

RESEARCH

Open Access



Gossen's first law in the modeling for demand side management: a thorough heat pump case study with deep learning based partial time series data generation

Chang Li^{1*}, Gina Brecher¹, Jovana Kovačević¹, Hüseyin K. Çakmak¹, Kevin Förderer¹, Jörg Matthes¹ and Veit Hagenmeyer¹

*Correspondence:
chang.li@kit.edu

¹ Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen 76344, Germany
Full list of author information is available at the end of the article

Abstract

Gossen's First Law describes the law of diminishing marginal utility. This paper aims to further verify the proposed hypothesis that Gossen's First Law also holds in the modeling for Demand Side Management (DSM) with a thorough heat pump case study. The proposed hypothesis states that in general the complexity-utility relationship in the field of DSM modeling could be represented by a diminishing marginal utility curve. On the other hand, in data based modeling, when utilizing a large dataset for validation, the data integrity is critical to the reliability of the results. However, the absence of partial time series data may occur during the measurement due to missing sensors or IT related issues. In this work, an extensive real-world open dataset of a ground source heat pump is utilized for the case study. In the raw data, one key variable namely the flow rate is missing. Thus, three different algorithms based on machine learning and deep learning architectures namely Random Forest (RF), Long Short-Term Memory (LSTM) and Transformer are applied to predict the flow rate by utilizing an open loop forecasting. The raw data are first pre-processed with a time interval of one hour and then used for training, validation and forecast. Furthermore, a modified persistence model as the baseline is also defined. The predicted flow rate using LSTM yields the lowest error of 7.47% nMAE and 10.56% nRMSE respectively. The forecast results are then utilized in the following step of modeling of a heat pump use case. With the introduced quantification method for complexity and a modified version for utility, we further verify the proposed hypothesis with a longer time horizon of 7 days.

Keywords: Ground source heat pump (GSHP), Deep learning, Open loop forecasting, Demand side management (DSM)

Introduction

In the course of the energy transition, residential buildings contain increasingly more electricity-related components such as heat pumps for heating and supplying domestic hot water, roof solar systems for self-generating electricity, home charging stations for

e-mobility and even fuel cell systems as home power plants (Thomas et al. 2020). These components, called as Distributed Energy Resources (DERs), provide the possibility of adapting the electricity load to electricity production, also known as Demand Side Management (DSM) (Energie-Agentur 2016). However, when modeling such components and their synergies for DSM, it is often unclear how detailed models need to be for different DSM applications since there is always an interaction effect between the utility and the complexity of a model. A complex model can usually provide more meaningful results than a simple model, but the effort required to modeling it increases accordingly. Therefore, it's necessary to investigate the relationship between the utility and complexity of a model in order to provide a quantified reference for different DSM applications. Mainly inspired by Gossen's First Law in economics and by research results in other modeling applications, e.g., in Building Information Modeling (BIM) (McArthur 2015), a novel approach and hypothesis was proposed by integrating Gossen's First Law into DSM modeling based on a first ground source heat pump study in Li et al. (2024). The proposed hypothesis states that in general the complexity-utility relationship in the field of DSM modeling could be represented by a diminishing marginal utility curve, thereby shedding light on the quantified relationship between model complexity and utility. However, there are two major limitations in this first study (Li et al. 2024), the first is that only one day, i.e., 24 h in February has been selected for validation, which could limit the robustness and generalizability of the proposed hypothesis, since different days might have different patterns. Secondly, potential applications of the findings, especially in real-world scenarios, should be discussed and summarized in more detail.

In order to tackle the mentioned limitations, it's necessary to select a larger real-world dataset with a longer time span for validation, where more temporal impacts throughout the time will be captured. However, the absence of some time series data may occur during the measurement. Before utilizing the data in modeling and analysis, it is important to generate or forecast the absence data as accurate as possible. This is especially inevitable when these data are crucial for decision making. For data generation or forecast, there have been several methods and approaches in literature such as single imputation (Zhang 2016) and machine learning approaches (Emmanuel et al. 2021). More details will be discussed in Sect. 2.

The main contribution of this paper is to further verify the proposed hypothesis in the previous work Li et al. (2024) with a longer time horizon of 7 days and more, i.e., 5 model classes. The proposed hypothesis in Li et al. (2024) is only validated for 24 h based on a Ground Source Heat Pump (GSHP) in a stand-alone house with 4 model classes as the preliminary work. In the present work, the heat pump modeling and thorough hypothesis validation are carried out based on an extensive real-time updated database from Switzerland [Meyer (<https://www.effiziente-waermepumpe.ch/messdaten/index.php>)], where historical raw data such as supply and return temperatures, thermal power and electrical power with a time interval of 15 min are extracted for the years 2021 and 2022. However, one key variable for modeling, namely the flow rate, is missing in the raw data. To tackle this problem as a necessary pre-step for the following model classification, utility comparison and validation, different machine learning (Random Forest) and deep learning (Long Short-Term Memory, Transformer) based approaches in partial time series data forecast with the modified persistence model as the baseline are utilized

and compared in the present work. The raw data are first pre-processed into time series data on hourly resolution. Then the data in January and February 2021 are selected for training with cross-validation and generation based on the frequency of zeros in the pre-processed data. A time horizon of 168 h, i.e., 7 days is determined for the time series data forecast and generation. By utilizing the descriptive statistics, i.e., nRMSE and nMAE, the accuracy of different approaches is compared and the best results for this use case are selected for the following step of heat pump modeling and simulation. With the generated data, the quantified relationship between model complexity and utility are illustrated with a longer time span and therefore the hypothesis is further explored. In addition, it's worth noting that the term data generation refers to an open loop forecasting in this context, which are interchangeably used throughout the text. The workflow of the present work is summarized in Fig. 1.

The remainder of the paper is divided into the following four parts. Section 2 presents related work on modeling of demand response or DSM technologies as well as in different approaches for data generation or forecast and proposes the selected approaches in this work. In Sect. 3, a brief description of each algorithm used for data generation is given. Besides, the methods and ideas for quantifying complexity and utility are also introduced. Section 4 describes the selected ground source heat pump system for the large real-world dataset and then the raw data together with the results of the pre-processing are presented. Section 5 presents, analyses and discusses the results of data

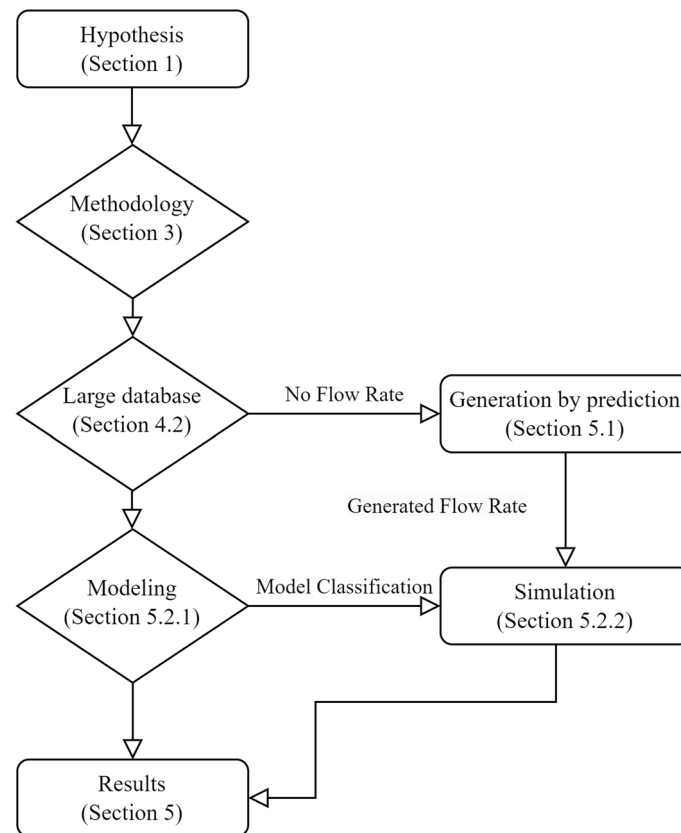


Fig. 1 Workflow of the present work

generation and heat pump simulation. Finally, the main conclusions of this work are highlighted in Sect. 6.

Related work

In recent years, there has been an increasing amount of literature on modeling of demand response or DSM technologies. For instance, in Turitsyn et al. (2011) a modeling framework for 4 types of individual devices which are expected to participate in future demand-response markets are introduced. The purpose is to pursue their optimal price-taking control strategy under a given stochastic situation. The models are differentiated into 4 types which are optimal and generic. Therefore, modeling of specific systems and synergies between different systems are not investigated. In 2013, a more generic taxonomy for modeling flexibility in Smart Grids are defined in Petersen et al. (2013), which divided all systems into three categories and used them to optimize and solve flexibility problems in Smart Grids. This type of modeling approach simplifies the modeling process and improves optimization efficiency. However, the challenges of considering different influencing factors in real energy systems such as temperature are not solved since the models are too abstract. For this reason, the models are hard to be directly applied to real energy systems on the demand side.

In contrast, Keeling and Butcher (2013) Peralta et al. (2021) Śliwa and Gonet (2005) used very detailed theoretical models and complex numerical techniques such as Lax-Wendroff finite difference approximations for a specified system, i.e., heat pump and its subsystems. These models are capable of delivering accurate results, however, yield very high complexity and low performance, meaning more computing resources and measurements are required, which limits the optimization efficiency as well as practical operations. This will limit the practical application in the field of DSM. In summary, we conclude that models of varying degree of complexity have different utilities, as mentioned in Sect. 1. However, there is no, to the best knowledge of the authors, straightforward investigation of the effect of model complexity on model utility in DSM. Hence, it's necessary to investigate the relationship between the utility and complexity of a model in order to provide a better reference for different DSM applications.

Moreover, dealing with partially missing data in modeling when utilizing large datasets for validation, has been an important topic, not only in engineering but also in other fields such as medicine for a long time. In order to address this problem more accurately and reliably, different approaches, from the common statistical techniques to machine learning based methods in recent years, are explored based on different use cases in many publications. In Zhang (2016), the implementation of R code to perform single imputation of missing data such as mean, median and mode imputations is conducted. However, no quantified results are summarized in the article. The authors in Austin et al. (2021) have developed a model based on Multiple Imputation (MI) to create imputed data and proven that the created values by using MI are plausible in their use case. Another new technique, which is a hybrid approach of single and multiple imputation techniques, is proposed in Khan and Hoque (2020) in two variations to impute categorical and numeric data. The experimental results show that the proposed algorithm achieves around 20% higher F-measure for binary data imputation and around 11% improvement in terms of

error reduction for numeric data. To handle the nonlinear associations between the variables in multilevel models, a flexible sequential approach based on Bayesian estimation techniques is proposed in Grund et al. (2021), which outperforms the conventional MI methods for multilevel models with nonlinear effects. In Weber et al. (2021), the authors have introduced a new Copy-Paste Imputation (CPI) method for imputing energy and power time series. The method takes into account the total energy of each gap and outperforms the selected three benchmark imputation methods in their work.

In addition to using statistical methods to reconstruct missing data, machine learning imputation methods are also widely used for imputation of missing data. For instance, the authors in Jerez et al. (2010) compare the performance of machine learning based techniques such as multi-layer perceptron (MLP) and k-nearest neighbor (KNN) with statistical techniques such as MI. The results reveal that the machine learning techniques lead to a significant enhancement of accuracy compared to statistical procedures. Similarly, eight statistical and machine learning imputation methods are compared based on real data and predictive models in Li et al. (2024). The most effective results are attained by KNN and Random Forest (RF). In the survey paper Emmanuel et al. (2021), the authors aggregate different imputation methods, particularly focusing on machine learning techniques. They evaluate the performance of KNN and missForest, which is an iterative method based on RF, by utilizing a power plant fan dataset. The results are promising for future research direction. Besides the common machine learning techniques, deep learning methods are also explored for dealing with missing data such as Long Short-Term Memory (LSTM). In Tian et al. (2018), a new model named as LSTM-M is proposed for managing missing data in the traffic flow, which outperforms several other methods such as Support Vector Regression (SVR) in terms of accuracy. Likewise, the authors in Ma et al. (2020) propose a LSTM-BIT model, which is a hybrid LSTM model with Bi-directional Imputation and Transfer Learning (BIT). The results show that the proposed model achieves a 4.24% to 47.15% RMSE under different missing rates.

Moreover, since Transformer was proposed in 2017 Vaswani et al. (2017), the exploration about applications based on its architecture is still ongoing. The huge success of this architecture in natural language processing (NLP) and computer vision (CV) motivates the exploration of its other potential such as handling time series data (Hertel et al. 2023). However, there have been very few works that focus on utilizing Transformer for handling data generation. Based on the related work above, three different approaches are selected for data generation in this work, namely RE, LSTM and Transformer. Furthermore, we propose a modified persistence model as the baseline for a better quantitative comparison and discussion.

Methodology

In this section, the algorithms for forecasting the flow rate in the heat pump modeling are first presented, including a modified persistence model as the baseline. And to visualize the relationship between complexity and utility, the method for quantification of complexity and utility are then discussed.

Prediction algorithms and modified persistence model

In the present work, three different algorithms are chosen for forecasting as mentioned in Sect. 2. In this subsection, each of them is briefly described. Besides, the definition of our modified persistence model as the baseline is also included in this subsection.

Random forest (RF)

As an ensemble learning method for classification and regression problems (Breiman 2001), RF has been widely used in many classification and regression problems. When dealing with data generation, it also shows promising results as stated in Emmanuel et al. (2021) Li et al. (2024). When the data is presented through time series, it requires transforming the time series dataset into a supervised learning problem first. Figure 2 shows this transformation process, i.e., sliding window, with an input size of one as an example, where Y is the value at each time step. However, there is a limitation of this method that cannot be ignored, i.e., random forest cannot extrapolate. It means that predicted values are always within the range of the training set. In this work, different input sizes are tested to find an ideal parameter. Finally, we create a bagged regression ensemble object with an input size of 5 together with the temporal features of days such as Monday, Tuesday etc. as the 6th input, to use bootstrap aggregation method for model training, since there are no significant improvements with further increased input sizes.

Long short-term memory (LSTM)

For predicting data based on time series while avoiding the vanishing gradient problem, LSTM has been developed as a modified version of traditional RNN. By introducing the so-called gates, LSTM can regulate the flow of information and maintain valuable information. In comparison to other RNN, LSTM can deal with large amounts of data and time steps more easily (Zhu et al. 2019). Besides, it's also powerful when managing missing data as presented in Tian et al. (2018) Ma et al. (2020). Based on these advantages, it's been chosen as one of the algorithms in the paper.

Transformer

For all RNNs, one major limitation is that the computations must be performed in the sequence's order, which makes parallel computation difficult and thus limits the efficiency when dealing with long sequences. The proposed Transformer architecture in Vaswani et al. (2017), which relies on the self-attention and multi-head attention mechanism, solved this limitation, making it more efficient than RNNs. While there is still debate about the advantages of Transformer in time series as remarked in Wen et al. (2022), the consideration and introduction of this new architecture to deal with time series data generation is worthwhile.

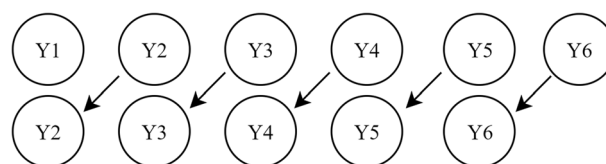


Fig. 2 Transformation of time series into a supervised learning problem with input size of one

Modified persistence model

The persistence model (Notton and Voyant 2018) is often used as a trivial reference model when different forecast models are compared. In this work, a modified version of the persistence model is defined by considering the temporal impacts. Instead of generating the future value by assuming that no changes happen between the current time step and next time step, we use the values a week ago of the same time period, i.e., same days in the week as presented in Fig. 3.

Method for quantification and visualization

The proposed hypothesis uses a diminishing marginal utility curve to represent the complexity-utility relationship in the field of DSM modeling. As with Gossen's First Law, the marginal utility itself is an inherently abstract concept and needs to be quantified first, such as income (Layard et al. 2008), in order to illustrate its relationship with consumption or other properties. Similarly, the method for quantifying the complexity and utility of DSM modeling is also crucial to visualize the interaction between them. This subsection discusses separately what kinds of quantitative options for complexity and utility are available and then explains those that have been chosen in the present work.

Quantification of complexity

In computer sciences complexity is measured in various ways, such as required time, number of operations, required memory and Big O notation. They do depend on the specific algorithms, their implementation, and the hardware they are running on. For instance, Big O notation is often used to classify the efficiency or complexity of algorithms according to how their run time grows as the input variable increases. However, for modeling we need other measures. Different from computational complexity theory or information theory, this work focuses on the modeling of physical structures and dynamic processes of energy components in DSM applications. Thus, an appropriate method in our scenario should help to understand the practical complexity of different models such as the measurement setup and how the model works, thereby promoting transparency and reliability in their practical application. Furthermore, an appropriate method for quantification applicable for all possible system components is required.

In Bao et al. (2014a, 2014b); Jiang and Wang (2012) different time scales are used in energy systems of different complexity. In the process of modeling, if transient processes within a system are non-decisive, we could neglect the details and use larger time scale to simplify the whole process. However, this option cannot differentiate the complexity of the same model because different time scales can also be chosen during the simulation for the same model.

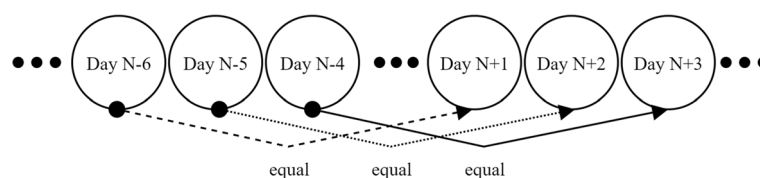


Fig. 3 Modified persistence model

Besides choosing different time scales, another option to quantify complexity would be by the power range that can cover the range from milliwatt (mW) to gigawatt (GW). Different power ranges would have an impact on dynamic responses of the model, leading to more complex model and corresponding controls (De Brito et al. 2011). However, the limitations of this option are also significant because the power range is generally determined for a given energy system. Therefore, the power range of a model cannot always be artificially changed to quantify its complexity.

A third way of quantifying complexity could be based on the number of required parameters in models. On a structural basis, any model is a combination of different input and output parameters. Furthermore, for the same model, the number of parameters could be adjusted according to the study objectives or experimental conditions, so that models of different complexity can be built.

Among the three methods mentioned above, the third method has the best applicability and feasibility. Besides, it aids in a better understanding of how the models work. Based on that, the included parameters of a model, i.e., the number of required parameters, has been chosen to quantify the complexity in our work.

Quantification of utility

The main goal of DSM applications is to improve the flexibility of a power system (Energie-Agentur 2016). In this context, the methods for the quantification of utility are as same as those for quantifying flexibility in DSM applications. In Péan et al. (2019) four typical ways for quantifying flexibility in DSM, namely load-shifting, peak shaving, reduction of energy use and valley filling, are explained and summarized. In De Coninck and Helsen (2016) two more specific approaches, namely daily primary energy use and daily energy costs are used to show the improved and quantified flexibility.

In addition, it is worth noting that the accuracy of a model must first be verified through offline simulations before the model is used to analyze flexibility in DSM applications. Models with high predictive and simulation accuracy can assist grid operators or DSM participants in optimizing recourse allocation, reducing unnecessary energy waste and effectively lowering operational costs (Panda et al. 2022), thereby improving the overall efficiency and profitability of DSM applications. According to ISO 5725-1, the general term “accuracy” describes the closeness of a measurement to the true value. Based on this definition, we can quantitatively describe the accuracy of a model with the help of some useful metrics in descriptive statistics such as normalized Root-Mean-Square Error (nRMSE) and normalized Mean-Absolute-Error (nMAE).

$$\text{nRMSE}(\hat{Y}) = \frac{\text{nRMSE}(\hat{Y})}{\max(Y_a) - \min(Y_a)} \quad (1)$$

$$\text{nMAE}(\hat{Y}) = \frac{\text{nMAE}(\hat{Y})}{\max(Y_a) - \min(Y_a)} \quad (2)$$

One focus of this work is on the accuracy of different models in an offline simulation and uses quantified accuracy to represent utility of models. In order to reduce the impact of absolute values on the accuracy analysis, two descriptive statistics namely nRMSE and

nMAE are defined in (1) and (2), where \hat{Y} is the generated or simulated value and Y_a is the ground truth.

Measurement system and data preprocessing

In this section, the overview and setup of the selected ground source heat pump system (GSHP) [Meyer (<https://www.effiziente-waermepumpe.ch/messdaten/index.php>)], which measures and stores the real-world dataset, is briefly described first. After that, the structure of the raw data is presented. In the second part, discussion of the necessary data preprocessing for the generation of flow rate is carried out.

Measurement system

The selected system uses a GSHP together with a smaller hot water tank for the domestic hot water supply and a larger hot water tank for the house heating. Figure 4 shows the schematic heat matrix of the overall heating system along with different positions of installed temperature sensors. It shows that 4 temperature sensors are installed at different layers in the large heating storage tank and 3 sensors are placed for the smaller one with the equal distance. This layout leads to the modification of the thermal model of heat pump storage, which will be discussed in Sect. 6.

Data preprocessing: generation of flow rate

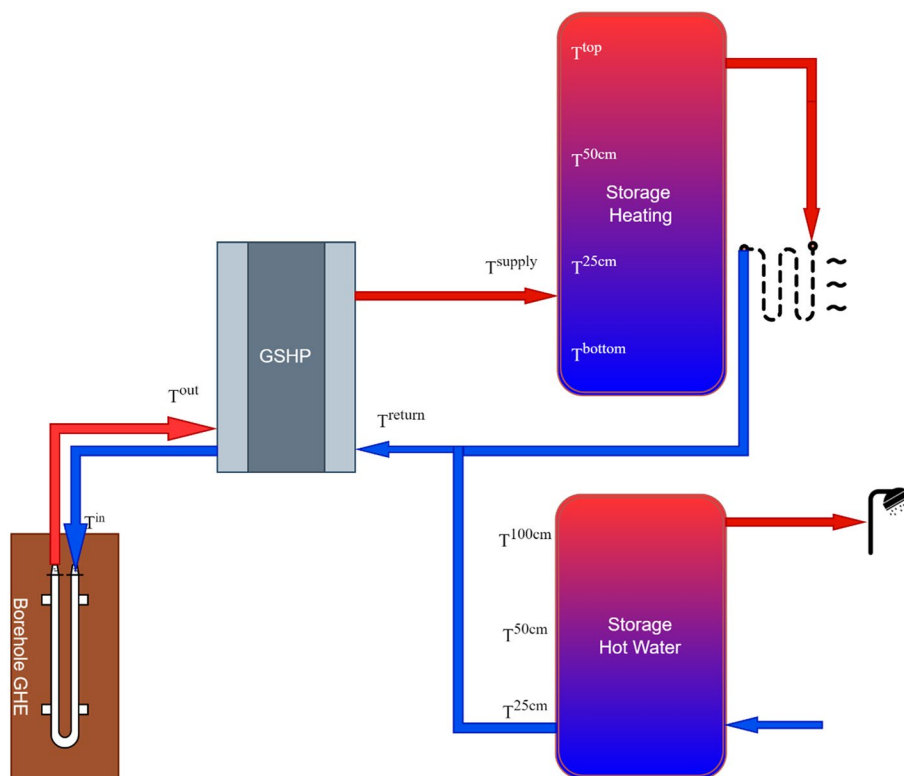
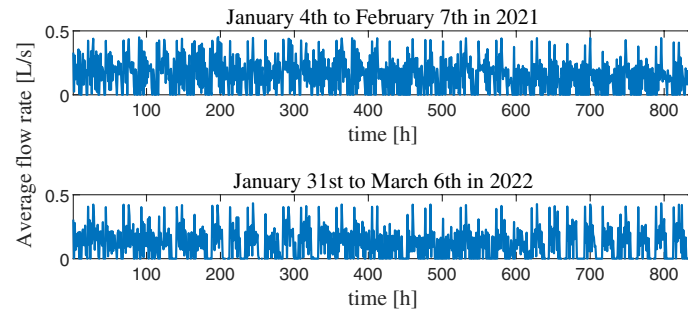


Fig. 4 Schematic heat matrix consisting of positions of installed temperature sensors

Table 1 Excerpt from the raw data

Date and time	T^{supply} [°C]	T^{return} [°C]	P_Q [W]	P [W]	COP
2021-01-02 03:00:00	42.4	34.6	13898	3440	4.04
2021-01-02 03:15:00	51.9	44.3	13494	4040	3.34
2021-01-02 03:30:00	58.3	51.1	12775	4612	2.77
...

**Fig. 5** Average flow rate during the selected time period in 2021 and 2022

The real-time updated databank has an update interval of 30 s to 60 s according to Meyer (<https://www.effiziente-waermepumpe.ch/messdaten/index.php>). In this work, the historical raw data with a time interval of 15 min are extracted for the years 2021 and 2022. Due to the space limitation, Table 1 shows an excerpt from the extracted raw data, where T^{supply} and T^{return} are the supply and return temperature of the heat pump respectively. The coefficient of performance (COP) presents heat pump's overall performance, which is defined as the ratio of P_Q and P , where P_Q is the thermal power and P is the consumed electrical power. However, one key variable is missing in the raw data, which is the flow rate, i.e., \dot{V}_w in (3), where c_w is the specific heat capacity of water and ρ_w is the density of water. This variable is used for calculating the thermal power and thus needs to be generated first for the following comparison and simulation.

$$P_Q = c_w \cdot \dot{V}_w \cdot \rho_w \cdot (T^{supply} - T^{return}) \quad (3)$$

According to the date and time, the raw data are pre-processed into time series data by hour at first. Besides, it's assumed that the thermal power and the electrical power are constant throughout each time interval. Moreover, it's worth noting that the thermal power will be equal zero when the heat pump is turned off, which means that the frequency of zeros in the pre-processed data should be as small as possible to avoid the case of sparse data. Based on these three conditions mentioned above, the data from January 4th to February 7th in 2021 and from January 31st to March 6th in 2022 are selected for the calculation of the average flow rate by hour. Each time period starts on Monday and ends on Sunday. The reason for choosing another month in 2022 is that several days of data are completely missing in January.

Figure 5 shows the results of calculated flow rate of the selected 5 weeks in 2021 and 2022. The frequency of zeros of the selected time period in 2021 and 2022 are 23.57%

and 32.38% respectively. It shows that the data in 2021 are less sparser than the data in 2022. Therefore, the chosen time period in 2021 will be determined for the following work.

Results and discussion

In this section, the results of the predicted flow rate by using different algorithms are first given and compared. In the following step, different model classes are defined based on the complexity, i.e., the number of required parameters. By utilizing the generated flow rate, the simulation results are then presented along with the discussion.

Flow rate generation results

As mentioned in Sect. 4, the selected time period in 2021 contains 5 weeks. The calculated flow rate in the first 4 weeks is used as training set with cross-validation. The subsequent week, namely a time horizon of 7 days, serves as the ground truth for the generated data. Different from predicting multiple subsequent time steps in a closed loop forecasting, we use an open loop forecasting for generating the data at the next time step. It means that for subsequent time steps, the true value, which is the calculated flow rate in our case, is collected until the last time step and used as input.

Compared to a conventional approach, which is to create forecast models for each measured variable namely the thermal power, the supply and return temperatures in (3) and then to use the predicted values to calculate the flow rate, the proposed pre-processing approach is more straightforward and less complex. The proposed approach calculates the flow rate in the past explicitly and only needs to create a forecast model for the flow rate directly.

To optimize the forecast results of each method, we have tuned the hyperparameters in different approaches separately, where the hyperparameters for RF are automatically optimized in MATLAB and the tuned hyperparameter settings for LSTM and Transformer in PyTorch are shown in Table 2. It's worth noting that hyperparameters such as Epoch and number of layers in LSTM and Transformer, which have a significant impact on the complexity and the run time of both approaches, are set to be the same in order to ensure that the complexity of both methods does not differ too much within the range of tuned values.

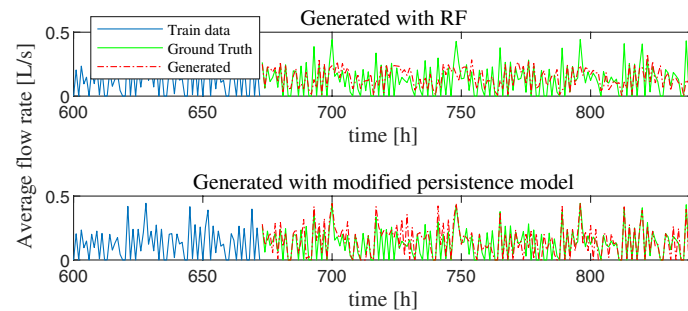
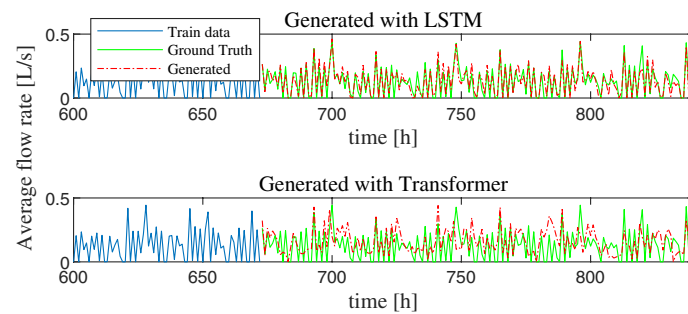
Two descriptive statistics, as described in Sect 3.2.2, are summarized in Table 3. The detailed plots are presented in Figs. 6 and 7. It should be noted that not all training data

Table 2 Hyperparameter setting for LSTM and Transformer

LSTM		Transformer	
Input size	1	Num heads	16
Hidden size	256	Hidden dimension	512
Hidden layer	1	Num layers	1
Batch size	128	Batch size	4
Epoch	200	Epoch	200
Learning rate	0.01	Learning rate	0.001
Optimizer	Adam	Regularization	Dropout = 0.1

Table 3 Summary of descriptive statistics for each algorithm

	Persistence model	RF	LSTM	Transformer
nRMSE	22.08%	22.00%	10.56%	24.46%
nMAE	15.14%	16.26%	7.47%	19.73%

**Fig. 6** Generated average flow rate with RF and the modified persistence model**Fig. 7** Generated average flow rate with LSTM and Transformer

are plotted in order to better demonstrate the comparison between the ground truth and the generated data.

According to the results in Table 3, the minimum error of the generated data is given by LSTM with a nRMSE of 10.56% and a nMAE of 7.47%. On the other hand, the results of RF are no better than the baseline with the modified persistence model. This demonstrates the limitation of RF when dealing with sparse data, although the input size of RF is longer compared to LSTM. In addition, it should be noted that the summarized results represent the capability of each machine learning algorithm under the current tuned hyperparameter settings in this scenario. For the model classification and utility comparison in Sect. 5.2, the LSTM generated results with the smallest error will be utilized.

Modeling and simulation results

In this subsection, the heat pump models are first briefly modified and described based on the selected heat pump system in Meyer (<https://www.effiziente-waermepumpe.ch/messdaten/index.php>). Afterwards, different model classes based on the number of required parameters by combining different mathematical models are defined. Then, the defined

model classes are used to perform offline simulations of the load profile for the following analysis. Lastly, the subsection concludes with a discussion of the hypothesis mentioned in Sect. 1.

Modification and classification of the models

In Li et al. (2024), the modeling of the ground source heat pump is carried out based on three main subsystems for heat transfer, namely the thermal model of the borehole ground heat exchanger (GHE), the thermal model of the heat pump itself and the thermal model of the heat pump storage. However, due to the new structure of the selected system in the current work, it's necessary to modify the models. The heat transfer in the borehole GHE is unchanged modeled in (4) and (5), where T^{in} and T^{out} are the inlet and outlet temperature of the borehole GHE as shown in Fig. 4. c_b is the specific heat capacity of the brine and \dot{m}_b is the mass flow of the brine. Besides, P_Q^{abs} is the absorbed thermal power, which is also the difference between P_Q and P .

$$T^{out} = T^{in} + \frac{P_Q^{abs}}{c_b \cdot \dot{m}_b} \quad (4)$$

$$P_Q^{abs} = P_Q - P \quad (5)$$

To model the performance of the heat pump itself, one simple way is to calculate the COP directly with the measured thermal and electrical power over a period of time and obtain an average value as presented in (6). Moreover, the thermal power can be obtained as mentioned in (1).

$$\text{COP}^{avg} = \frac{1}{n} \sum_{t=1}^n \frac{P_{Q,t}}{P_t} \quad (6)$$

In this work, the system contains two different hot water tanks for different purposes as described in Sect. 3. As the central storage for thermal energy, the temperature and corresponding energy changes have a significant impact on the overall system. Therefore, it's necessary to consider the energy changes of the storage separately. In general, the thermal energy change in the storage between two successive time steps could be calculated in (7) under the assumption that the density and the specific heat capacity of hot water as constant. In (7), the V_s is the volume of the hot water tank and $(T_t^{mean} - T_{t-1}^{mean})$ donates the average temperature change of the hot water, which are determined in (8) and (9) for the small and the large storage respectively with the assumption that the temperature is evenly distributed in each layer at every time step.

$$\Delta Q_s = c_w \cdot V_s \cdot \rho_w \cdot (T_t^{mean} - T_{t-1}^{mean}) \quad (7)$$

$$T_t^{mean,s} = \frac{T_t^{25cm} + T_t^{50cm} + T_t^{100cm}}{3} \quad (8)$$

Table 4 Model classification with respect to parameters

Model class	Combination	Number of required parameters
Model A	(3)(4)(5)(7)(8)(9)	16
Model B	(3)(4)(5)(7)(9)	14
Model C	(3)(4)(5)(7)(8)	13
Model D	(3)(4)(5)	9
Model E	(3)(6)	7

Table 5 Overview of the applied parameters to each model class

Parameter	Model A	Model B	Model C	Model D	Model E
T^{supply}	■	■	■	■	■
T^{return}	■	■	■	■	■
\dot{V}_w	■	■	■	■	■
\dot{m}_b	■	■	■	■	■
T^{out}	■	■	■	■	■
T^{in}	■	■	■	■	■
T^{bottom}	■	■	■	■	■
T^{25cm}	■	■	■	■	■
T^{50cm}	■	■	■	■	■
T^{100cm}	■	■	■	■	■
T^{top}	■	■	■	■	■
ρ_w	■	■	■	■	■
c_w	■	■	■	■	■
c_b	■	■	■	■	■
V_s	■	■	■	■	■
V_l	■	■	■	■	■
P_Q	■	■	■	■	■
P	■	■	■	■	■

$$T_t^{mean,l} = \frac{T_t^{bottom} + T_t^{25cm} + T_t^{50cm} + T_t^{top}}{4} \tag{9}$$

Using the modified models, we introduce five different model classes (A, B, C, D and E) with decreasing complexity in terms of the number of required parameters. All model classes utilize (3) to calculate the thermal power with the generated average flow rate to further obtain the electrical power, while Model A considers the energy changes in both storages, Model B and Model C neglect the impact of the small and the large hot water tank respectively. Moreover, Model D is further simplified by ignoring the energy changes in both storage. The last model class directly uses the average COP to calculate the consumed electrical power. Table 4 presents the model classification and the number of required parameters and an overview of the individual parameters that apply to each model class is given in Table 5.

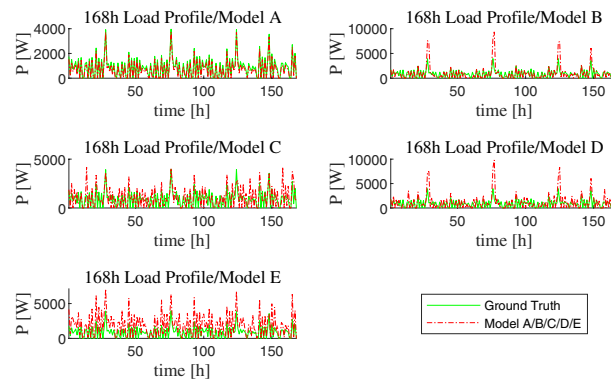


Fig. 8 Comparison between model results and measured results

Table 6 nMAE and Utility of each model class

	Model A	Model B	Model C	Model D	Model E
nMAE	3.77%	10.62%	16.99%	20.78%	26.79%
U	96.23%	89.38%	83.01%	79.22%	73.21%

Results and utility comparison

The quantification of the utility of the models is modified with the new definition in (10), where U represents the utility of a model in percentage. The reason to use nMAE instead of MAPE as described in Li et al. (2024) is that the ground truth contains zeros, which makes the calculation of MAPE not feasible.

$$U = (1 - (\text{nMAE})) \cdot 100[\%] \quad (10)$$

As mentioned in Sect. 1, a time horizon of 168 h is determined for the simulation and analysis. Besides, different from the initialization in the previous work (Li et al. 2024), the initial value of the consumed electrical power is calculated by utilizing the generated flow rate. Figure 8 shows the results of different models along with the differences between them and the ground truth.

The diagram shows that the results of Model A are the closest to the measured results, whereas Model B and Model D show several large deviations at some time steps as shown in some tips of the curve. What these two models have in common is that neither considers the energy changes in the small storage for domestic hot water. Therefore, one possible reason for this behavior is that the usage patterns of the domestic hot water are more dynamic than heating. In addition, the simplest Model E in our case presents a larger value than the ground truth in most cases, which could be caused by the underestimated average COP in (6), since COP is equal zero when the heat pump is turned off.

In order to describe the overall statistic features of the simulation results and the utility of the models as defined above, we calculate the nMAE and the corresponding U , yielding the results presented in Table 6. Model A, with the highest complexity in terms of the required parameters, has the lowest nMAE of 3.77% compared to

other four model classes and thus has the highest utility among all models. Besides, it's worth noting that Model B has a lower nMAE than that of Model C despite the large deviations at some time steps, which means the overall impact of the large hot water storage is greater than that of the small one.

With the definition in (10), the relationship between the utility and the complexity of all five model classes are illustrated in Fig. 9. This demonstrates that the results with a longer time horizon of 7 days are further verifying the proposed hypothesis in the previous work (Li et al. 2024), which is that the complexity-utility relationship in the field of DSM modeling could be represented by a diminishing marginal utility curve. However, it should be noted that the graph line is not as smooth as an approximated diminishing marginal utility curve by using a polynomial curve of degree 2, which is also presented in orange dashed line as a reference in Fig. 9. The deviation between the simulation and the approximation results, such as the data point of Model C, reveals that there could exist gaps between the simulation and an ideal value by approximation, which is reasonable.

Conclusion

This paper investigates thoroughly the proposed hypothesis of diminishing marginal utility in DSM modeling with a heat pump case study according to Gossen's First Law in economics. The simulation results are basically in line with the diminishing marginal utility curve and further verify our proposed hypothesis. In this process, a large real-world dataset with the predicted flow rate data is utilized as the input. To handle the problem of the absence of time series data in the dataset, we first utilize and compare three different machine learning algorithms together with our modified persistence model, which serves as the baseline. The results show that generation with LSTM delivers the smallest error, i.e., a nRMSE of 10.56% and a nMAE of 7.47%, by utilizing the open loop prediction as the generation method. With the generated flow rate, we then carry out the heat pump system modeling, model classification based on the complexity namely the number of required parameters and load profile simulation for a time horizon of 7 days with different patterns. Due to the zero values of electrical power in our dataset, we modify the definition of utility of models in the present work compared to Li et al. (2024) and then illustrate the relationship between the complexity and utility among all five classified model classes. With these findings, potential applications could be identified in real-world scenarios. For instance, if we have a pre-defined range

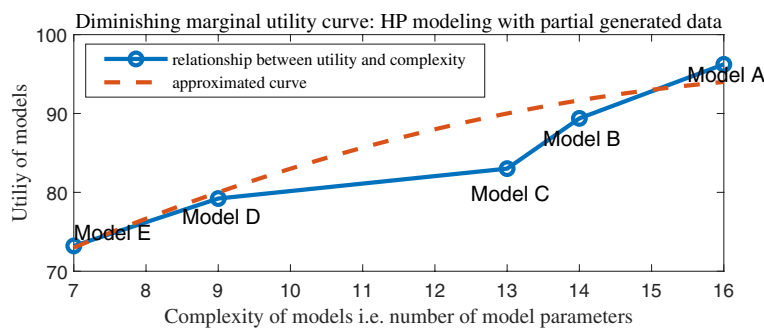


Fig. 9 Diminishing marginal utility curve based on the complexity of models

of acceptable error, we could use the curve to find a balanced modeling solution, which satisfies the error range and contains less complexity at the same time.

Author contributions

CL: Conceptualization, Methodology, Programming, Validation, Visualization and Writing of the original manuscript. GB: Preparation of the original dataset. JK: Review and Editing. HC: Review and Editing. KF, JM, VH: Funding acquisition, Supervision, Review and Editing.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Energy System Design (ESD) Program of the Helmholtz Association (HGF) within the structure 37.12.02..

Availability of data and materials

Original dataset is available in Meyer (<https://www.effiziente-waermepumpe.ch/messdaten/index.php>). Code is available on GitHub via <https://github.com/KIT-IAI/Gossen-Heat-Pump>.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no conflict of interest.

Received: 23 May 2024 Accepted: 19 June 2024

Published online: 24 June 2024

References

- Austin PC, White IR, Lee DS, Buuren S (2021) Missing data in clinical research: a tutorial on multiple imputation. *Can J Cardiol* 37(9):1322–1331
- Bao Z, Zhou Q, Yang Z, Yang Q, Xu L, Wu T (2014) A multi time-scale and multi energy-type coordinated microgrid scheduling solution part i: model and methodology. *IEEE Trans Power Syst* 30(5):2257–2266
- Bao Z, Zhou Q, Yang Z, Yang Q, Xu L, Wu T (2014) A multi time-scale and multi energy-type coordinated microgrid scheduling solution part ii: optimization algorithm and case studies. *IEEE Trans Power Syst* 30(5):2267–2277
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- De Brito MA, Sampaio LP, Luigi G, Melo GA, Canesin CA (2011) Comparative analysis of mppt techniques for pv applications. In: 2011 International Conference on Clean Electrical Power (ICCEP), pp. 99–104. IEEE
- De Coninck R, Helsen L (2016) Practical implementation and evaluation of model predictive control for an office building in brussels. *Energy Build* 111:290–298
- Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O (2021) A survey on missing data in machine learning. *J Big Data* 8:1–37
- Energie-Agentur D (2016) Studie: Roadmap demand side management
- Grund S, Lüdtke O, Robitzsch A (2021) Multiple imputation of missing data in multilevel models with the r package mdmb: a flexible sequential modeling approach. *Behav Res Methods* 53(6):2631–2649
- Hertel M, Beichter M, Heidrich B, Neumann O, Schäfer B, Mikut R, Hagenmeyer V (2023) Transformer training strategies for forecasting multiple load time series. *Energy Inf* 6(Suppl 1):20
- Jerez JM, Molina I, Garcia Laencina PJ, Alba E, Ribelles N, Martin M, Franco L (2010) Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif Intell Med* 50(2):105–115
- Jiang Q, Wang H (2012) Two-time-scale coordination control for a battery energy storage system to mitigate wind power fluctuations. *IEEE Trans Energy Convers* 28(1):52–61
- Keeling S, Butcher K (2013) Ground Source Heat Pumps. CIBSE
- Khan SI, Hoque ASML (2020) SICE: an improved missing data imputation technique. *J Big Data* 7(1):37
- Layard R, Mayraz G, Nickell S (2008) The marginal utility of income. *J Public Econ* 92(8–9):1846–1857
- Li J, Guo S, Ma R, He J, Zhang X, Rui D, Ding Y, Li Y, Jian L, Cheng J et al (2024) Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Med Res Methodol* 24(1):41
- Li C, Förderer K, Moser T, Spatafora L, Hagenmeyer V (2024) Gossen's first law in the modeling for demand side management: a first heat pump case study. In: Jørgensen BN, Silva LCP, Ma Z (eds) *Energy Informatics*. Springer, Cham, pp 111–125
- Ma J, Cheng JC, Jiang F, Chen W, Wang M, Zhai C (2020) A bi-directional missing data imputation scheme based on lstm and transfer learning for building energy data. *Energy Build* 216:109941
- McArthur J (2015) A building information management (bim) framework and supporting case study for existing building operations, maintenance and sustainability. *Proc Eng* 118:1104–1111

- Meyer J. Aktuelle Messwerte der Sole-Wasser Wärmepumpen Anlage. <https://www.effiziente-waermepumpe.ch/messdaten/index.php>
- Notton G, Voyant C (2018) Forecasting of intermittent solar energy resource 77–114
- Panda S, Mohanty S, Rout PK, Sahu BK, Bajaj M, Zawbaa HM, Kamel S (2022) Residential demand side management model, optimization and future perspective: a review. *Energy Rep* 8:3727–3766
- Péan TQ, Salom J, Costa-Castelló R (2019) Review of control strategies for improving the energy flexibility provided by heat pump systems in buildings. *J Process Control* 74:35–49
- Peralta D, Cañizares CA, Bhattacharya K (2021) Ground source heat pump modeling, operation, and participation in electricity markets. *IEEE Trans Smart Grid* 13(2):1126–1138
- Petersen MK, Edlund K, Hansen LH, Bendtsen J, Stoustrup J (2013) A taxonomy for modeling flexibility and a computationally efficient algorithm for dispatch in smart grids. In: 2013 American Control Conference, pp. 1150–1156. IEEE
- Šliwa T, Gonet A (2005) Theoretical model of borehole heat exchanger
- Thomas JM, Edwards PP, Dobson PJ, Owen GP (2020) Decarbonising energy: the developing international activity in hydrogen technologies and fuel cells. *J Energy Chem* 51:405–415
- Tian Y, Zhang K, Li J, Lin X, Yang B (2018) Lstm-based traffic flow prediction with missing data. *Neurocomputing* 318:297–305
- Turitsyn K, Backhaus S, Ananyev M, Chertkov M (2011) Smart finite state devices: A modeling framework for demand response technologies. In: 2011 50th IEEE Conference on Decision and Control and European Control Conference, pp. 7–14. IEEE
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Adv Neural Inf Proc Syst* 30
- Weber M, Turowski M, Çakmak HK, Mikut R, Kühnapfel U, Hagenmeyer V (2021) Data-driven copy-paste imputation for energy time series. *IEEE Trans Smart Grid* 12(6):5409–5419
- Wen Q, Zhou T, Zhang C, Chen W, Ma Z, Yan J, Sun L (2022) Transformers in time series: a survey. arXiv preprint [arXiv:2202.07125](https://arxiv.org/abs/2202.07125)
- Zhang Z (2016) Missing data imputation: focusing on single imputation. *Ann Transl Med* 4(1)
- Zhu J, Yang Z, Guo Y, Zhang J, Yang H (2019) Short-term load forecasting for electric vehicle charging stations based on deep learning approaches. *Appl Sci* 9(9):1723

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.