the measurements in time are carried out unevenly and the amount of data obtained is quite large, for convenience we will obtain a time series with a uniform time step. In our case, the time step was taken equal to 5 min. The step is chosen in order to reduce the amount of data being processed, and its value should be determined by the researcher on the basis of experimental data or on the basis of additional information about the speed of ongoing processes (for example, based on mathematical modeling of the process). As a recommendation, when choosing a time step, it should be indicated that its value should be no more than half the time interval during which the fastest process of transition from one system operation mode to another can occur in the system under study, for example, from the normal operation mode to the critical one. some node or system as a whole (see Kotelnikov's theorem).

This data transformation solves the problem of short-term gaps. Due to a sufficiently large amount of initial data and the continuity of the process, gaps can be replaced either by the result

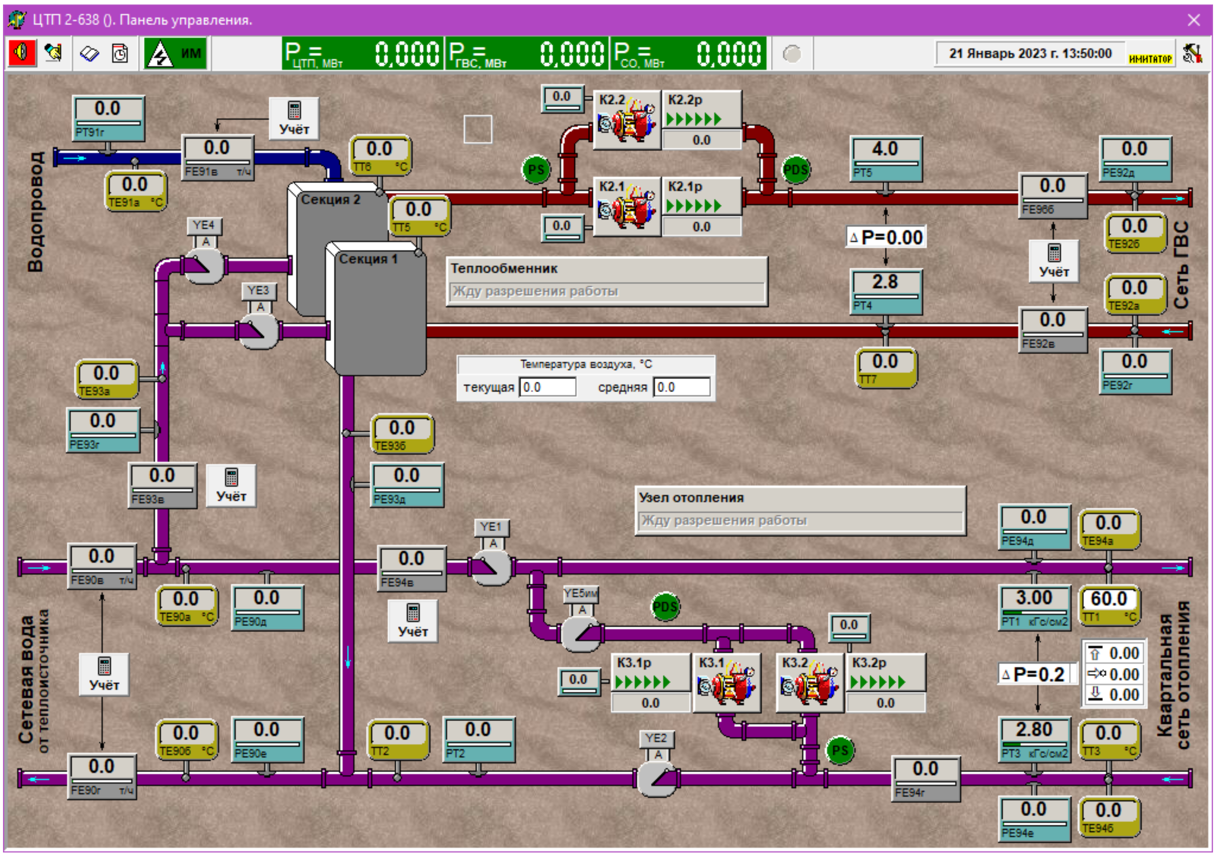

FIG. 1: (color online) The main screen of the automated control system of the central heating station.

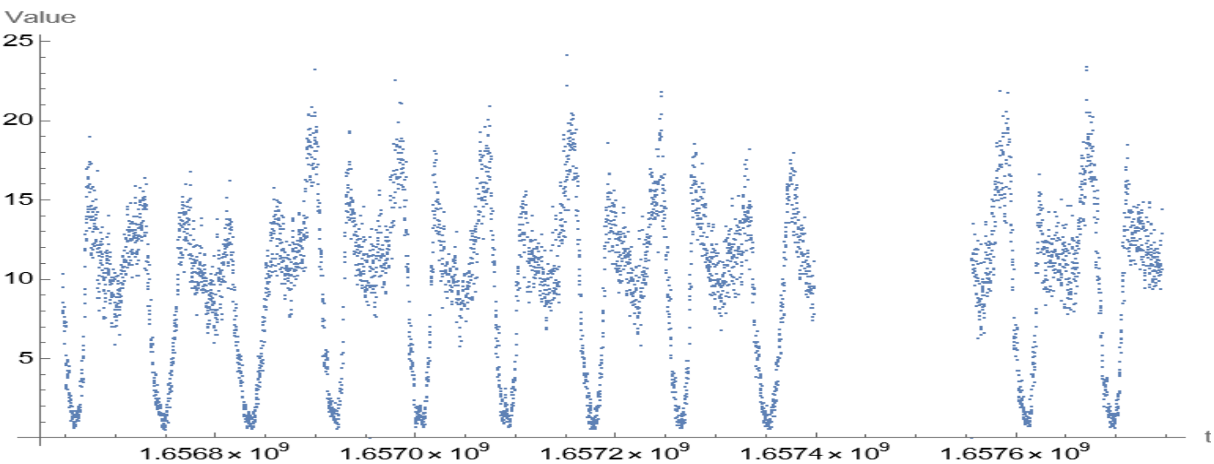

FIG. 2: The consumption of cold water in the hot water supply system.

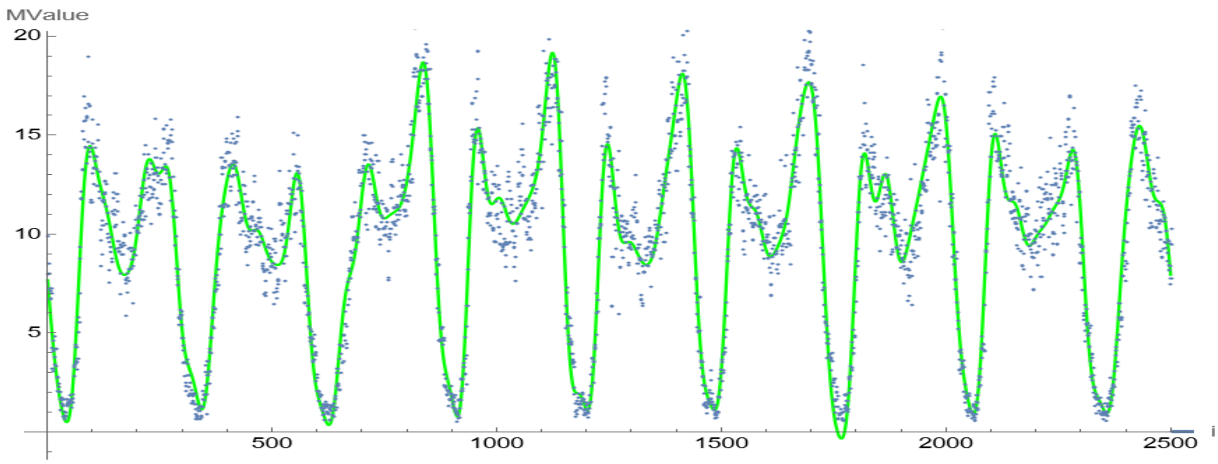

FIG. 3: The result of smoothing by the Fourier transform.

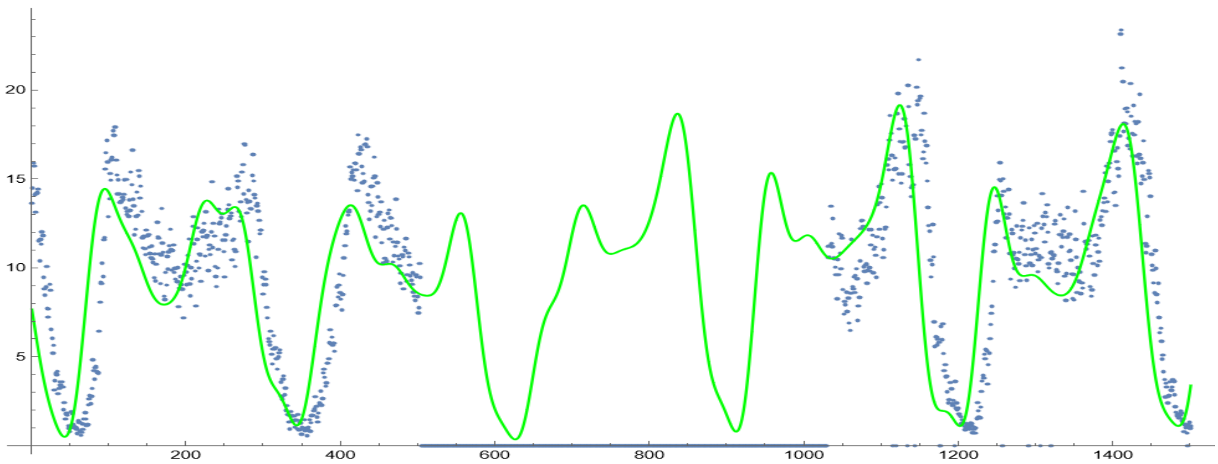

FIG. 4: The result of algorithm of data restoring (threshold for the sum of used coefficients is 50%).

of averaging other measurements that fall into the same interval, or by a linear approximation of the value built on the basis of values lying on opposite sides of the gap.

Further, to isolate the periodic structure, we apply the Fourier transform for that part of the time series in which gaps are absent or have already been restored by this moment, and in the resulting spectrum we leave only selected values. Further, we will assume that N is the number of elements of the series used in the Fourier transform, and v_k are the values of the observations. In the case under consideration, we sorted the obtained values of the coefficients in descending order and took into account only those coefficients that are large in absolute value, the total contribution of which to the sum of the absolute values of the spectrum was equal to approximately half of the total contribution. In this case, we need to strike a balance between the speed of calculations and the noise filtered out in this way. The issue of establishing such a limit should be studied separately. After filtering, we restore the values

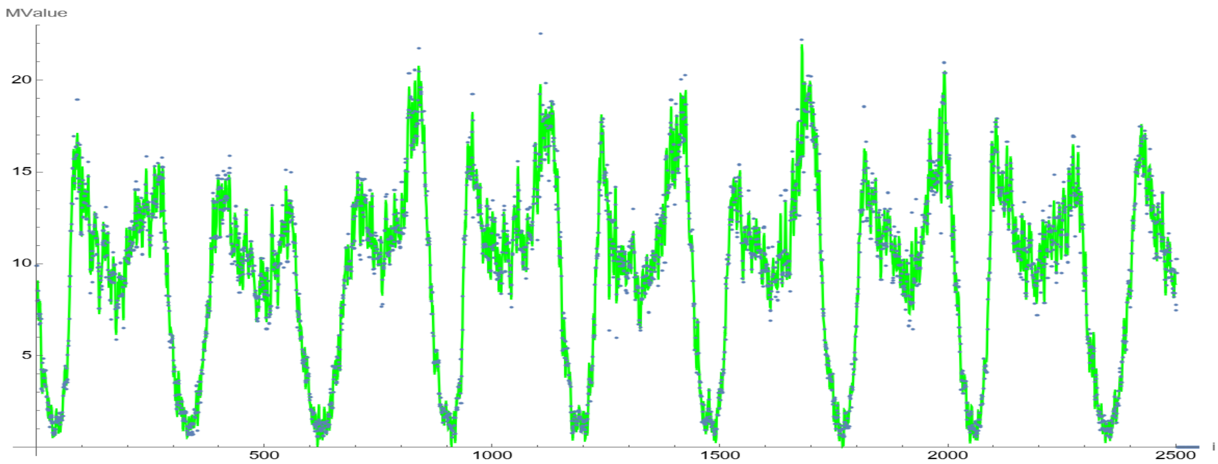

FIG. 5: The smoothed data.

using the inverse Fourier transform. The result of transformations is represented at the figure 3

The figure 5 shows the approximation for which we takes into account above 74% of the total sum of the absolute values of the coefficients. In the figure, the experimental data are marked in blue, and the result of smoothing is shown in green. The i axis denotes the interval number, the $MValue$ axis denotes the value used for the calculation (built at the preliminary stage).

The resulting smoothed signal will be used to restore data on the skip interval.

The main problem here is the fact that, over a significant time interval, there are no measurements at all, at least single ones. To restore the data, we will use the data for the analyzed period, but not directly, since, due to the random nature, we can include outliers in the approximating values, but the corresponding constructed approximation. To automate this task, we again use the assumption of the continuity of the process and form a window of observations with a gap. When forming the window, it is recommended to add to it time periods before and after the skip, the value of which is not less than the period of the slowest selected harmonic. To determine the shift position of the predicted window, we will solve the following minimization problem

$$\min_{i=1,N-S} \sum_{k=i}^{S} |(W_{k-i+1} - v_k)\mathrm{H}(W_{k-i+1})|$$

where S is a window size, $W_k, k = \overline{1, S}$ are values from the observation window used, and

$$\mathrm{H}(x) = \begin{cases} 0, x = 0, \\ 1, x \neq 0. \end{cases}$$

In this case, to reduce the influence of possible outliers, we used the modulus function, but in the general case, you can use any other norm, at your discretion.

The figure 4 shows the result of the algorithm for the case when the terms of the expansion with a specific share of 50% of the total sum of the absolute values of the coefficients are taken into account Green indicates the approximation, blue indicates the experimental data.

The figure 6 shows the result of the algorithm for the case when the terms of the expansion with a specific share of 74% of the total sum of the absolute values of the coefficients are taken into account.

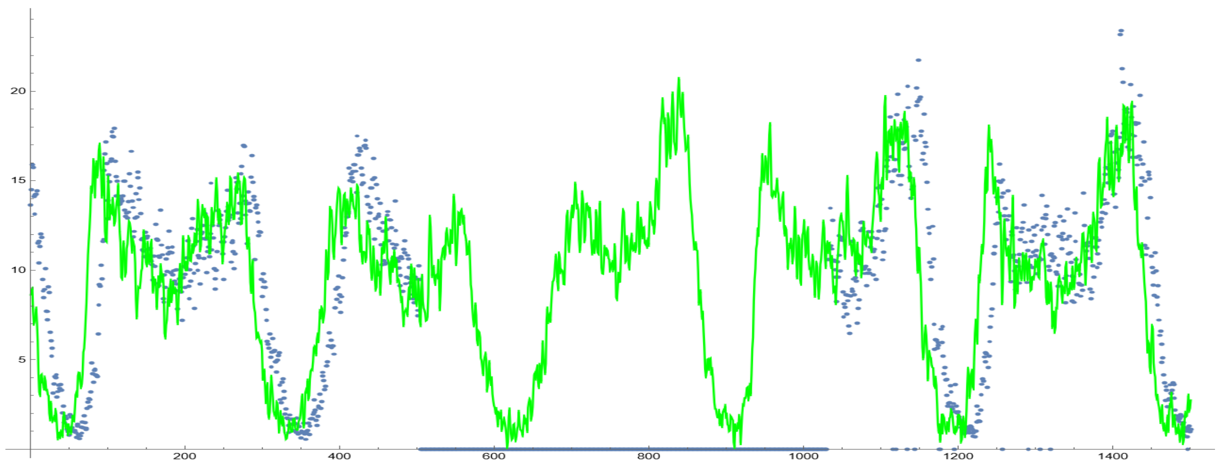

FIG. 6: The result of algorithm of data restoring (threshold for the sum of used coefficients is 74%).

3. Conclusion

The main purpose of this work was to pay an attention not to the selected aspects of data collection or processing, but to formation of principles for constructing a data collection system focused on their subsequent processing, taking into account existing technologies. The proposed approaches can be used in formation of technical requirements for industry-specific systems focused on the use of cloud technologies.

Acknowledgement

This work has been partially carried out within the framework of the joint scientific project of the Belarusian Republican Foundation for Fundamental Research and the Ministry of Innovative Development of the Republic of Uzbekistan, BRFFR MIRRU-2022 (Project T22UZB-052).

References

[1] N.N. Novitskii, Z.I. Shalaginova, A.V. Alekseev, O.A. Grebneva, V.V. Tokarev, A.V. Lutsenko, O.V. Vanteeva. Intellectualization of heat-supply systems: Current state, trends and tasks (A review). Thermal Engineering. **69(5)**, 367 (2022).

[2] N.N. Novitsky, Z.I. Shalaginova, A.A. Alekseev, V.V. Tokarev, O.A. Grebneva, A.V. Lutsenko, O.V. Vanteeva, E.A. Mikhailovsky, R. Pop, P. Vorobev, M. Chertkov. Smarter Smart District Heating. Proc. IEEE. **108(9)**, 1596 (2020).

[3] V.A. Sednin, A.V. Sednin. Analysis of the State and Main Trends in the Development of District Heating Systems in Belarus. Energy and management. **6(93)**, 2 (2016). (in Russian).

[4] V.A. Sednin, A.V. Sednin. Integration Trends in Heat Supply Systems. Power efficiency. **6**, 23 (2021). (in Russian).

[5] A.V. Sednin, A.V. Zherelo. An approach to data processing for the smart district heating system. Energetics. Proc. CIS higher education institutions and power engineering associations. **65(3)**, 240 (2022).

[6] V.A. Sednin, A.A. Gutkovskii. Modernization and automation of the heat supply system. Belarus experience. Energy saving. **8**, 52 (2016). (in Russian).

[7] V.A. Sednin, A.V. Sednin. Reconstruction of

heat sources of the Branch “Minsk Heat Networks” RUE “Minskenergo”. Energy Strategy. **6**, 16 (2020). (in Russian).

[8] M. Neumayer, D. Stecher, S. Grimm, A. Maier et al. Fault and anomaly detection in district heating substations: A survey on methodology and data sets. Energy. **276** (2023). https://doi.org/10.1016/j.energy.2023.127569.

[9] H. Johra, D. Leiria, P. Heiselberg, A. Marszal-Pomianowska, T. Tvedebrink. Treatment and analysis of smart energy meter data from a cluster of buildings connected to district heating: A danish case. E3S Web Conf. **172** (2020). DOI: https://doi.org/10.1051/e3sconf/202017212004.

[10] M. Schaffer, T. Tvedebrink, A. Marszal-Pomianowska. Three years of hourly data from 3021 smart heat meters installed in danish residential buildings. Sci Data. **9**, 420 (2022).

[11] J. van Dreven, V. Boeva, S. Abghari, H. Grahn, J. Al Koussa, E. Motoasca. Intelligent approaches to fault detection and diagnosis in district heating: Current trends, challenges, and opportunities. Electronics. **12(6)**, 1448 (2023).

[12] T.A. Grigor’eva, A.Yu Bayeva. Application of Predictive Analytics in the Energy Sector. Proc. Bratsk State University. Series: Natural and engineering sciences. **1**, 16 (2021). (in Russian).

[13] D. Geysen, O. De Somer, Ch. Johansson, J. Brage, D. Vanhoudt. Operational thermal load forecasting in district heating networks using machine learning and expert advice. Energy and Buildings. **162**, 144 (2018).

[14] H.G. Bergsteinsson, J.K. Møller, P. Nystrup, Ól.P. Pálsson, D. Guericke, H. Madsen. Heat load forecasting using adaptive temporal hierarchies. Applied Energy. **292**, (2021). https://doi.org/10.1016/j.apenergy.2021.116872.