Transient Modelling and Simulation for Optimal Future Management of a District Heating Network.

Olamilekan E. TIJANI [1*], Sylvain SERRA[1], Sabine SOCHARD[1], Hugo VIOT[2], Aurélien HENON[2], Rachid MALTI[3], Patrick LANUSSE[3], Jean-Michel RENEAUME[1].

[1]Universite de Pau et des Pays de l'Adour, E2S UPPA, LaTEP, Pau, France.
[2]Nobatek/INEF4, 67 rue de Mirambeau, F-64600 Anglet, France.
[3]Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France.
[*]*(Corresponding author: tijani.oe@etud.univ-pau.fr)*

Abstract - The share of fossil fuel in some District Heating Networks (DHN) makes Dynamic Real-Time Optimisation (DRTO) techniques paramount in improving the energy efficiency of such networks. This research project is a fraction of the RESEAUDATA project, which aims to improve the heat network management technique through machine learning and dynamic optimisation approaches. This paper then focuses on simulating the DHN to accomplish the optimal planning phase for the DRTO and to generate the Big Data to form and train the machine learning model. The procedures for designing a Blackbox model generator are also established in this paper.

Index and exponent
k- instance or observation, k = 1, 2, … n_inst
n_inst- number of instances
ot- an output variable,
in- an input feature variable
inp- total number of input feature variable
outp- total number of output variables

1 Introduction
1.1 Background Information

50% of the world's energy consumption is used for heating, representing 40% of the global CO_2 emission[1]. These statistics show the need and the effects of heat generation. According to [1], 50% of this thermal energy is used in the industry. The other 50% is used for space heating, heating buildings, hot water supplies in buildings, and in the agricultural sector. A district heating system can efficiently generate and distribute thermal (heat) energy from the energy generation source to various end users. An inefficient DHN would require the excess generation of energy to meet up the power demands of the consumers, which will increase the cost of energy generation, reduce the life span of the generation components, and increase CO_2 and other greenhouse gases emission (if fossil fuel resources are the energy sources). This illustrates the necessity of an effective district heating distribution and production system and gives a solid backup for optimising such a system. Therefore, this research focuses on the Dynamic Real-Time Optimisation (DRTO) of a district heating system because it allows real-time control and operation for design or operation improvement purposes.

The algorithm of the dynamic real-time optimisation of any system relies on swift computational time for model resolution. The models that can be developed to simulate any process parameter in a district heating network can only be achieved using the knowledge-based energy and mass balance general principles. However, the simulation of these models requires high computational time and supercomputers for large-scale district heating systems [2]. The simulation time of these models can be drastically decreased by finding the equivalent Machine Learning (ML) models, which can be integrated into the dynamic real-time optimisation algorithms. Also, the lack of sufficient data as a ML model training challenge can be abridged thanks to the availability of big data from an industrial partner of this research.

1.2 Objectives of the RESEAUDATA Project

As a branch of the RESEAUDATA Project, this research paper considers the first two sections of the outlined pathways shown in Figure 1; model development and simulation of DHN and the preliminary steps in designing a Blackbox Model Generator explained in section 2 and section 3, respectively. This technique can also be applied to other energy generation and distribution sectors.

https://doi.org/10.25855/SFT2023-122

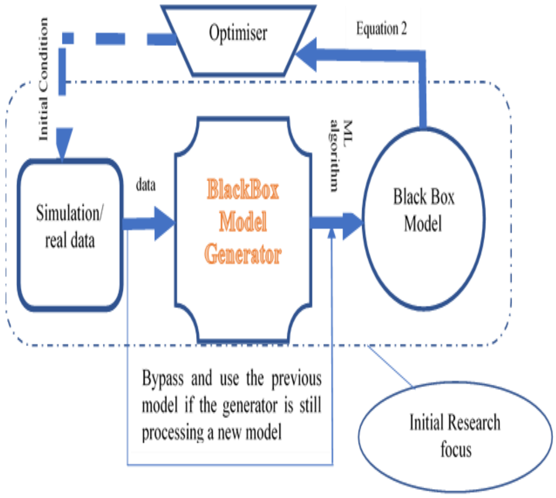

Figure 1: *pathways for the optimal management of a District Heating Network (DHN) considering real-time optimisation.*

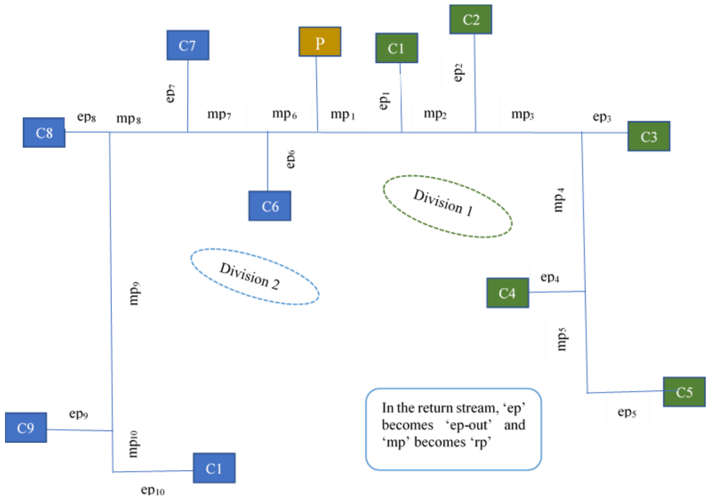

Figure 2*: A DHN network of 10 consumers.*

Several researchers have worked vastly on the thermal and hydraulic models of DHNs [2], [3] and [4], so the models are pretty classical. The developed models are based on the general mass and energy balance principles around the nodes (splitter and mixer), heat transfer fluid (HTF) transporting pipes, consumer's substations and around the thermal energy generation station, and they are therefore adapted to simulate the case study presented in Figure 2. The simulation results are shown in section 2.

2 Result Analysis

The result analysis is based on the given data and the analysis of the simulated result

2.1 Given Data

The given data for the DHN simulation are the pipe dimensions shown in Table 1, the meteorological data (ground surface temperature), the consumers' power demand shown in Figure 3, and the thermophysical properties of the heat transfer fluid (water) and that of the pipe materials.

Based on the assumptions made by previous researchers in modelling DHNs, the thermophysical properties of water and that of the insulation and protective material are assumed to be constant within the range of operation of the district heating Network. The property values include: specific heat capacity, density, and dynamic viscosity of water are 4187 J/(kg.K)), 965.3 (kg/m^3) and 0.000315 kg/(m.s) respectively, the thermal conductivity of water, pipe, insulation material and soil are 0.65, 54,0.024, and 1.2 W/(m.K), and the depth of buried pipes in the ground is 1m.

2.2 Supply Temperature and Mass flow rates

In an adiabatic DHN, these values of supply temperature should be equal to the outlet temperature of the combined thermal generation plant $T_{0,pmp(t)}$ = 90ºC. However, due to heat losses along the pipe length in the time domain, Figure 4 and Figure 5 depict that the supply temperature at each of the consumers reduces as a function of the overall heat transfer coefficient, which is dependent on the pipe diameter, the insulation thickness and the thermal conductivity of the insulation material. The supply temperatures also vary with time as the ground surface temperatures and the mass flow rates (function of the power demands) vary. Therefore, the trajectories of the supply temperatures in the DHN reflect the changes in the power demand, meteorological conditions and the thermophysical properties of the heat transfer fluid and the pipe material.

The mass flow rate in each extended pipe of a consumer is a function of the power demand of the consumer. Therefore, its trajectory varies in the same manner as the power demand, as shown in

Figure 6. This is quite reasonable in the explanatory physics of the DHN. Suppose the supply temperatures dynamically change based on the weather condition and power demand. In that case, the changes in the mass flow rates should be proportional to that of the power demands to ensure the power demands of each consumer are met.

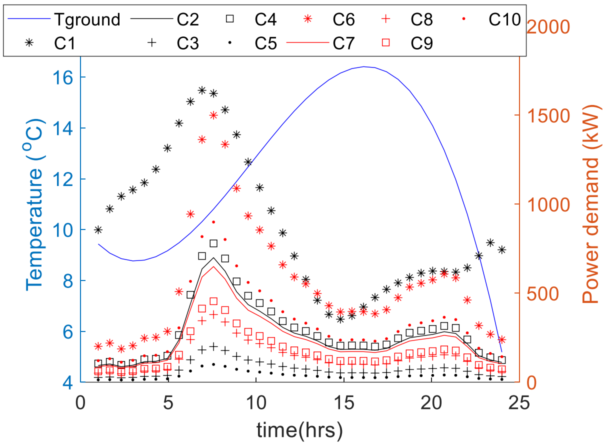

Figure 3: *Power demands and surface ground temperature*

Consumer	Φ_{mp}(in)	L_{mp}(m)	Φ_{ep}(in)	L_{ep}(m)
C1	8.00	279.93	5.00	95.21
C2	6.00	720.07	4.00	150.00
C3	5.00	176.46	2.00	40.00
C4	5.00	124.69	3.50	30.00
C5	1.50	397.32	1.50	100.00
C6	10.00	190.01	5.00	25.00
C7	8.00	97.04	3.50	10.00
C8	6.00	268.79	3.00	90.20
C9	5.00	1280.60	3.00	200.00
C10	5.00	198.28	4.00	50.00
P1-10	12.00	200.00		

Table 1: *pipe dimensions with thickness and insulation thickness of 0.01m and 0.053m, respectively*

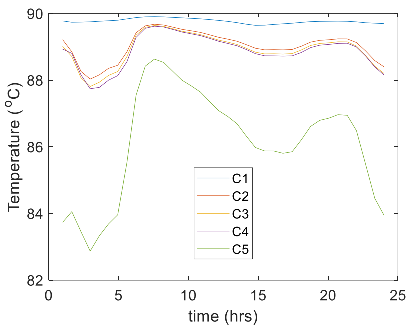

Figure 4: *supply temperatures in division 1*

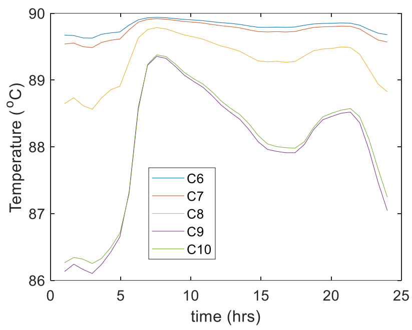

Figure 5: *supply temperatures in division 2*

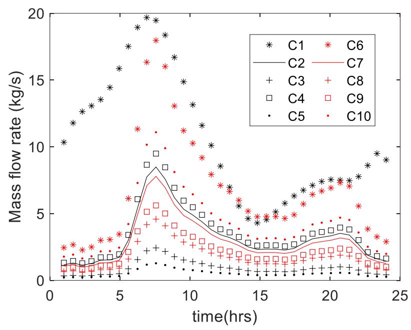

Figure 6: *Supply Mass flow rates to the consumers*

2.3 Overall Network temperature during the forward and return flow distribution

Figure 7 and Figure 8 represent the dynamic spatial temperature distribution through the main and return pipes, excluding the production pipes during the forward and return flow distribution, respectively. The configuration of these network temperature plots is similar to the main and return pipes arrangement in Figure 2. The maximum temperature in Figure 7 illustrates the outlet temperature of the heat transfer fluid from the production pipe. The temperature decreases towards the positive and negative x-axis. This implies that the longer the network, the more heat losses to the surroundings. However, concluding the temperature chronology during the return flow distribution is quite complex. Due to the mixing of the heat transfer fluid from a preceding consumer with the heat transfer fluid from the extended-out pipe of this consumer (the outlet temperature of each consumer's substation is assumed to be 70°C), a decrease or an increase in the fluid temperature can be observed. This explains the non-uniform temperature distribution along the network length during the return flow distribution.

2.4 Power Generation and the Overall Return Temperature

The dynamic power demands of each consumer can be theoretically validated using the time profile supply mass flow rates values in Figure 6, and the transient supply temperature values in Figure 4 and Figure 5. The question that might catch up with this study is what amount of power must be generated in the combined thermal plant to meet up the total demand of the network. The total generated power during the simulation of the DHN is always greater than the total power demand due to the overall network heat losses that must be compensated for, as shown in Figure 9. The variation of the overall return temperature is also affected by the same parameters that change the supply temperatures. Therefore, the dynamic of the overall return temperature of the network is majorly influenced by meteorological conditions and power demands.

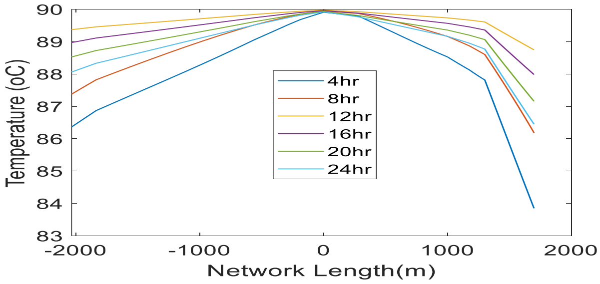

Figure 7: *forward flow transient temperature variation along network length.*

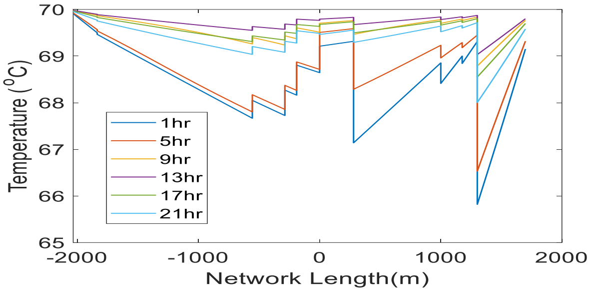

Figure 8: *return flow transient temperature variation along network length.*

2.5 Hydraulic Result Analysis

The primary purpose of the hydraulic analysis is to determine the dynamic variation of the pumping power required in divisions, which depends on the pressure drop in the divisions. One of the independent parameters of the pressure drop calculation is the transient fluid velocity, and its constraints must be respected to prevent pipe erosion [2], [4].

Although a single pump or pump based on the number of divisions can be used in a DHN, this research considers two pumps based on the number of divisions in the case study.

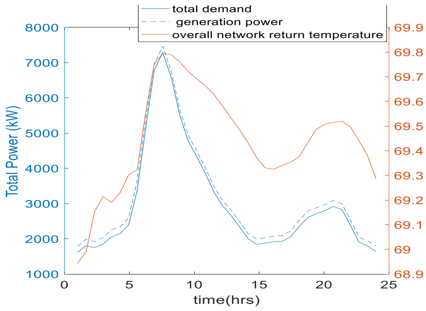

Figure 9: *total generated power, total demand and the overall network return temperature*

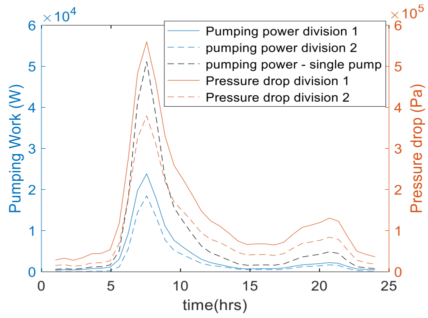

Figure 10: *Pressure drop and pumping power of the pumps*

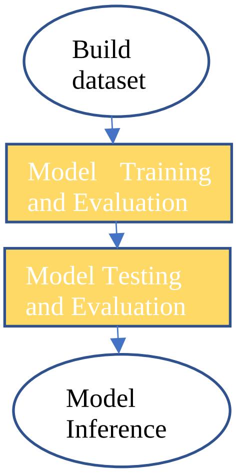

Figure 11: steps in designing a ML black box model

Figure 10 depicts the variation of the pressure drops and pumping powers with time, the pumping powers and pressure drops majorly depend on the mass flow rate profiles. Therefore, their time profiles will be similar in shape to that of the total power demand in each division. Thus, increasing demand for all the consumers will increase the mass flow rates, leading to high pressure drop in the division and thereby requiring pumps with huge pumping powers to compensate for the pressure drop in the system.

3 Designing a Blackbox Model Generator

The aim of designing a Blackbox Model Generator is to formulate a reduced and fast computing model that can predict the output variables (supply temperatures and mass flow rates, final network return temperature, spatial overall network temperatures, pressure drops and pumping powers), using the input feature variables which include the time series features (power demands, outlet generation temperature, outlet substation temperatures), meteorological conditions (ambient or soil temperature) and physical network parameters (pipe dimensions, thermophysical properties of network heat transfer fluid and pipes). The simulation of the case study DHN over 24 hours took more than an hour using a Windows 10 DELL PC, 32Gb RAM, 12th Gen Intel(R) Core (TM) i7-12700H, 2.30 GHz

Processing speed, the Blackbox model to be generated must simulate the case study DHN in seconds if possible, for efficient DRTO of the network.

The steps in developing a model for resolving any Machine Learning Problem are similar and can be presented in Figure 11.

3.1 Definition of feature variables and problem analysis

Indeed, this research has developed thermal and hydraulic models for predicting the output variables. Therefore, it is easy to build the dataset for training the Blackbox Model generator. This kind of dataset has output variables for all the input feature variables. In machine learning techniques, the best ML channel for solving such a problem is the ***Supervised Learning technique*** due to the presence of labels (output results) for all data input. Since the datasets are continuous, the ML techniques with ***Regression Analysis*** will be considered [4]. This is in contrast to the other ML channel counterparts, Non-supervised learning; there is no available output label for the input feature variables, and it encourages methods such as clustering for dataset grouping [4] and Reinforcement Learning, where a ML agent learns based on experience. Its learning algorithm encourages maximising total cumulative rewards [4].

3.2 Generating Dataset

Building the dataset involves collating results from simulating a DHN of different values of the feature variables to generate as many observations (k) as possible. A typical form of the compiled dataset can be visualised in Table 2.

3.3 Model Training

This is the main section for developing a Blackbox model where the entire dataset is initially split into the training and test datasets. Before diving into the details, let us analyse the model relationship based on the dataset.

$$y_k^{ot} = f\left(x_k^1, x_k^2, \dots, x_k^{inp}, c\right) \quad \forall\ ot = 1,2, \dots, out \qquad (1)$$

Equation (1 shows the generic form of the Blackbox model that must be developed, c- vector of constants that must be obtained during the training algorithm.

In developing a Blackbox model, the selection of essential feature variables and the training algorithm are the two important characteristics that must be carefully analysed [5] – [7]. For this reason, this research section will review previous literature to examine the best ML training algorithm and the principle of essential feature variable selection. Considering all the feature variables might increase the computational time in some training algorithms [6], which might nullify the importance of using ML models, so it is imperative to select the essential feature variables if possible.

As the dataset is regressive and the machine learning task is supervised, the ML algorithms that can be employed include Linear Regression (LR) or Multiple Linear Regression (MLR), Support Vector Regression (SVR), Artificial Neural Network (ANN), Recurrent Neural Network (RNN) or (Long Short-Term Memory (LSTM) network), Random Regression Forest (RF), Gradient Boosting or Extreme gradient boosting (XGBoost), and ensemble models [5] – [7]. Table 2 gives the operational principle of each algorithm, their validated research inferences and limitations. To the test of our knowledge, no research paper has been published on designing a Blackbox model generator for the DRTO of a DHN to eliminate the impractical ML techniques based on the validated results of the past study. Nevertheless, the non-linearity of a DHN thermal and hydraulic model depicts that using LR or MLR algorithm is impossible.

A grey box model based on optimal parametric identification can also be considered along with the Blackbox ML Algorithms.

3.4 Model Evaluation Metrics

Model Evaluation is a process of validating model performance. Some ML experts consider using the training dataset for the model evaluation to check the effectiveness of both the model and the training phase. However, [7] considered splitting the dataset into three (3) sections rather than the traditional two divisions, which include 60%, 20% and 20% for training, evaluation and testing. The work of [7] intended to choose the best model through cross-validation using the evaluation dataset before testing the model with the test dataset.

According to [5] – [7], the evaluation metrics for the regression supervised learning technique include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared (RMSE), Mean Absolute Percentage Error (MAPE), Coefficient of determination (R^2) and Coefficient of Variation (CV). In developing a Blackbox model, several training algorithm models can be obtained and compared with one another, the model with the least cumulative error will be considered the optimum Blackbox model [7] for the model generator.

3.5 Model Testing

The obtained models can then be tested with the test dataset to be considered for further analysis or can be tested after a cross-validation analysis. The model algorithm with the lowest bias and variance is the best algorithm to generate the black box model. It should be noted that evaluation metrics are also employed to evaluate the model effectiveness or variance based on the test dataset.

3.6 Model Inference

The obtained model (1), with the estimated parameters, is then used to predict each output variable at instance k.

$$y_k^{ot} = f\left(x_k^1, x_k^2, \dots, x_k^{inp}\right) \;\; \forall\, ot = 1,2, \dots, out \tag{2}$$

Equation (2) can predict each output variable at instance k given values of the input variables at instance k. It can then be considered as the generated ***Blackbox model*** by the ***Model generator.***

	x_k^{in} (input feature variables)					y_k^{ot}(output variables)			
k	$P_{demand(t)_i}$	$T_{0,ep\_out_i(t)}$	$T_{0,pmp(t)}$	T_{mn}	dimensions & thermo-ppts	$\dot{m}_{(t)_{ep_i}}$	$T_{n,ep_i(t)}$	W_{pump_DHN}	others

Table 2: *a sample of the Generated dataset for building the Blackbox Model Generator.*

Algorithm	Principle	Validated Inferences	Limitations
LR or MLR	• used to find a linear relationship between each output variable and several predictor input variables	• simple and quick during model inference[6], [7]. • quite good for weather prediction	• high bias and high variance when predicting nonlinear relationships • low precision [6]
SVR	• it employs kernel functions to change the dimensional space of the instances to a new space where LR can be employed [5]–[7].	• reasonably low bias and variance when predicting nonlinear relationships [6]. • Owing to its accuracy, it is recommended for Short Time Load Forecast (STLF) [6], [8].	• Better predicting aggregated energy consumption than individual building energy consumption with reasonably low bias and variance when predicting nonlinear relationships [6], [9].
ANN	• Based on human learning techniques to train a machine learning agent.	• Owing to the intensive mathematical mapping of information among very	• It is computationally intensive when all the

RNN or LSTM	• LSTM uses backpropagation to work on sequential data, as it also combines traditional neural networks [6]	• LSTM makes better predictions than RF, SVR, XGBoost and LR during STLF [6].	• Inefficient in a Long time load forecasting [6].
RF	• Based on repetitive bagging of the dataset and creation of decision trees [6], [7]. In contrast to the random forest, the regressive random forest uses a range of data to classify decision tree leaves. • The depth of the trees influences model performance [7].	**Advantages** • Owing to its ability to randomly bootstrap datasets and develop multiple trees, it has resistance to overfit and is insensitive to outliers [6], [7] • It produces the lowest error when compared to the non-linear autoregressive model (NARM) and linear model stepwise regression (LMSR) during energy prediction[6], [10]. • It can be employed for essential feature selection [5], [6].	**Disadvantage** • Suffers tremendously from covariate shift [11]. • Ineffective in real-time prediction at large tree numbers [11]
XGBoost	• In contrast to RF, which builds multiple trees in parallel, it makes trees sequentially [7]. • An ensemble algorithm that is based on gradient boosting to convert ineffective ML agents to effective learned ML agents [6], [7] • The depth of the trees influences the model performance. Therefore, stepwise searching technique can be used to determine the best tree depth [7].	**Advantages** • High prediction accuracy and speed [6]. • Has excellent resistance to overfitting due to the integrated regularisation technique [6]. • It outperformed SVM, RF and LSTM during energy prediction [6], [7]. • It can handle energy prediction in HVAC systems [6].	**Disadvantages** • XGBoost with deeper trees can have high variability [12].
Ensemble Model	• Combination of several algorithms [5], [6]	**Advantages** High prediction accuracy by leveraging the accuracy of each algorithm [6]	

Table 3: *ML Algorithms for formulating a data-driven model*

4 Conclusion

This research focused on the pathway for the Dynamic Real-Time Optimisation of District Heating Networks for their efficient management in real-time operation.

A case study simulation of a ten (10) consumer District Heating Network was initially considered, and the MATLAB simulation algorithm ran for more than 1hour on a standard ***Dell*** laptop. Practical DRTO approaches require such simulation to run within seconds because the input parameter of the model (meteorological conditions and consumer power demands) changes as such. To achieve this objective of the optimal operation of DHN, this research primarily simulated DHN with a combined generation station. Then it proposed the method for building big datasets from the simulation results to train an equivalent ML model.

Having confirmed the Supervised Regression learning technique as the best ML channel for designing a Blackbox model generator to facilitate the optimal operation of DHNs, this research reviewed the principle, validated inferences and limitations of LR or MLR, SVR, ANN, RNN or LSTM, RF, XGBoost, and ensemble models as the possible ML algorithms for the training phase. MAE, MSE, RMSE, MAPE, R^2 and CV were reviewed as the evaluation metrics for comparing the ML algorithms.

Reference

[1] IEA Report, “International Energy Agency, Renewables 2019,” 2019, Accessed: Aug. 18, 2022. [Online]. Available: www.iea.org/renewables2019

[2] M. Betancourt Schwarz, B. Lacarrière, C. Santos Silva, P. Alexandre Haurant, P. Haurant, and M. Tahar Mabrouk, “Dynamic Modeling of Heat Transport in District Heating Networks,” 2018. [Online]. Available: https://www.researchgate.net/publication/333774780

[3] M. Vallée, L. Giraud, R. Bavière, C. Paulus, and J.-F. Robin, “Dynamic Modelling, Experimental Validation and Simulation of a Virtual,” 2015. [Online]. Available: https://www.researchgate.net/publication/277007243

[4] S. Dridi, “Supervised Learning-A Systematic Literature Review SUPERVISED LEARNING-A SYSTEMATIC LITERATURE REVIEW A PREPRINT Supervised Learning-A Systematic Literature Review A PREPRINT,” 2021, doi: 10.13140/RG.2.2.34445.67049.

[5] Y. Sun, F. Haghighat, and B. C. M. Fung, “A review of the-state-of-the-art in data-driven approaches for building energy prediction,” *Energy and Buildings*, vol. 221. Elsevier Ltd, Aug. 15, 2020. doi: 10.1016/j.enbuild.2020.110022.

[6] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, “Physical energy and data-driven models in building energy prediction: A review,” *Energy Reports*, vol. 8. Elsevier Ltd, pp. 2656–2671, Nov. 01, 2022. doi: 10.1016/j.egyr.2022.01.162.

[7] A. A. Ahmed Gassar, G. Y. Yun, and S. Kim, “Data-driven approach to prediction of residential energy consumption at urban scales in London,” *Energy*, vol. 187, Nov. 2019, doi: 10.1016/j.energy.2019.115973.

[8] Y. Yang, J. Che, C. Deng, and L. Li, “Sequential grid approach based support vector regression for short-term electric load forecasting,” *Appl Energy*, vol. 238, pp. 1010–1021, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.127.

[9] P. Vrablecová, A. Bou Ezzeddine, V. Rozinajová, S. Šárik, and A. K. Sangaiah, “Smart grid load forecasting using online support vector regression,” *Computers and Electrical Engineering*, vol. 65, pp. 102–117, Jan. 2018, doi: 10.1016/j.compeleceng.2017.07.006.

[10] T. Ahmad and H. Chen, “Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems,” *Sustain Cities Soc*, vol. 45, pp. 460–473, Feb. 2019, doi: 10.1016/j.scs.2018.12.013.

[11] “A limitation of Random Forest Regression | by Ben Thompson | Towards Data Science.” https://towardsdatascience.com/a-limitation-of-random-forest-regression-db8ed7419e9f (accessed Feb. 14, 2023).

Acknowledgement

This research was performed within the framework of the RESEAUDATA project, which was funded by the Regional Council of Nouvelle Aquitaine in France.