Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 7 (2021) 119–127
www.elsevier.com/locate/egyr

17th International Symposium on District Heating and Cooling, DHC2021, 6–9 September 2021, Nottingham, United Kingdom Country

Backward simulation of temperature changes of District Heating networks for enabling loading history in predictive maintenance

Pakdad Pourbozorgi Langroudi[a,*], Ingo Weidlich[a], Stefan Hay[b]

[a] HafenCity University Hamburg, Henning-Voscherau-Platz 1, 20457 Hamburg, Germany
[b] AGFW |Der Energieeffizienzverband für Wärme, Kälte und KWK e.V., Stresemannallee 30, 60596 Frankfurt, Germany

Received 23 August 2021; accepted 12 September 2021

Abstract

District Heating (DH) networks, like most of industries, are in transition to the fourth industry age and they are retrofitting themselves with different sensing and inspection technologies to enable cyber connectivity for different purposes, such as system optimization, failure detection, maintenance, etc. Since DH pipes show different ageing behaviour under different conditions and initially the pre-insulated bounded pipes had been designed for a minimum of 30 years life span, a long-term loading history is required for predictive maintenance (PdM) purposes and it is necessary to understand the ageing of the DH pipes. These historical temperature changes of the networks are not available for such a long period and they are usually limited to the past few years. To exploit the available implemented technologies for PdM , the missing data must become available to understand the ageing patterns and expand the ageing model to the pipes in use. In this research, various Machine Learning (ML) techniques such as Support Vector Machine (SVM), Random Forest algorithm (RF), Artificial Neural Networks (ANN) have been tested to train a model and backward simulate the temperature changes of the system based on recorded weather data. Various none-temperature variables have been used to enhance the prediction qualities to the real-world data. The historical temperature changes of the system shall be used for different ageing estimation such as fatigue cycles or remaining useful life of the polyurethane (PUR) foam.
© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the scientific committee of the 17th International Symposium on District Heating and Cooling, DHC2021, 2021.

Keywords: Backward simulation; Loading history; Predictive maintenance; Machine learning; Asset management; Artificial neural networks; System reliability; District Heating

1. Introduction

District Heating (DH) is an infrastructure, which transfers energy from a mass production heat source to numerous consumers. The DH system generally constitutes of three main parts: 1- Heat source, 2- Transmission network, 3- Consumer. The current distribution networks, specially the urban grids, consist of various pipe system, including 2nd,

* Corresponding author.
E-mail address: pakdad.langroudi@hcu-hamburg.de (P. Pourbozorgi Langroudi).

https://doi.org/10.1016/j.egyr.2021.09.031
2352-4847/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the scientific committee of the 17th International Symposium on District Heating and Cooling, DHC2021, 2021.

Nomenclature

AI	Artificial Intelligence
ANN	Artificial Neural Networks
DH	District Heating
DL	Deep Learning
DTR	Decision Tree Regression
DWD	Deutscher Wetterdienst
HGBR	Hist Gradient Boosting Regression
KNN	K-Nearest Neighbours Regression
IoT	Internet of Things
LAS	LassoCV
LR	Linear Regression
LSTM	Long Short-Term Memory
MEA	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perceptrons
MSE	Mean Squared Error
PdM	Predictive Maintenance
PLS	Partial Least Squares Regression
PUR	Rigid Polyurethane
RDG	Ridge Regression
R2	Coefficient of Determination
RF	Random Forest
RFR	Random Forest Regression
RNN	Recurrent Neural Networks
SVM	Support Vector Machine
SVR	Support Vector Regression

3rd, and 4th generation [1]. The grids have been expanded with time based on the demand the asset owners expanded their networks. This brought more complexity to the system in terms of maintenance and asset management. The utilities changed their operating mode from constant temperature to dependent mode. While this change in the system contributed in energy efficiency and drop of temperature in return pipes, it introduced a new force to this complex system as thermal fatigue and cyclic load. All the mentioned evolutions led to different ambiguities for understanding the ageing mechanism and the influence of various causes to the system.

The transition to industry 4.0 provided immense amount of data ready to be translated to information. On the one hand, the data collection from different sensing and inspection technologies, and interconnectivity with Internet of Things (IoT), and on the other hand, the new artificial intelligence (AI) approaches such as machine learning (ML) and deep learning (DL), have introduced a new term in the maintenance dictionary as Predictive Maintenance (PdM). The data quality plays a major role in PdM.

The data availability in DH sector is limited or restricted. The DH asset owners and grid operators are in the digitalization transition. Therefore, a big fraction of information is not available and mostly they are limited to the past few years. However, as mentioned before, DH has shown a great stamina and the pipes are sustaining more than expected. Therefore, HCU developed a new model to backward simulate the temperature changes of the system with the help of ML. This model uses the temperature and non-temperature variables to find the correlation among them and the real-world operation data of the past few years. This model enables to derive and extract multiple ageing parameters such as steel fatigue, Polyurethane foam thermo-oxidation, full load temperature cycle, etc. These derivations could be used as input data for PdM model. Alternatively, the damage statistics provided by multiple grid operators will be utilized to find the correlation among different ageing models and cause of damage at the designated time.

In this paper, we would like to discuss a developed approach to overcome the scarcity of the available data and backward simulate the temperature changes of DH networks to prepare it to adopt the PdM models for the network. The aim of this research is not only PdM but also ease of asset management for the grid operators. This paper is not discussing the complete PdM model in DH systems, it is only a sequence of a PdM model, which enables to have failure predictions in DH systems.

1.1. Predictive maintenance

PdM enables to learn from the past data in order to predict the future and it is able to adopt the big-data approach. Therefore, the more experienced industries in data processing have targeted to collect the necessary data in order to have higher accuracy predictions. In terms of maintenance, the main goal of predictions is about failures and breakdowns. Therefore, as latest maintenance strategy, many sectors have employed PdM for their business. This trend has mainly observed in fields where reliability is of paramount importance, such as power plants, oil and gas industry, and utilities [2].

1.2. Role of computer science in predictive maintenance

Based on the form, size, and the quality of the data availability there are different ANN architecture that can be used in PdM application. Therefore, a review on the state of the art of the ANN architecture is required for developing the PdM model and find the correlation based on different input data for predictions. Meireles et al. conducted a comprehensive review on ANN use in industrial application divided the ANNs in to eleven groups, which the Multilayer Perceptrons (MLPs) were at rank one in usage [3]. In recent years, with development of IoT equipment and data availability in real-time, employment of Recurrent Neural Networks (RNN) has been observed for real-time monitoring. This real-time data is not only useful for monitoring, but also PdM could get benefit of it. Sharma employed RNN in combination with a long short-term memory (LSTM) for predictive maintenance and asset management in oil and gas industry [4]. In DH sector, ANNs mainly have been used for operation optimizations and the PdM models are in research and development phase.

2. Material and methods

In this research, various supervised Machine Learning (ML) methods have been tried for backward-simulation purpose. Three different grid operators have provided the sample data for temperature-changes of the DH networks. The data format was hourly based and the length of the time-series was for three continues years. The data was logged in a substation of each DH network. To form the training set additional weather data was required, which has been obtained from Deutscher Wetterdienst (DWD) ftp server. Based on the availability and quality of the weather data, different features have been used for training and prediction. These features are not only the temperature variables, but also non-temperature variables as well such as relative humidity, cloud cover, precipitation height and form, sunshine duration, mean wind speed, month, hour, and off days.

The criteria to choose a weather station was to compromise between the proximity to the DH substations and completeness of the data and date of the records. However, some weather stations in some cases were closer to the DH substations, but the recorded data was not sufficient to cover the age of the DH network pipes.

There are some missing data in temperature change record of the DH substations. The reason is unknown. This could be a potential failure of the logging device or a shutdown for maintenance, which could have effect on the simulation results. As to not lose much data, a large negative number has replaced the missing variables. This helped to handle null values as outliers.

The mentioned outdoor temperature and non-temperature variables have been used as features and the temperature change record of the DH substations has been used as label. Based on the availability and cleanness of the feature data, different features have been used for backward simulation for each DH network.

2.1. Data evaluation

The correlations among features have been calculated in two Pearson and Spearman methods to see if there is any monotonic effect. No significant change has been observed between correlation of the supply and return to the rest of the variables in both methods. Non-temperature variables have less correlation but they increase the prediction quality considerably. Among non-temperature variables, relative humidity, sunshine duration, month, and hour contributed most in simulation.

The quality of each data set was different in terms of completeness and availability of all features for a long period. However, including the hydraulic model in data-sets increase the accuracy of predictions significantly but it cannot be used because these data is also limited to the past few years. In addition, only one network were able to deliver the hydraulic model for investigation.

2.2. Regressor selection

Since predicting temperature is a continuum, the regression algorithms have been employed for predictions. To choose the best solution for backward-simulation, a cross validation with 10 fold have been iterated through all data-sets. This process has been repeated for different algorithm to make the comparison available with the help of multiple metrics for scoring. The algorithms which have been tested are as follow:

- Linear Regression (LR)
- Decision Tree Regression (DTR)
- Ridge Regression (RDG)
- K-Neighbours Regression (KNN)
- Partial Least Squares Regression (PLS)
- Support Vector Regression (SVR)
- Random Forest Regression (RFR)
- LassoCV (LAS)
- Hist Gradient Boosting Regression (HGBR)

As Fig. 1 illustrates, RFR shows the best scoring among the named algorithms. In all three metrics that have been chosen for comparison, including mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R^2) metrics, RFR is showing an outstanding performance.

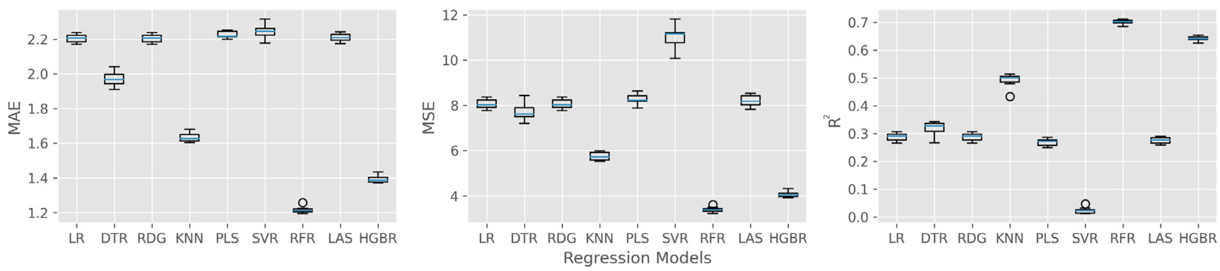

Fig. 1. Model comparison and evaluation with three MEA, MSE, and R^2 metrics.

After RFR, the HGBR and KNN respectively have shown the best results. The rest of the algorithms were approximately in the same tolerance of predictions. Based on the above results RFR has been chosen for the backward simulation and further predictions. Each data-set has been split into training and testing sets with the test size of 20%.

3. Results and discussion

The backward simulation has been run for three different utilities with Random Forest algorithm. The quality of predictions in supply pipes result always higher accuracy between 0.81 to 0.92 and for the return pipes the accuracy observed between 0.59 to 0.77 (see Table 1).

Table 1. Sample description.

Network	Event counts	Outdoor temp. correlation to supply	Outdoor temp. correlation to return	Supply accuracy	Return accuracy	Training algorithm
A	25537	−0.66	0.41	0.81	0.69	RFR
B	26280	−0.64	0.52	0.92	0.75	RFR
C	24394	−0.86	0.45	0.88	0.59	RFR
D	25171	−0.66	0.30	0.92	0.77	RFR

The difference of prediction qualities between supply and return is because of user behaviour and consumption habits. Grid operators know the supply temperature based on the outdoor temperature and in high demand cases with more pumping, it is possible to compensate for the exceeded demand. In return pipes, the temperature depends on the instantaneous consumption, which this depends on user behaviour and living habits. Even though the return predictions are not as accurate as the supply, the observed results are inside the expected fluctuation range. This enables the estimation of the remaining useful life of the PUR foam with calculating the subjected hours of the foam at each temperature range. The Fig. 2 illustrates the quality of the simulation values compared to their original ones. As it is visible in this graph, the simulated values are mostly within the ranged of peak and trough. Therefore, it was plausible to miss some cycles on higher ranges and in return, an increase counts in lower ranges was expected. Thus, a comparison between cycles ran through all for both real and simulated data. Thermal cycle counting is at importance in this research, because it is the key to reveal the damage accumulation for the steel medium pipe. For this purpose, the Rainflow-counting algorithm has been employed [5]. According to EN 13941-1:2019 [6], the number of equivalent full temperature cycles can be calculated from:

$$N_o = \frac{\sum n_i \cdot (\Delta T_i)^m}{(\Delta T_{ref})^m} \tag{1}$$

Where

n_{i} is the number of cycles with temperature range ΔT_{i}

ΔT_{ref} is the reference temperature at which N_{o} is calculated

m is the constant in the SN-curve

There is no clear advice at which step of degree centigrade, the ΔT shall be set and cycle counts shall be calculated. Various step sizes such as 1 °C, 2 °C, and 5 °C had been used for fatigue analysis of DH systems [7]. A calculation with 5 °C intervals could be found in Table 3 in Appendix. In this research, the one degree centigrade step size has been used for cycle counting and the equivalent full temperature cycles has been calculated with the formula (1). As it was expected, a shift is occurred on higher temperature differences on every network (see Fig. 3). In all four networks, for each real count with the delta temperatures bigger than 20 degree, we see the corresponsive simulated cycle count but in a lower temperature range for both supply and return time-series. Since higher ranges have greater influence especially on damage accumulation of the steel pipes, it is decisive to identify this deviation between the real and simulated values and optimize it on each network. In addition, the resolution of the data and step size for cycle counting has great influence on calculation of the equivalent full load cycles. This requires further investigations to standardize the mentioned parameters to obtain the equivalent results with the real physical phenomenon. To make the results comparable, the interruptions in real-data have not been considered in the calculations.

These interruptions could be caused by failure of the logging device, maintenance, or a failure in the network, which is unknown to us. In the case the cause of interruptions was due to system shutdowns, it should be considered in the cycle calculations and usually it has a considerable effect on the system.

Despite the higher accuracy in the supply pipes, less discrepancy has been detected in the full load cycle of the return pipes. In contrast, the number of full load cycles in the networks B, C, and D are almost double than the simulated values (see Table 2).

The calculated full load cycles with the $m = 4$ and $\Delta T_{ref} = 110$ °C were less than expected, which might be justified by the recent development of operation optimization technics. Since, these values are going to be calculated for each pipe in the network, these values will anyway internally be scaled and normalized and presumably, it will not affect the end goal, which is the failure prediction of the grid components.

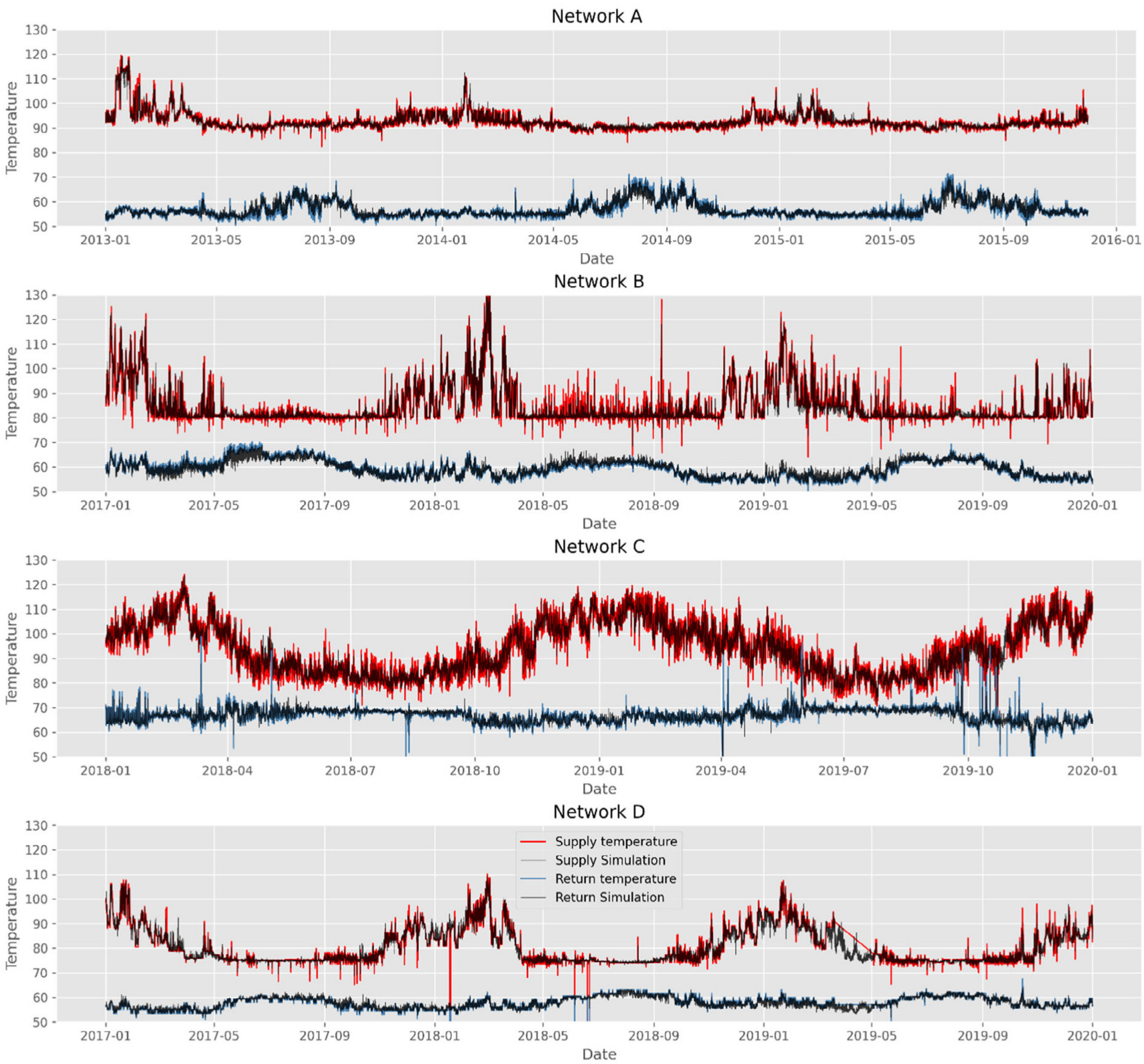

Fig. 2. Comparison of network sub-station temperature changes between real-world data and simulations in four networks.

Table 2. Equivalent full load cycles in four networks.

	Supply		Return	
	Real-world	Simulation	Real-world	Simulation
Network A	0.03	0.03	0.01	0.01
Network B	0.76	0.44	0.004	0.006
Network C	0.44	0.20	0.24	0.06
Network D	0.10	0.06	0.001	0.001

4. Conclusion

Backward-simulation enables to revive the loading history on pipes. The loading history is at importance in PdM approach to understand the ageing behaviour of the DH pipes. This method facilitates the formation of a combined ageing model and bond the both deterministic and statistical approaches together. The available current logged data from the grid operators are limited to the past few years and it is not sufficient to see the entire loading history

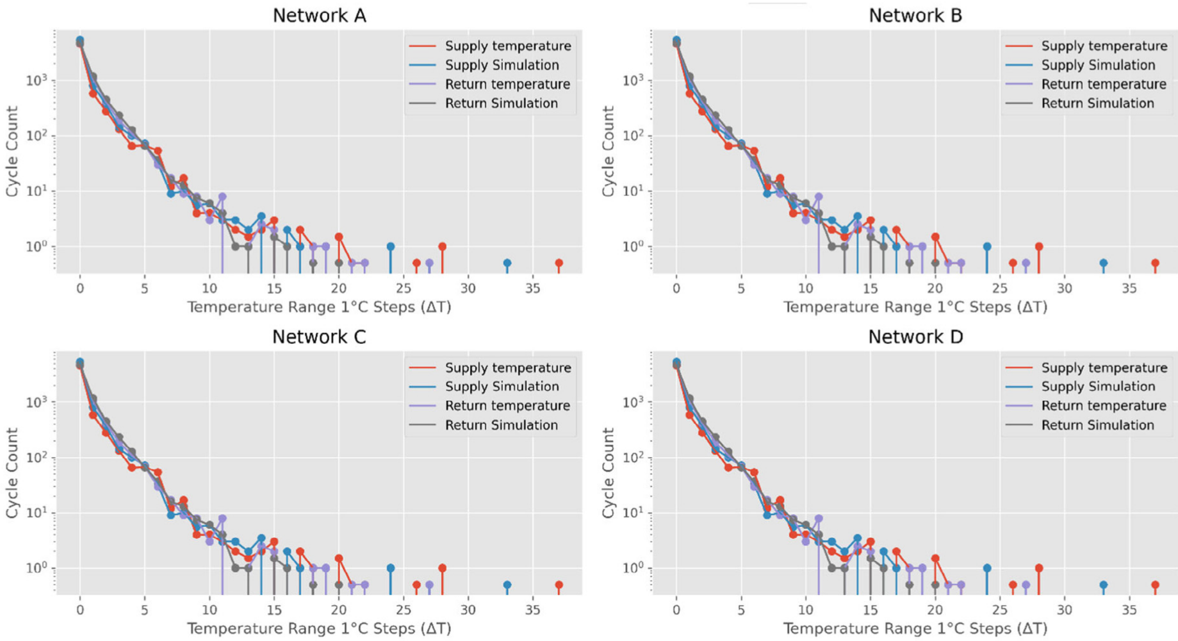

Fig. 3. Comparison of cycle count for real and simulated data in four networks.

on pipes, which is an essential element for PdM. Therefore, backward-simulation aid to fill the data scarcity with artificial data points. This method enables to overcome the unavailability of the network historical temperature data. The historical temperature could be used for estimation of remaining useful life of the PUR foam and fatigue analysis of the steel medium pipe in pre-insulated bounded pipes. Simulations ran for four different DH networks in Germany. The Random Forest regressor showed an outstanding performance in predictions among variety of algorithms that have been tested. The accuracy of predictions were always higher in supply, rather than return pipes. No significant cycles have been detected in return pipes. The number of full load cycles in three networks were almost double as the simulation values. Further investigations are required for finding the resolution effect in time-series and setting proper parameters for cycle counting.

CRediT authorship contribution statement

Pakdad Pourbozorgi Langroudi: Design, Conceptualization, Data curation, Formal analysis, Writing – original draft, Critical discussion of the results and editing of the manuscript. **Ingo Weidlich:** Supervision, Critical discussion of the results and editing of the manuscript. **Stefan Hay:** Practical experience and previous research results in the field of maintenance of DH networks, Critical discussion of the results and editing of the manuscript, Scientific supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors acknowledge the financial support by the Federal Ministry for Economic Affairs and Energy of Germany in the project Instandhaltung-FW (project number 03ET1625B).

Appendix. Number of cycles in 5 degree temperature ranges

See Table 3.

Table 3. Cycle comparison and temperature ranges. The empty cells means that no cycle detected in this range.

Data form	Mode	Minimum temperature (°C)	Maximum temperature (°C)	Number of cycles in 5 degree temperature ranges (°C)													
				0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	55–60	60–65	65–70
Network A																	
Real-world	Supply	82.28	119.57	5651.0	154.0	12.5	6.0	2.5	1.5	0.0	0.5						
Simulation	Supply	85.59	118.76	6739.5	126.0	19.5	6.5	0.0	1.5	0.5							
Real-world	Return	44.19	71.42	6543.0	132.0	14.5	4.0	1.0	0.5								
Simulation	Return	49.21	69.55	6846.5	126.5	11.0	3.0	0.5									
Network B																	
Real-world	Supply	62.83	130.77	5117.5	272.5	113.0	72.0	43.0	10.5	7.0	2.0	3.0	2.0	0.5	1.5	0.5	1.0
Simulation	Supply	68.47	130.18	6308.0	233.5	100.0	65.0	33.0	7.5	3.5	4.5	0.0	2.0	1.0	0.0	0.5	
Real-world	Return	49.66	70.18	4473.0	60.0	4.5	1.0	0.5									
Simulation	Return	52.55	68.79	6493.0	123.5	8.5	0.5										
Network C																	
Real-world	Supply	64.20	124.105	2737.5	570.5	275.5	116.0	35.0	14.0	5.5	0.0	1.0	1.0	0.0	1.0		
Simulation	Supply	70.54	123.65	3659.5	514.0	196.5	42.0	15.5	4.5	0.5	0.0	1.0	0.5	0.5			
Real-world	Return	35.61	99.04	2856.5	162.0	30.5	8.5	5.0	2.5	3.0	1.5	1.0	0.0	0.5	0.0	1.0	
Simulation	Return	43.82	89.37	4117.5	168.5	23.5	4.0	1.5	2.5	0.0	0.5	0.5	0.5				
Network D																	
Real-world	Supply	65.24	110.21	1103.5	140.5	46.0	15.5	6.0	2.0	0.5	0.0	2.5					
Simulation	Supply	68.61	109.35	7172.5	214.5	50.0	11.5	2.0	0.5	0.0	2.0	0.5					
Real-world	Return	49.42	67.51	670.0	14.5	1.5	0.5										
Simulation	Return	51.92	63.85	7373.5	25.0	1.5											

References

[1] Lund Henrik, Werner Sven, Wiltshire Robin, Svendsen Svend, Thorsen Jan Eric, Hvelplund Frede, Mathiesen Brian Vad. 4th Generation District Heating (4GDH). Energy 2014;68:1–11. http://dx.doi.org/10.1016/j.energy.2014.02.089, PII: S0360544214002369, In press.
[2] Selcuk S. Predictive maintenance, its implementation and latest trends. Proc Inst Mech Eng B 2017;231:1670–9.
[3] Meireles M, Almeida P, Simoes MG. A comprehensive review for industrial applicability of artificial neural networks. IEEE Trans Ind Electron 2003;50:585–601.
[4] Sharma Anupam. PyTorch on Azure: Deep learning in the oil and gas industry. Azure Blog and Updates. Microsoft Azure. 2019, https://azure.microsoft.com/en-us/blog/pytorch-on-azure-deep-learning-in-the-oil-and-gas-industry/?cdn=disable. [Accessed 24 March 2021].
[5] Murakami Y. The rainflow method in fatigue: The Tatsuo Endo memorial volume. Burlington: Elsevier Science; 1992, 251 pp.
[6] EN 13941-1:2019. District heating pipes – Design and installation of thermal insulated bonded single and twin pipe systems for directly buried hot water networks – Part 1: Design; German and English version EN 13941-1:2019. Beuth Verlag GmbH, Berlin.
[7] Christensen R. Fatigue analysis of district heating systems. Novem; 1999.