Environmental and Climate Technologies
2024, vol. 28, no. 1, pp. 670–685
https://doi.org/10.2478/rtuect-2024-0052
https://content.sciendo.com

Exploring Heat Demand Forecasting in District Heating Networks Using Random Parameter Linear Regression Model

Hesham ALI[1*], Andrei DEDOV[2], Anna VOLKOVA[3]

[1–3]*Department of Energy Technology, Tallinn University of Technology, Ehitajate Tee 5, 19086 Tallinn, Estonia*

Received 25.04.2024; accepted 21.10.2024

Abstract **– Accurate forecasting of heat demand in district heating networks is essential for their efficient and sustainable operation. This paper presents a novel approach using a random parameter linear regression model to forecast heat demand, distinguishing itself from classical linear regression models by its ability to address unobserved heterogeneity among parameters. Through a case study in Estonia and utilizing data from 2018 to 2023 and considering seasonality and consumption patterns, the study investigates determinants of heating demand in district heating networks. Two models were trained for heating and non-heating seasons. Results indicate significant impacts of weather conditions, energy prices, time of day, and network infrastructure on heat supply during the heating season, while only time of day and electricity prices were significant drivers during the non-heating season, with no notable influence of weather conditions. Prediction accuracy was slightly enhanced using the random parameter linear regression model, with a mean absolute percentage error of 9.66 % compared to 9.99 % for the Multi Linear Regression Model on the testing set.**

Keywords **– District Heating Networks; forecasting models; heat demand; Random Parameter Linear Regression Model (RPLRM); seasonal variation; weather impact.**

Nomenclature	
DHN	District Heating Network
ML	Machine Learning
MLRM	Multi-Linear Regression Model
RPLRM	Random Parameter Linear Regression Model
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
AIC	Akaike Information Criterion

1. INTRODUCTION

District heating technologies show their dominancy in various regions around the world. For instance, around 15 % of the heating demand in Europe was supplied by district heating

* Corresponding author.
E-mail address: Hesham.ali@taltech.ee

networks (DHN) in 2020 with a steady increase in their share due to their economic and environmental advantages compared to traditional individual heating systems, especially in urban areas [1]. The employed technologies in DHN are witnessing a continuous development contributing to enhance the utilization of sustainable energy sources and reduce the carbon emissions from the energy sector [2], [3].

The utilization of surplus heat from various sources in the DHN is one of the biggest advantages of the DHN technology [4], [5]. For instance, the waste heat from the incineration plant, industrial waste heat, thermal heat from combined heat and power plants, and other sources of waste heat can be integrated into the DHN reducing the fuel consumption for heat [6]. In addition, utilizing renewable energy sources including wind, solar, and geothermal energy in DHN can be promising and significantly contribute to the reduction of carbon emissions [7]. A promising solution for enhancing the utilization of sustainable energy sources is the utilization of large-scale heat pumps, where these technologies enable the system to expand the utilization of other sources of free heat [8], [9]. However, many of these sources of sustainable heat can be exposed to fluctuations due to various reasons, same as the demand for heat in the network. The importance of effective operational management of the DHN appears to secure the supply of heat in the most cost effective, efficient, and eco-friendly way [10], [11].

Accurate and reliable heat demand forecasting is crucial for district heating optimization, operational management, and transition toward more sustainable technologies [12], [13]. Forecasting heat demand in DHNs plays a pivotal role in various strategic aspects including, budgeting, heat generation optimization, network stability, and it guides technology integration decisions, such as thermal energy storage and renewable energy utilization. Understanding heat demand determination factors allows the DHN operators to anticipate the heating demand fluctuations, and response to the heating demand more effectively with minimal energy wastage. In addition, by matching heat supply and demand, the scheduling of energy production can be optimized, reducing the production of excess energy, and maximizing the utilization of renewable energy sources which contribute to maintain the optimal cost of energy production [14].

As heating demand in the DHN is originated from space heating and hot water demand, the main factor affecting the space heating demand is mainly the weather conditions and outside temperature, while the hot water demand might be influenced by more complex factors related to social behavior [15]. Generally, heating demand in the DHN can be influenced by a wide and complex set of factors that interact. Climate conditions (outside temperature, wind, humidity, radiation, etc.) [16], Technical factors (Network renovation and improvement), Economic factors (Heat prices, prices of other heating solutions), Social behaviour factors (working time, holiday time, day activities, etc.), and demographic factors (building densities and size of the family) [13]–[16].

The forecasting of heat demand can be classified into two categories, short-term forecasting which predicts the hourly or minutely demand for the next few days, and long-term forecasting, which predicts from weekly to annual heat demand for different future period [19]. Several techniques of heating demand forecasting are used and investigated in the literature, including statistical approaches, machine learning methods, and hybrid methods as shown in Fig. 1. ML models, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forests (RF), offer the advantage of capturing complex nonlinear relationships and handling large datasets efficiently [17]–[19]. However, they may require extensive data pre-processing and tuning to achieve optimal performance [24]. On the other hand, statistical models like Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing Methods are adept at capturing temporal patterns and seasonality in energy data but may struggle with capturing nonlinear relationships [25]. Hybrid models, such as Long Short-Term Memory (LSTM) networks with exogenous variables and ARIMA combined with machine learning residuals, attempt to provide the strengths of both ML and statistical approaches to enhance forecasting accuracy [26].

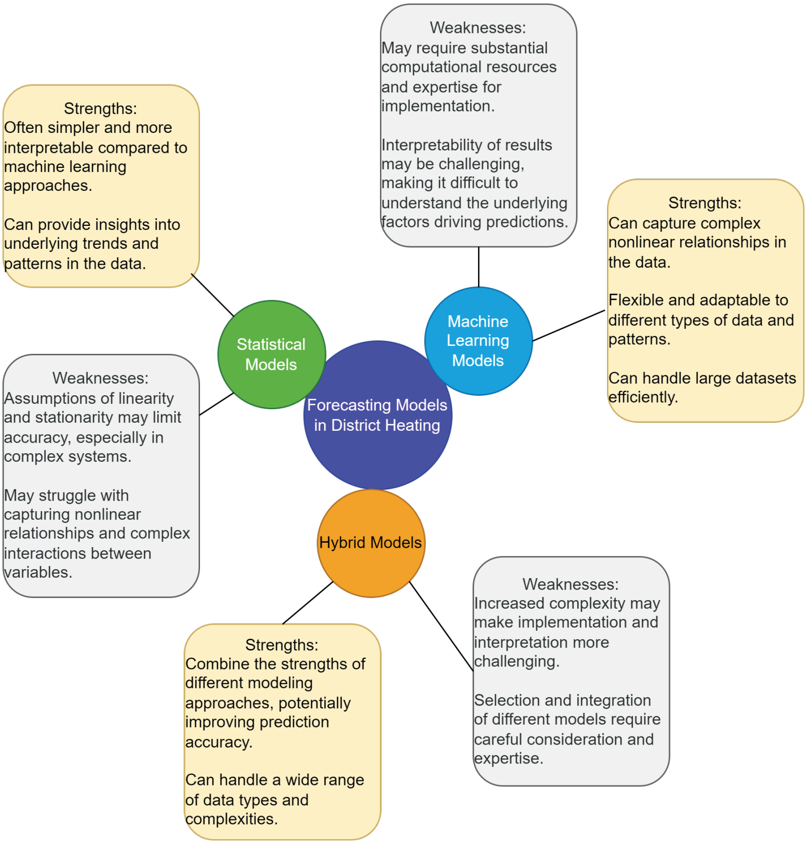

Fig. 1. Forecasting Models for Heating Load in DHN and Assessing Their Strengths and Weaknesses.

Dotzauer [27] has developed a linear regression model to predict the heating load and they considered outside temperature and social behavior as influencing factors. Shakeel *et al.* [28] proposed a novel approach using FB-Prophet model for DHN load forecasting, demonstrating promising results in enhancing forecast accuracy. Zdravkovic *et al.* [29] developed a high-resolution heat demand forecasting model validated at the national level, offering insights into hourly heat demand variations. Additionally, Kemper *et al.* [19] evaluated models for predicting residential unit heat demands, contributing to the optimization of energy management systems.

Studies also investigate explainable forecasting models. Dand *et al.* [30] utilized data from a cogeneration DH plant to build analysis and forecast models, enhancing understanding and predictability of heat load patterns. Dand *et al.* [30] have investigated the influence of five parameters on heating load, namely, wind speed, humidity, outdoor temperature, and holidays. Furthermore, operational strategies integrating forecast results have been proposed.

Golla *et al*. [31] implemented a control strategy for DHNs with heat pumps and storage systems, enhancing forecast outcomes to optimize network operations. Machine learning is widely used and studied in the literature, Ntakolia *et al*. [32] presented a review study about the various ML tools that have been used in DH load forecast and the various parameters that have been considered. Chaganti *et al*. [33] developed a ML model to predict the future heating and cooling load in buildings and Saloux *et al*. [34] used ML algorithm to predict the heating load for a solar DHN with 52 homes. The results of [34] have shown a privilege of ML algorithm over LR methods in heating load forecasting.

Existing econometric models for heating demand forecasting often overlook the multifaceted impact of various factors beyond ambient temperature and time of the day [35]. In addition, existing forecasting techniques are often based on fixed-parameter approach, which assume that impact of each parameter on heating demand is always the same. This paper introduces a novel Random Parameter Linear Regression Model (RPLRM) that addresses this gap by incorporating a broader range of determinants and addresses the heterogeneity for the parameters which potentially will lead to enhancement of the accuracy of heating demand forecasting within DHN. RPLRM has been used in various fields and sectors. Mannering *et al.* [36] developed a RPLRM for the accident on highways and estimated the unobserved heterogeneity in the explanatory factors. Hamed *et al.* [37] utilized RPLRM to investigate vehicle ownership decisions following traffic accidents, focusing on the impact of the current car fleet on individuals' choices. Ghiasi *et al.* [38] applied RPLRM in their statistical assessments of higher education ranking methodologies in the USA, aiming to unravel the quantitative metrics utilized in ranking educational institutions. Provencher *et al.* [39] employed RPLRM in their economic analysis of fisheries production, aiming to understand market dynamics, resource management strategies, and economic ramifications within the fisheries industry.

RPLRMs have emerged as a promising approach in energy forecasting, offering enhanced accuracy and robustness in predicting energy consumption patterns. Studies such as [40] have developed heterogeneity-based econometric models that leverage RPLRM to forecast annual electricity consumption, showcasing the effectiveness of this methodology. RPLRM enables the incorporation of diverse factors and individual characteristics into the forecasting process, leading to more nuanced and accurate predictions. Additionally, Verwiebe *et al.* [41] conducted a systematic literature review on energy demand modelling, highlighting RPLRM as one of the notable methodologies utilized in the field. This indicates the growing recognition of RPLRM's relevance and applicability in addressing the complexities of energy forecasting tasks. Furthermore, Hamed *et al.* [42] explored RPLRM's utility in forecasting annual electric power consumption, emphasizing its capability to account for heterogeneity in both means and variances. The integration of RPLRM into energy forecasting frameworks enables analysts to capture the inherent variability and uncertainty present in energy consumption data, leading to more robust and reliable predictions. Moreover, innovative approaches, such as combining RPLRM with inverse matrix calculation, have been proposed to enhance energy demand forecasting in remote areas [42]. This demonstrates the versatility of RPLRM in adapting to different forecasting contexts and addressing specific challenges in energy prediction tasks.

The literature review shows the importance of accurate and reliable heating demand prediction and the significance of understanding the impact of various parameter on heating demand. By developing a robust forecasting model that considers diverse factors influencing heating demand, this research offers a valuable contribution to enhancing the efficiency and reliability of DHN. The introduction of the RPLRM addresses the limitation of the conventional econometric models by providing a framework to account for unobserved heterogeneity for the determinants, resulting in more accurate and reliable forecasting. The paper proceeds with a description of the dataset used, followed by an explanation of the methodology employed to develop the forecasting models. Subsequently, the results of the model's application are presented, followed by a discussion of their implications and conclusions drawn from the study.

2. DATASET DESCRIPTION

The dataset used in this study was sourced from the DHN in Narva, located in north-eastern Estonia. The DHN operates with a Combined Heat and Power (CHP) plant with a heat capacity of 160 MW_{th} and an electrical capacity of 215 MW_{el}. In 2022, the annual heat demand of the network was approximately 400 GWh. With a network length of 80 km, it serves over 730 buildings, maintaining a heat supply and return temperature of 120/70 [43].

The dataset comprises hourly data spanning from 2018 to 2023, including heat supply to the network, and the heat consumed by consumers in the previous 24 hours as well as weather conditions such as temperature, wind speed, and radiation. Energy price data, including heat price and electricity price, were also collected. Additionally, binary variables were created to represent seasonality and patterns for different times of the day, weekends, and holidays. The study segmented the time of day into four distinct periods: morning hours from 6 to 9, working hours from 9 to 17, evening hours from 17 to 22, and night hours from 22 to 6. Infrastructure data, specifically the share of pre-insulated pipes in the network, was also included. Table 1 presents the statistical overview of the dataset.

TABLE 1. STATISTICAL DESCRIPTION OF THE DATA

	Unit	**Mean**	**Standard deviation**	**Max**	**Min**
Heat supply	MW	49.02	34.15	163.00	5.00
Temperature	°C	6.57	9.40	34.30	-28.10
Night hours	Binary	0.33	0.47	1.00	0.00
Morning hours	Binary	0.12	0.33	1.00	0.00
Working hours	Binary	0.33	0.47	1.00	0.00
Evening hours	Binary	0.21	0.41	1.00	0.00
Weekend and holidays	Binary	0.30	0.46	1.00	0.00
Heat price	€/MWh	39.43	8.52	73.35	35.33
Electricity price	€/MWh	82.83	85.96	4000.00	-60.04
Irradiance	W/m	116.13	198.22	908.00	0.00
Wind Speed	m/s	3.60	1.90	14.60	0.10
Share of pre-insulated pipes	%	36	0.01	37	35

3. METHODOLOGY

The study conducted an analysis of the impact of ten factors on heat supply and developed a novel forecasting model, which was then compared with the classical MLRM. The initial phase of the research involved collecting data on heat supply and consumption from the Narva DHN in Estonia for the years 2018 to 2023. Hourly historical data on various parameters, including weather conditions, heat price, electricity price, and the number of connected buildings, were collected.

The primary objective of this research was to assess the performance of the RPLRM in forecasting heat supply and estimating the precise impact of the determinants on heat supply. This research distinguishes itself from classical models by considering unobserved heterogeneity in the independent factors. Therefore, the RPLRM was compared with the classical MLRM.

The dataset was split into training data (2018–2022) and testing data (2023) to facilitate model development and evaluation. Additionally, recognizing the different dynamics of heat supply between cold and warm seasons, the dataset was further split into heating and non-heating seasons. The beginning and end of the heating season were determined based on observations of several consecutive hourly heat supply surpassing or falling below a threshold of 25 MW, respectively. The threshold of 25 MW was determined based on observations of the heat supply dynamics between the warm and cold seasons. It was noted that when the heat supply dropped below 25 MW, it stabilized at an average of 13 MW during the non-heating season. Conversely, when the heat supply exceeded 25 MW, it averaged around 70 MW during the heating season.

Fig. 2 illustrates the framework developed for the creation of data-driven forecasting models in DHNs. The figure outlines the sequential process involved in the development and evaluation of these models. Beginning with data input, including parameters such as hourly heat supply, weather conditions, energy prices, and infrastructure data, the process proceeds through data processing stages.

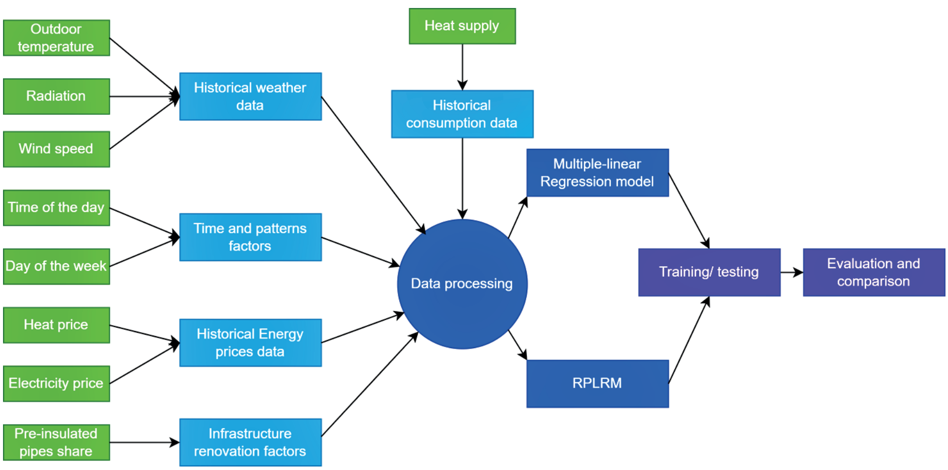

Fig. 2. Framework for Developing the Data-Driven Forecasting Models in District Heating Systems.

The selection of the factors for the forecasting models was guided by a comprehensive review of existing literature and preliminary data analysis. Factors such as weather conditions, time and patterns, and infrastructure factors were chosen based on their established significance in prior studies [13]–[16]. Electricity prices and heat prices were included due to their impact on operational cost dynamics within the heating systems. The framework acknowledges the impact of factors such as consumer behaviour, occupancy patterns, response to pricing, external events, heating preferences, and social behaviour. However, due to the difficulty in measuring these factors, they were not explicitly included in the model development process. Subsequently, the Multi-Linear Regression Model (MLRM) and RPLRM are constructed, followed by testing and evaluation phases to assess the models' forecasting accuracy and performance.

The modelling process commenced with the construction of MLRM and RPLRM for all parameters, followed by iterative refinement to eliminate insignificant variables. The goodness of fit and forecasting accuracy of the models were evaluated, and a comparative analysis between MLRM and RPLRM was conducted. All analyses in this study were conducted using the R programming language. The results of the developed models were then applied to the testing dataset to assess their predictive performance.

3.1. Forecasting Models

3.1.1. Multi-Linear Regression Model

MLRM is one of the most traditional and simplest used techniques to predict the relationship between the dependent variable and one or more independent variables. The linear regression model can be represented by Eq. 1.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon_i, \tag{1}$$

where Y_i is the dependent variable for the i^{th} observation, in our case it is the heating demand. β_0, β_1, β_k are the coefficient to be estimated, and ε_i is the error term.

The coefficient β_0, β_1, to β_k were estimated using the ordinary least square method as shown in Eq. 2.

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \tag{2}$$

where $\hat{\beta}$ is the estimated coefficient, X is the matrix of independent variables, X^T is the transpose matrix of X, and Y is the vector of the dependent variable (heat supply).

3.1.2. Random Parameter Linear Regression Model

RPLRM is a variant of regression model that account for heterogeneity in the coefficient of the independent variables. Unlike the traditional linear regression models where coefficients are fixed across the observations, RPLRM allows for variability in these coefficients, making it more flexible and adaptable to diverse datasets. Eq. 3 show the general equation of RPLRM:

$$Y_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k + \xi_i, \tag{3}$$

where Y_i is the dependent variable for the i^{th} observation, α_o, α_1, α_k – the coefficient to be estimated, and ξ_i is the error term.

The distinguished characteristic about the RPLRM is the estimation of the coefficients α_o, α_1, α_k. In order to estimate the coefficients, Eq. 4 was used by assuming normal distribution:

$$\alpha_{ji} \sim N\left(\mu_j \sigma_j^2\right), \tag{4}$$

where μ_j is the mean of the distribution for the j^{th} coefficient, and σ_j is the variance of the distribution for the j^{th} coefficient.

The estimation of the RPLRM is conducted by the maximum likelihood estimation technique (MLE). where the likelihood function is maximized to find the parameter values that best fit the observed data. The likelihood function for the RPLRM is shown in Eq. 5.

$$L(\theta|Y) = f(y_1|\theta), f(y_2|\theta) \ldots f(y_n|\theta), \tag{5}$$

where θ represents the vector of model parameters, including the means μ_i, and variances, $Y = y_1, y_2, \ldots y_n$ represents the observed data, $f(y_i|\theta)$ is the probability density function for the i^{th} observation.

3.2. Performance Metrics

The performance of both the MLRM and RPLRM was rigorously evaluated using established metrics, notably the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) as shown in Eq. 6 and Eq. 7. These metrics served as benchmarks for assessing the accuracy and efficacy of the forecasting models.

$$\text{MAPE} = \frac{1}{\text{n}}\left(\left|\frac{\Sigma e_i}{Y_i}\right|\right) \cdot 100\%, \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{\text{n}}\sum_{i=1}^{n}(e_i)^2}, \tag{7}$$

where, e_i is the residual between the predicted value and the actual value at point i, and Y_i is the actual value for the same point, and n is the number of observations.

4. RESULTS

This section shows the estimated results of the performed models in this study. The estimated results include the influence of the significant parameters on heat supply, the accuracy of both models, MLRM and RPLRM in forecasting heating demand, the difference in parameters impact on heat supply in heating season and non-heating season and shows the prediction accuracy in the testing set.

4.1. Training Set

As a result, both MLRM and RLPRM demonstrated significant findings in estimating the determinants of heat supply in the network. Table 2 illustrates the close results in parameter coefficient estimates and their impact on the DHN's heat supply during both heating and non-heating seasons. The results in Table 2 are elaborated in the following subsections, which focus on the heating season, non-heating season, random parameters, and forecasting accuracy.

TABLE 2. RESULTS OF HEAT SUPPLY MODELLING BY THE FOUR MODELS

Variable	**Heating season**				**Non-heating season**			
	MLRM		**RLPRM**		**MLRM**		**RLPRM**	
	Coefficient estimate, β_n	t-stat	Coefficient estimate, α_n	t-stat	Coefficient estimate, β_n	t-stat	Coefficient estimate, α_n	t-stat
Intercept	78.80	31.11	82.64	32.29	0.82	10.43	1.78	4.64
Hourly temperature	–2.20	–116	–2.20	–110	–	–	–	–
Standard deviation of the random parameter – normal distribution	–	–	1.24	–	–	–	–	–
Night hours	–4.72	–36.94	–4.81	–37.85	–1.80	–33.08	–1.82	–46.03
Morning hours	3.30	19.58	3.26	19.45	0.89	12.87	0.90	5.15
Standard deviation of the random parameter – normal distribution	–	–	–	–	–	–	0.12	–
Working hours	–3.53	–26.08	–3.57	–26.62	–0.11	-2.14	–	–
Weekend and holidays			–0.26	–2.63	–0.14	-3.34	–0.15	–3.66
Heat price	–0.05	–5.19	–0.04	–4.04	–	–	–	
Electricity price	0.01	10.06	0.01	9.16	0.01	2.56	0.01	3.52
Irradiance	0.00	–2.92	0.00	–3.31				
Wind Speed	0.63	25.22	0.62	17.19	–	–	–	–
Standard deviation of the random parameter– normal distribution	–	–	0.40	–	–	–	–	–
Percentage of pre–insulated pipes	–0.88	–10.66	–0.97	–11.66	–	–	–	–
Model statistics								
Log-likelihood at convergence	–95 469.80		–95 358.04		–36 861.67		–36 888.69	
Number of observations	27 864		27 864		15 936		15 936	
Number of parameters	9		12		5		5	
Goodness of fit								
R-squared	0.91		0.92		0.75		0.80	
AIC	190 963.60		190 748.10		73 793.38		73 739.33	
Accuracy measures								
MAPE	9.33 %		8.99 %		10.10 %		9.72 %	
RMSE	7.44		7.19		2.12		1.89	

4.2. Heating Season

As anticipated, the primary driver of heat supply during the heating season is the hourly temperature, with a 1-degree decrease corresponding to a 2.2 MW increase in heat supply. Moreover, time of day also significantly influences heat supply, with morning hours witnessing an increase of 3.26 to 3.30 MW, while working and night hours' experience decreases of approximately 3.5 MW and 4.75 MW, respectively.

Additionally, the results reveal a statistically significant impact of heat and electricity prices on heat supply. Despite the stability of heat prices in Narva, with only three changes observed over a six-year period from 2018 to 2023, a 1 % increase results in a 0.05 MW decrease in hourly heat supply during the heating season. Similarly, a 1-euro increase in electricity prices leads to a 0.01 MW increase in hourly heat supply, possibly due to consumers who have more than one heating option, switch to district heating options when electricity prices rise.

The findings also highlight the significant influence of wind speed during the heating season, with a one-unit (1 m/s) increase contributing to a 0.63 MW increase in heat supply. This suggests that higher wind speeds facilitate outdoor wind infiltration into buildings, increasing heating demand. Moreover, infrastructure factors, such as the share of pre-insulated pipes in the network, exhibit a significant impact, with a 1% increase resulting in a reduction of heat supply by 0.88 to 0.97 MW hourly, attributed to decreased heat loss in the network. This result clearly indicates the effectiveness of preinstalled pipes and their impact on heat supply. Furthermore, the RPLRM effectively captures the influence of weekends and holidays, indicating a reduction of 0.26 MW per hour in heat supply during these periods.

4.3. Non-Heating Season

Forecasting heat supply during the non-heating season proved challenging due to the different consumption patterns. Unlike the heating season, where space heating is a primary usage of heat and measured conditions are main drivers, the non-heating season relies on domestic hot water (DHW) usage. This DHW patterns is influenced more by behavioural factors rather than directly measurable or observed variables. Interestingly, our analysis revealed that neither the fluctuation in hourly temperature nor the average daily temperature were statistically significant factors affecting heat supply during this period. Instead, the primary influencers were found to be seasonality and time of day. Specifically, the time of day and weekends emerged as the main drivers of heat supply fluctuations during the non-heating season, overshadowing other factors such as weather conditions, pricing dynamics, or infrastructural variations. Our results indicated a notable increase in heat supply by around 0.89 MW during morning hours, contrasted by a decrease of 1.80 MW and 0.11 MW during night and working hours, respectively. Moreover, a decrease of 0.14 MW was observed during weekends and holidays.

4.4. Random Parameter Analysis

The RPLRM analysis for the heating season revealed significant random parameters associated with two variables: hourly temperature and wind speed. This finding underscores the nuanced impact of wind speed and temperature fluctuations during the heating season on heat supply dynamics. Specifically, both variables were found to exert a negative influence on heat supply, aligning with prior discussions. However, the analysis revealed heterogeneity in their effects, with standard deviations of 1.24 and 0.40, respectively. This variability underscores the presence of unobserved factors that modulate their impact on heat supply.

Moreover, the analysis of the non-heating season identified a significant random parameter associated with morning hours, characterized by a standard deviation of 0.12. This finding suggests that the heat supply during morning hours in the summer season is influenced by additional, unobserved factors. These factors contribute to the observed heterogeneity in the impact of morning hours on heat supply dynamics. Further exploration of these latent influences could provide valuable insights into the factors driving heat demand during non-heating periods.

4.5. Forecasting Accuracy

The assessment of forecasting accuracy highlighted a slight advantage of the RPLRM over the MLRM in both heating and non-heating seasons. During the heating season, the RPLRM achieved a MAPE of 8.99 %, compared to 9.33 % for the MLRM, indicating superior predictive performance of the former. Similarly, in the non-heating season, the RLPRM exhibited a slightly lower MAPE of 9.72 % compared to 10.10 % for the MLRM, reflecting its higher accuracy in forecasting heat demand dynamics during the summer months.

It is important to acknowledge that despite the RPLRM incorporating the consideration of unobserved heterogeneity, the improvement in prediction accuracy compared to the MLRM was not notably significant. Although the RPLRM demonstrated slightly enhanced accuracy during both seasons, it remained very close to the performance of the MLRM.

The marginal difference in forecasting accuracy between the RPLRM and MLRM suggests that while the RPLRM offers a more sophisticated framework for accounting for unobserved heterogeneity, this additional complexity may not always translate into substantially improved prediction accuracy. Despite this, both models exhibited commendable performance in predicting heat supply within the DHN across various seasons.

4.6. Models' Validation

The validation process involved assessing the forecasting accuracy of the models using data from 2023, which had not been trained in the modelling phase. Results from the testing set indicated a slight advantage of the RPLRM over the MLRM, with the former achieving a MAPE of 9.66 % compared to 9.99 % for the latter. This disparity underscores the superior predictive capabilities of the RPLRM in capturing the nuanced variability within the DHN. Fig. 3 provides a comprehensive visualization of the forecasted heat supply by both the MLRM and RPLRM during the validation period.

Upon scrutinizing the models' predictions for the testing set, it was observed that the maximum residual and higher prediction error occurred in the last hours of the year, suggesting a potential influence from New Year's Eve events and festivities, which was not accounted for in this study. While national holidays were considered collectively with weekends and holidays, the unique impact of major celebrations like New Year's Eve warrants further investigation. Fig. 4, depicting the hourly residuals in the predictions made by both models during the testing set year, may provide further insight into whether the highest residual indeed corresponds to the last observations.

Fig. 5 illustrates the histogram of the residual of the predicted heat supply by both models for the testing set. Notably, the frequencies and values of the residuals in both models exhibit close resemblance, as depicted in the Fig. 5.

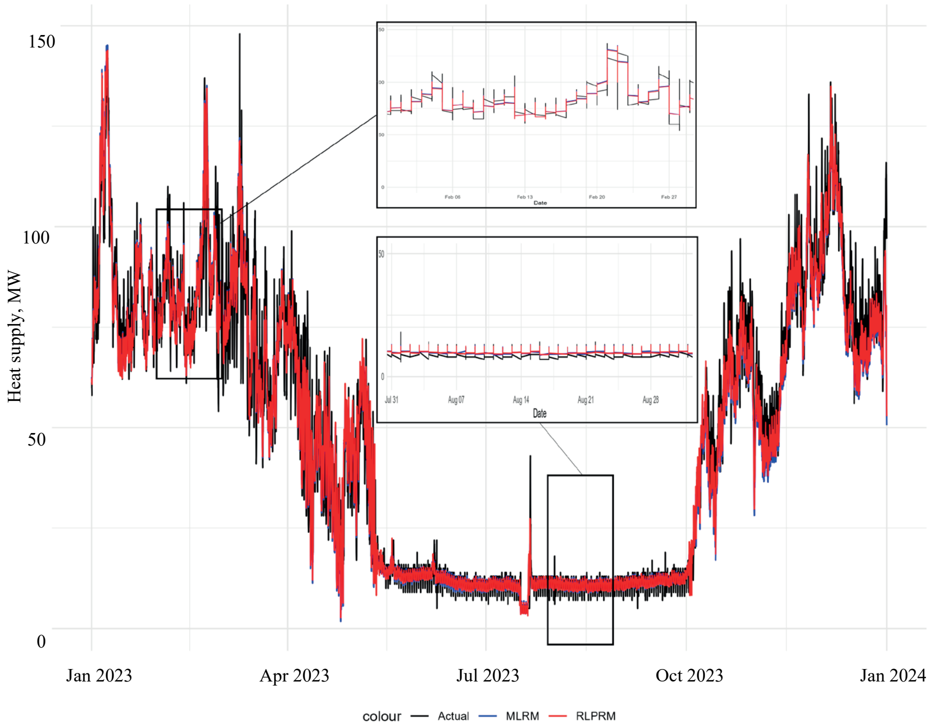

Fig. 3. Actual and Predicted heat supply for the year 2023 by MLRM and RPLRM.

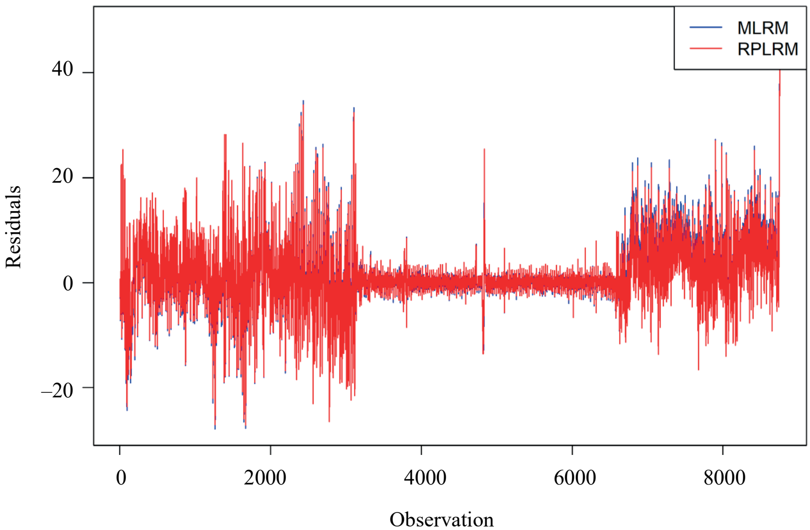

Fig. 4. Hourly residual of the prediction by the two models MLRM and RPLRM for the testing set (the year 2023).

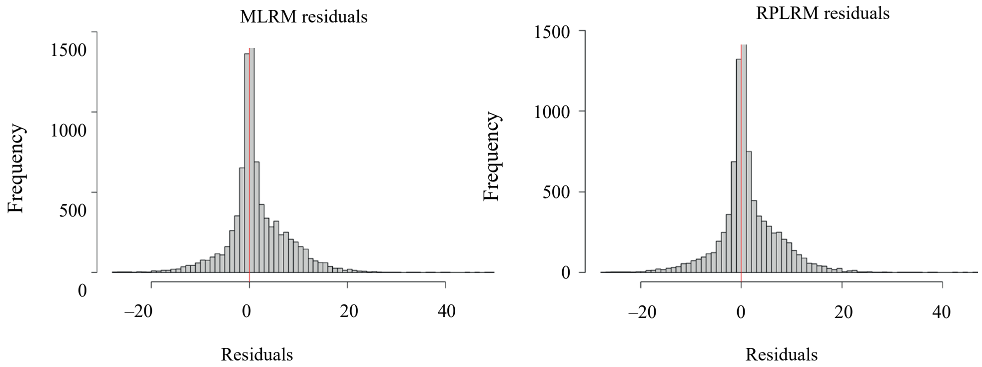

Fig. 5. Histograms depicting the distribution of forecasting residuals for the MLRM and RPLRM for the testing year 2023.

5. Discussion and Conclusions

This study sheds light on the dynamics of heat supply in district heating networks (DHNs) and introduces the RLPRM as a novel approach for predicting heat supply variations.

Both the MLRM and RLPRM effectively estimate heat supply determinants, with close results in parameter coefficient estimates. Temperature is a primary driver of heat supply during the heating season, and hourly variations significantly impact demand. Additionally, time of day reflects consumer behaviour patterns, while heat and electricity prices significantly influence heating demand [44]. Moreover, wind speed and infrastructure factors, such as pre-insulated pipes, played significant roles in determining heat supply dynamics.

Forecasting heat supply during the non-heating season posed challenges due to different consumption patterns, primarily driven by DHW usage. The dominance of seasonality and time of day over weather conditions, pricing, or infrastructure factors underscores the behavioural aspect of heat demand during this period. Despite the absence of statistically significant temperature fluctuations, mornings emerged as a peak period for heat demand, indicating consumer routines [15].

Interestingly, various parameters found to have significant impact on heat supply during heating season but their impact demolished when the heating season ends. Which can be interpret by the different use of heat from the DHN during these two seasons. Our observation about the impact of wind speed on DHN aligns with the finding of [45] where it was found cold wind speed affects heat supply while warm wind speed does not.

The analysis of random parameters in the heating season's RPLRM revealed significant effects of hourly temperature and wind speed on heat supply dynamics. This highlights the intricate relationship between weather variables and heat demand, with both factors exerting a negative influence on heat supply. The observed heterogeneity underscores the presence of unobserved factors that modulate their impact, emphasizing the need for comprehensive modelling approaches to capture these complexities. Similarly, the identification of significant random parameters associated with morning hours during the non-heating season suggests additional latent influences on heat demand, warranting further investigation into seasonal variations and behavioural factors.

The evaluation of forecasting accuracy demonstrated a slight advantage of the RPLRM over the MLRM in both heating and non-heating seasons. While the RPLRM exhibited marginally lower MAPE values, the difference in prediction accuracy compared to the MLRM was not notably significant. Despite incorporating unobserved heterogeneity, the RPLRM's improvement in accuracy was minimal, indicating that the added complexity may not always translate into substantial gains in predictive performance. Nonetheless, both models displayed commendable accuracy in forecasting heat supply within the DHN compared to with used models in the literature [19], [30], [41], [46], highlighting their utility in energy management and planning.

The findings of this study contribute to advancing the understanding of heating demand forecasting in DHNs. However, while this study employed RPLRM to forecast heat supply in DHNs, certain limitations persist. Firstly, despite utilizing the RPLRM's capabilities to account for unobserved heterogeneity, it still operates within the framework of linear modelling, which may not fully capture the intricate relationships within the data. Moreover, the presence of collinearity among the parameters used in the model could potentially distort the estimation of coefficients and undermine the model's predictive performance. Future research endeavours should explore more sophisticated modelling techniques beyond linear modelling to enhance the accuracy and reliability of heat supply forecasting in DHNs.

In conclusion, this study demonstrates the effectiveness of both MLRM and RLPRM in forecasting heat supply in DHNs. While temperature remains a significant factor, other variables such as time of day, economic conditions, and infrastructure play crucial roles in shaping heat demand dynamics. The insights gained from this research can inform policy decisions and infrastructure planning to optimize energy usage and enhance sustainability in urban heating systems.

ACKNOWLEDGEMENT

This work was co-funded by the European Union and Estonian Research Council via project TEM-TA78. The analysis and literature review of the impact factors were specifically analysed in the project RE-PEAK, financed by IEA DHC within Annex XIV. Project partners are Energieinstitut an der Johannes Kepler Universität Linz, Austria, Tallinn University of Technology, Estonia, and Euroheat & Power, Belgium.

REFERENCES

[1] Fallahnejad M. *et al.* District heating potential in the EU-27: Evaluating the impacts of heat demand reduction and market share growth. *Applied Energy* 2024:353:122154. https://doi.org/10.1016/j.apenergy.2023.122154

[2] Werner S. International review of district heating and cooling. *Energy* 2017:137:617–631. https://doi.org/10.1016/j.energy.2017.04.045

[3] Krumins A., Lebedeva K., Tamane A., Millers R. Possibilities of Balancing Buildings Energy Demand for Increasing Energy Efficiency in Latvia. *Environmental and Climate Technologies* 2022:26(1):98–114. https://doi.org/10.2478/rtuect-2022-0009

[4] Pakere I., Feofilovs M., Lepiksaar K., Vītoliņš V., Blumberga D. Multi-source district heating system full decarbonization strategies: Technical, economic, and environmental assessment. *Energy* 2023:285:129296. https://doi.org/10.1016/j.energy.2023.129296

[5] Rieksta M., Zarins E., Bazbauers G. Potential Role of Green Hydrogen in Decarbonization of District Heating Systems: A Review. *Environmental and Climate Technologies* 2023:27(1):545–558. https://doi.org/10.2478/rtuect-2023-0040

[6] Fritz M., Savin M., Aydemir A. Usage of excess heat for district heating – Analysis of enabling factors and barriers. *J Clean Prod* 2022:363:132370. https://doi.org/10.1016/j.jclepro.2022.132370

[7] Madurai Elavarasan R., Pugazhendhi R., Irfan M., Mihet-Popa L., Khan I. A., Campana P. E. State-of-the-art sustainable approaches for deeper decarbonization in Europe – An endowment to climate neutral vision. *Renewable and Sustainable Energy Reviews* 2022:159:112204. https://doi.org/10.1016/j.rser.2022.112204

[8] Volkova A., Koduvere H., Pieper H. Large-scale heat pumps for district heating systems in the Baltics: Potential and impact. *Renewable and Sustainable Energy Reviews* 2022:167:112749. https://doi.org/10.1016/j.rser.2022.112749

[9] Pieper H. *et al.* Optimal usage of low temperature heat sources to supply district heating by heat pumps. Proceedings of ECOS 2017: 30th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, 2017.

[10] Pardo-Bosch F., Blanco A., Mendoza N., Libreros B., Tejedor B., Pujadas P. Sustainable deployment of energy efficient district heating: city business model. *Energy Policy* 2023:181:113701. https://doi.org/10.1016/j.enpol.2023.113701

[11] Rashidi Zadeh D., Derakhshan G., Mehdi Hakimi S., Abdi B. An Economic and Environmental Optimization Model in a Micro Grid with Demand Response. *Environmental and Climate Technologies* 2022:26(1):730–741. https://doi.org/10.2478/rtuect-2022-0056

[12] Bujalski M., Madejski P., Fuzowski K. Heat demand forecasting in District Heating Network using XGBoost algorithm. In *E3S Web of Conferences, EDP Sciences*. 2021. https://doi.org/10.1051/e3sconf/202132300004

[13] Kurek T. *et al.* Heat demand forecasting algorithm for a Warsaw district heating network. *Energy* 2021:217:119347. https://doi.org/10.1016/j.energy.2020.119347

[14] Zhang J., Hu Y., Yuan Y., Yuan H., Mei N. Accuracy improvement of the load forecasting in the district heating system by the informer-based framework with the optimal step size selection. *Energy* 2024:291:130347. https://doi.org/10.1016/j.energy.2024.130347

[15] Johansson C., Bergkvist M., Geysen D., De Somer O., Lavesson N., Vanhoudt D. Operational Demand Forecasting in District Heating Systems Using Ensembles of Online Machine Learning Algorithms. *Energy Procedia* 2017:116:208–216. https://doi.org/10.1016/j.egypro.2017.05.068

[16] Apostolopoulou A., Jimenez-Bescos C., Cavazzi S., Boyd D. Impact of Climate Change on the Heating Demand of Buildings. A District Level Approach. *Environmental and Climate Technologies* 2023:27(1):900–911. https://doi.org/10.2478/rtuect-2023-0066

[17] Ruseljuk P. *et al.* Factors Affecting the Improvement of District Heating. Case Studies of Estonia and Serbia. *Environmental and Climate Technologies* 2021:24(3):521–533. https://doi.org/10.2478/rtuect-2020-0121

[18] Latõšov E., Volkova A., Hlebnikov A., Siirde A. Technical improvement potential of large district heating network: Application to the Case of Tallinn, Estonia. *Energy Procedia* 2018:149:337–344. https://doi.org/10.1016/j.egypro.2018.08.197

[19] Kemper N., Heider M., Pietruschka D., Hähner J. Forecasting of residential unit's heat demands: a comparison of machine learning techniques in a real-world case study. *Energy Systems* 2023. https://doi.org/10.1007/s12667-023-00579-y

[20] Etxebarria M., Oregi X., Grijalba O., Hernández R. J. Relationship between energy demand, indoor thermal behaviour and temperature-related health risk concerning passive energy refurbishment interventions. *Environmental and Climate Technologies* 2020:24(2):348–363. https://doi.org/10.2478/rtuect-2020-0078

[21] Motuzienė V., Bielskus J., Lapinskienė V., Rynkun G. Office building's occupancy prediction using extreme learning machine model with different optimization algorithms," *Environmental and Climate Technologies* 2021:25(1):525–536. https://doi.org/10.2478/rtuect-2021-0038

[22] Potočnik P., Škerl P., Govekar E. Machine-learning-based multi-step heat demand forecasting in a district heating system. *Energy Build* 2021:233:110673. https://doi.org/10.1016/j.enbuild.2020.110673

[23] Eseye A. T., Lehtonen M. Short-Term Forecasting of Heat Demand of Buildings for Efficient and Optimal Energy Management Based on Integrated Machine Learning Models. *IEEE Trans Industr Inform* 2020:16(12):7743–7755. https://doi.org/10.1109/TII.2020.2970165

[24] Moradzadeh A., Mohammadi-Ivatloo B., Abapour M., Anvari-Moghaddam A., Roy S. S. Heating and Cooling Loads Forecasting for Residential Buildings Based on Hybrid Machine Learning Applications: A Comprehensive Review and Comparative Analysis. *Institute of Electrical and Electronics Engineers Inc.* 2022:10. https://doi.org/10.1109/ACCESS.2021.3136091

[25] Kontopoulou V. I., Panagopoulos A. D., Kakkos I., Matsopoulos G. K. A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet* 2023:15(8):255. https://doi.org/10.3390/fi15080255

[26] Ma T., Antoniou C., Toledo T. Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transp Res Part C: Emerg Technol* 2020:111:352–372. https://doi.org/10.1016/j.trc.2019.12.022

[27] Dotzauer E. Simple model for prediction of loads in district-heating systems. *Applied Energy* 2002:73(3–4):277–284. https://doi.org/10.1016/S0306-2619(02)00078-8

[28] Shakeel A., Chong D., Wang J. Load forecasting of district heating system based on improved FB-Prophet model. *Energy* 2023:278:127637. https://doi.org/10.1016/j.energy.2023.127637

[29] Zdravković M., Ćirić I., Ignjatović M. Explainable heat demand forecasting for the novel control strategies of district heating systems. *Annu Rev Control* 2022:53:405–413. https://doi.org/10.1016/j.arcontrol.2022.03.009

[30] Dang L. M. *et al.* Toward explainable heat load patterns prediction for district heating. *Sci Rep* 2023:13:7434. https://doi.org/10.1038/s41598-023-34146-3

[31] Golla A., Geis J., Loy T., Staudt P., Weinhardt C. An operational strategy for district heating networks: application of data-driven heat load forecasts. *Energy Informatics* 2020:3(1):22. https://doi.org/10.1186/s42162-020-00125-5

[32] Ntakolia C., Anagnostis A., Moustakidis S., Karcanias N. Machine learning applied on the district heating and cooling sector: a review. *Energy systems* 2022:13:1–30. https://doi.org/10.1007/s12667-020-00405-9

[33] Chaganti R. *et al.* Building Heating and Cooling Load Prediction Using Ensemble Machine Learning Model. *Sensors* 2022:22(19):7692. https://doi.org/10.3390/s22197692

[34] Saloux E., Candanedo J. A. Forecasting District Heating Demand using Machine Learning Algorithms. *Energy Procedia* 2018:149:59–68. https://doi.org/10.1016/j.egypro.2018.08.169
[35] Bünning F., Heer P., Smith R. S., Lygeros J. Improved day ahead heating demand forecasting by online correction methods. *Energy Build* 2020:211:109821. https://doi.org/10.1016/j.enbuild.2020.109821
[36] Mannering F. L., Shankar V., Bhat C. R. Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal Methods Accid Res* 2016:11:1–16. https://doi.org/10.1016/j.amar.2016.04.001
[37] Hamed M. M., Al-Eideh B. M. An exploratory analysis of traffic accidents and vehicle ownership decisions using a random parameters logit model with heterogeneity in means. *Anal Methods Accid Res* 2020:25:100116. https://doi.org/10.1016/j.amar.2020.100116
[38] Ghiasi A., Fountas G., Anastasopoulos P., Mannering F. Statistical assessment of peer opinions in higher education rankings: The case of US engineering graduate programs. *Journal of Applied Research in Higher Education* 2019:11(3):481–492. https://doi.org/10.1108/JARHE-09-2018-0196
[39] Provencher B., Moore R. A discussion of 'using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model. *Environ Resour Econ (Dordr)* 2006:34(1):117–124. https://doi.org/10.1007/s10640-005-3793-8
[40] Hamed M. M., Ali H., Abdelal Q. Forecasting annual electric power consumption using a random parameters model with heterogeneity in means and variances. *Energy* 2022:255:124510. https://doi.org/10.1016/j.energy.2022.124510
[41] Verwiebe P. A., Seim S., Burges S., Schulz L., Müller-Kirchenbauer J. Modeling energy demand—a systematic literature review. *Energies* 2021:14(23):7859. https://doi.org/10.3390/en14237859
[42] Klyuev R. V. *et al.* Methods of Forecasting Electric Energy Consumption: A Literature Review. *Energies* 2022:15(23):8919. https://doi.org/10.3390/en15238919
[43] Ruseljuk P., Lepiksaar K., Siirde A., Volkova A. Economic dispatch of chp units through district heating network's demand-side management. *Energies (Basel)* 2021:14(15):4553. https://doi.org/10.3390/en14154553
[44] Neagu O., Teodoru M. C. The relationship between economic complexity, energy consumption structure and greenhouse gas emission: Heterogeneous panel evidence from the EU countries. *Sustainability (Switzerland)* 2019:11(2):497. https://doi.org/10.3390/su11020497
[45] Wojdyga K. An influence of weather conditions on heat demand in district heating systems. *Energy Build* 2008:40(11):2009–2014. https://doi.org/10.1016/j.enbuild.2008.05.008
[46] Huang Y., Zhao Y., Wang Z., Liu X., Liu H., Fu Y. Explainable district heat load forecasting with active deep learning. *Applied Energy* 2023:350:121753. https://doi.org/10.1016/j.apenergy.2023.121753