15TH IEA HEAT PUMP CONFERENCE
MAY 26–29, 2026 | HOFBURG
VIENNA 2026
DECARBONISATION THROUGH INNOVATION

Estimation of parameter values in heat pump systems using neural networks

Jan Furtwengler[a,*], Ana Constantin[a], Janis Keuper[b]

[a]Robert Bosch GmbH, Corporate Research, Robert-Bosch-Campus 1, 71272 Renningen, Germany
[b]Offenburg University of Applied Sciences, Institute for Machine Learning and Analytics, Badstraße 24, 77652 Offenburg, Germany

Abstract

Heat pump systems are a key component in the decarbonization of the building sector. However, efficient control of these systems requires technical parameters that are often unknown or undocumented in real-world installations. Relevant parameters include pipe length, pipe inner diameter, and buffer storage volume. This paper investigates the use of artificial neural networks (ANNs) to estimate these parameters based on time-series data from selected sensor signals. Several network architectures are examined, including Multi-Layer Perceptrons (MLPs), Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Generative Adversarial Networks (GANs). The models are trained on a simulation model of a building energy system comprising a heat pump, a buffer storage tank, a domestic hot water tank, and the relevant hydraulic components. By varying the pipe length, pipe inner diameter, and buffer storage volume, different system configurations are simulated. The resulting sensor time series serve as input to the networks, while the corresponding system parameters represent the target outputs. In the first study, the models' ability to estimate individual parameters is investigated. The CNN and RNN achieve the lowest mean errors, whereas the GAN exhibits the highest error values. In the second study, the simultaneous estimation of all three parameters is analyzed. Again, the CNN and RNN achieve the highest accuracy, with mean errors of 0.8 % for pipe length, 1.7 % for pipe inner diameter, and 0.9 % for buffer storage volume. The results highlight the potential of neural networks for data-driven identification of technical parameters in building energy systems. Future research should investigate the transferability of the proposed approach to real measurement data and explore methods for uncertainty quantification.

Keywords: Building energy system; heat pump; parameter identification; neural networks; time series analysis

1. Introduction

The building sector is responsible for approximately 30 % of global final energy consumption and accounts for about one quarter of energy-related CO_2 emissions [1]. Around half of these emissions are generated by space heating and domestic hot water [1]. Electrically driven heat pumps represent a promising alternative to fossil-based heating systems. They utilize environmental heat from air, ground, or groundwater and can be operated in a climate-friendly manner, particularly in power grids with a high share of renewable energy [2]. According to a scenario by the International Energy Agency, the number of installed heat pumps worldwide will rise to 600 million by 2030, more than tripling compared to 2020 [3].

With their increasing deployment, questions concerning the operational efficiency and system integration of heat pumps are gaining greater attention. In combination with thermal storages or photovoltaics, complex systems emerge that offer a wide range of control possibilities. Accordingly, the literature proposes various control strategies, ranging from conventional methods (e.g., PID controllers) to model-based optimization approaches (e.g., model predictive control), data-driven methods (e.g., reinforcement learning), and hybrid

* Corresponding author. Tel.: +49-711-811-14775
E-mail address: Jan.Furtwengler@de.bosch.com

approaches [4]. Regardless of their methodological differences, these approaches all rely — either explicitly or implicitly — on assumptions about the underlying system dynamics. These dynamics are fundamentally determined by physical system parameters, such as the sizing of piping networks, the capacity of thermal storage units, and the thermal properties of the building envelope. In practical installations, many of these parameters are unavailable or poorly documented. This lack of reliable parameter information directly limits achievable control performance and may reduce the operational efficiency of heat pump systems [5, 6, 7, 8]. Consequently, inverse methods are commonly employed to infer such unknown parameters from measured data [6]. Established inverse methods include trial-and-error approaches, optimization techniques, and stochastic methods such as the Bayes filter, Kalman filter, or Bayesian estimation [6]. These methods typically rely on analytical models that mathematically describe the physical behavior of the underlying components. However, constructing such models in practice is challenging, since building energy systems are characterized by high structural complexity and numerous interactions among subsystems [6, 9]. A promising alternative is the use of artificial neural networks [10]. Their training is not based on an explicit system model but on measurement or simulation data. Moreover, they can capture complex relationships between input and output variables. Their high performance, however, comes at the cost of limited transparency, which complicates the physical interpretation of the results [10]. Research on neural network applications in building energy systems has largely focused on inferring building-related parameters from time series data, such as the insulation level of the building envelope [6, 9]. In contrast, generation-side components remain comparatively underexplored. This part of the system typically includes the heat pump unit, buffer storages, and the hydraulic distribution network. Previous work has shown that the pipe length between the heat pump and the building can be successfully estimated from time series data using neural networks [11]. This paper extends that approach to additional generation-side parameters, specifically the pipe inner diameter and the buffer storage volume. The choice of these parameters is motivated by two practical considerations. First, their values are often unknown or poorly documented in real-world systems. Second, they have a significant influence on the dynamic behavior of the overall system and are therefore important for control strategy design. The central research question is whether a single model can jointly identify all three parameters or whether separate models are required for each. To address this question, various neural network architectures are evaluated, including Multi-Layer Perceptrons (MLP), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Generative Adversarial Networks (GAN). The models are trained on data generated by a simulation model of a building energy system including a heat pump, thermal storage, and the associated hydraulic network. This study constitutes a first step towards improved control and efficiency by focusing on parameter identification. The integration into control strategies and the resulting efficiency impact are left for future work.

Section 2 describes the investigated system and the applied methodology. Section 3 presents and discusses the results, followed by Section 4, which concludes the work with a summary and final remarks.

2. Methodology

2.1. System Description

The following section presents the building energy system under investigation. The main components and their interactions are described. Figure 1 shows a schematic representation of the system. It is divided into a generation side and a load side. The load side comprises the heating system and the domestic hot water (DHW) supply. On the generation side, there is a heat pump system consisting of an outdoor and an indoor unit.

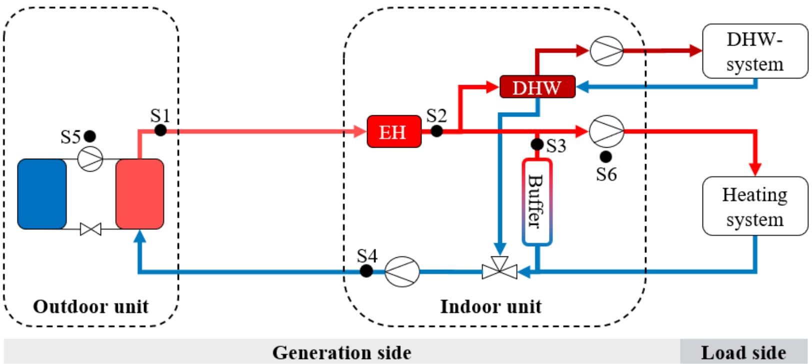

Figure 1. Schematic representation of the system

The outdoor unit is installed outside the building, while the indoor unit is located inside. The two units are hydraulically connected by a pipe with a length L typically ranging from $1\ \mathrm{m}$ to $30\ \mathrm{m}$. The standardized inner diameter D of the pipe can take discrete values of $16\ \mathrm{mm}$, $26\ \mathrm{mm}$, or $33\ \mathrm{mm}$, which are specific to the German market. The outdoor unit is an electrically driven air-to-water heat pump that extracts thermal energy from the ambient air. The indoor unit comprises a buffer storage tank, a domestic hot water tank, an electric heating element (EH), and additional hydraulic and control components. The DHW circuit supplies domestic hot water, while the buffer tank provides heat to the space heating system. Depending on the system configuration, the buffer storage volume V typically ranges from $50\ \mathrm{dm}^3$ to $400\ \mathrm{dm}^3$. The focus of this paper is the identification of the parameters L, D, and V.

In addition, the building energy system is equipped with several sensors for monitoring and control purposes. Figure 1 illustrates all relevant measurement points. Sensors S1–S3 measure the supply temperature at different locations in the hydraulic system, while sensor S4 records the return temperature. Sensor S5 monitors the compressor speed, and sensor S6 measures the rotational speed of the heating circuit pump.

2.2. Workflow

In this paper, an inverse method for parameter estimation is investigated. The objective is to determine the geometric system parameters — pipe length L, pipe inner diameter D, and buffer storage volume V — from time series data $y(t)$ (see Figure 2). To achieve this, the mapping $y(t) \rightarrow (L,\ D,\ V)$ is approximated using several neural network architectures. Time series data from sensors S1 to S6 are used as input (see Figure 1). The integration of data from all sensors yields a more comprehensive representation of the system's dynamics, thereby improving the network's ability to accurately infer the underlying parameters.

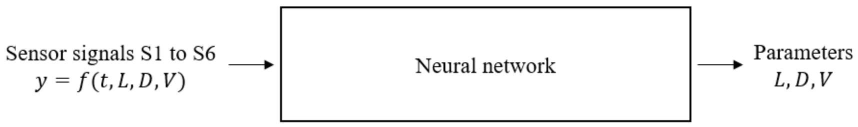

Figure 2. Input and output variables of the neural network

Two investigations are performed to assess the predictive performance of different neural network architectures. In Study 1, a one-dimensional prediction approach is applied, where a separate model is trained for each target parameter (L, D, V). This allows for an analysis of the models' ability to estimate individual parameters independently. In Study 2, a multidimensional prediction approach is examined, in which a single model estimates all three parameters simultaneously. This investigation enables an assessment of the model performance when multiple system parameters are estimated at the same time. Both investigations follow a uniform workflow, illustrated in Figure 3. The starting point is a simulation model that serves as a virtual test bench for a real building energy system. In this work, it is used to generate synthetic time series data (from sensors S1–S6). After appropriate preprocessing of the simulation data, these are made available for model development. For the creation of data-driven identification models, the datasets are divided into training, validation, and test sets. The individual steps of this process are described in detail in the following sections.

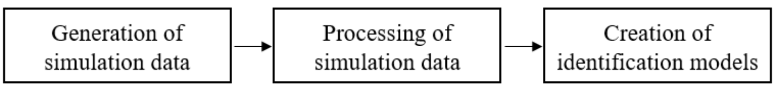

Figure 3. Overview of the workflow

2.3. Generation of Simulation Data

In this paper, a simulation model is employed to generate synthetic time series data. The model captures the dynamic behavior of the building energy system based on fundamental physical equations. It is implemented in MATLAB/Simulink and utilizes the Building Technology Simulation Library (BTSL), developed by the Bosch Home Comfort Group. The BTSL is an internal Simulink library comprising validated components for the dynamic simulation of thermal energy systems in buildings. It includes both elementary models, such as valves, pipes, and air ducts, and complex systems, such as heat pumps. Furthermore, the library provides functionalities for modeling various boundary conditions, including user behavior and weather data [12]. Figure 4 illustrates the input and output variables of the simulation model. For the generation of training, validation, and test data, the geometric system parameters — pipe length L, pipe inner diameter D, and buffer storage volume V — are treated as variable model parameters. The corresponding value ranges are provided in Section 2.1. It is assumed that the pipe is straight and does not include any bends.

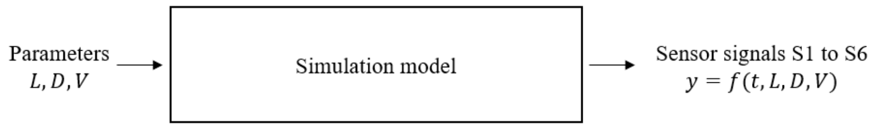

Figure 4. Input and output variables of the simulation model

The time-dependent sensor signals S1–S6 serve as the relevant output variables of the simulation. To accurately represent the sensor signals as a function of the system parameters L, D, and V, the simulation model accounts for the entire building energy system, including all components shown in Figure 1. The electric heating element (EH) is implemented in the model but remains deactivated during the simulation. For the heating system, a simplified model is used that accounts for the mass flow rate and the thermal inertia of the building. The model is controlled according to a predefined sequence of operating scenarios to ensure reproducible data generation. The procedure starts from a thermally balanced initial condition, which is achieved by pre-cooling the building. Subsequently, the heat pump system is activated. After the compressor startup phase is completed, a setpoint step of the supply temperature is applied to induce a dynamic system response. The resulting transient behavior then serves as the basis for identifying the system parameters L, D, and V.

2.4. Processing of Simulation Data

This section describes the individual steps for processing the simulation data, which are based on the simulation model explained in Section 2.3. First, the relevant observation period is defined, excluding all data points outside the interval $[t_{\min}, t_{\max}]$. Since the first-time step of the simulation model contains numerical initialization processes, $t_{\min} = 1$ is chosen. From this point onward, physically consistent sensor data are available. The upper time limit is set to $t_{\max} = 6000$. After this period, the system is largely in a steady-state condition, allowing the relevant dynamics to be fully captured. Subsequently, temporal downsampling of the simulation data is performed by considering only every second time step. The objective is to reduce the data volume and computational effort for the subsequent model development. In the next step, all input and output variables are standardized. This includes the time-resolved sensor signals (S1–S6) as well as the target parameters pipe length L, pipe inner diameter D, and buffer storage volume V. Standardization is carried out using z-score normalization, ensuring that all features have a mean of 0 and a standard deviation of 1. This procedure enhances numerical stability during training and mitigates the dominance of individual features arising from differing value ranges [13]. Figure 5 shows three processed time series generated by the simulation model. These examples serve to qualitatively evaluate the signal characteristics used for training the neural network models. The figure depicts the normalized supply temperature measured at sensor S2, located downstream of the pipe. Among the six available sensor signals (S1–S6), S2 was chosen as it most clearly represents the system's dynamic response and operational behavior.

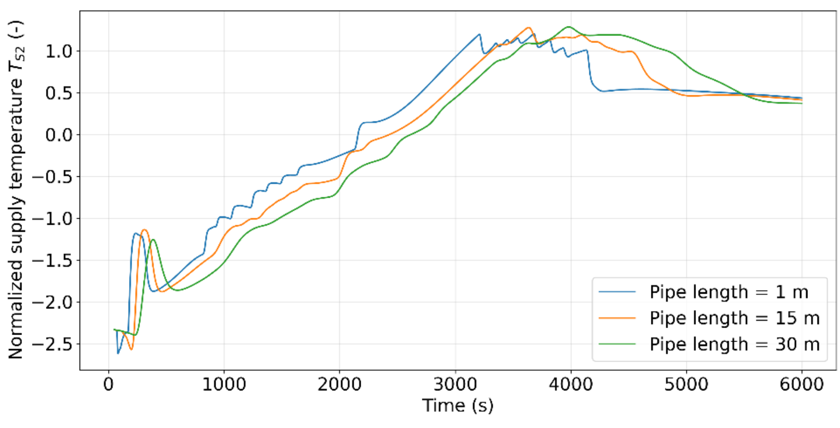

Figure 5. Simulation data after processing

The time series are presented for three pipe lengths: $L = 1$ m, 15 m, and 30 m. The remaining parameters are kept constant, with a buffer storage volume of $V = 200\ \mathrm{dm}^3$ and a pipe inner diameter of $D = 33$ mm. Similar temporal structures are observed for other values of D and V. As described in the previous section, all simulations start from a thermally balanced condition with uniform initial temperature. After the heat pump system is activated, an initialization phase occurs during which the compressor ramps up (approximately 500 s). Subsequently, the supply temperature T_{S2} increases continuously until the setpoint is reached. Due to the implemented PID control, a slight overshoot occurs before the temperature decreases again. Figure 5 illustrates that the pipe length has a significant impact on the dynamic behavior of the system. A short pipe (e.g., $L = 1$ m) exhibits a low thermal storage capacity, allowing temperature variations to propagate rapidly throughout the entire circuit. Consequently, each adjustment of the compressor speed induces rapid and pronounced temperature fluctuations, as evidenced by the sharp spikes observed in Figure 5. With increasing pipe length (e.g., $L = 15$ m and $L = 30$ m), the thermal inertia of the system increases. The larger pipe mass acts as a thermal buffer, damping temperature peaks and reducing the system's response speed. As a result, the temperature evolution becomes smoother, and the overall system responds more gradually to disturbances. Overall, a clear trend toward reduced dynamic behavior with increasing pipe length can be observed. The differences in temporal system dynamics evident in Figure 5 form the foundation for the subsequent parameter estimation, in which the neural network models aim to infer the parameters L, D, and V from such time-series data.

2.5. Creation of Identification Models

The models used in the two investigations are based on artificial neural networks. In the first study, a separate model is trained for each parameter, whereas in the second study, all three parameters are identified simultaneously using a single network. The following section introduces the network architectures employed for this purpose and describes their fundamental characteristics. Subsequently, the procedure for model development is explained. The analysis includes four model architectures that differ in structure and learning behavior: Multi-Layer Perceptron, Convolutional Neural Network, Recurrent Neural Network, and Generative Adversarial Network. These architectures are briefly characterized below to facilitate their classification. The descriptions follow the representations given in [14]. For more detailed information on their functionality and applications, the reader is referred to [15].

A Multi-Layer Perceptron [14] represents a fundamental architecture of artificial neural networks. As a feedforward network, it processes information only in one direction. It consists of an input layer, one or more hidden layers, and an output layer. All neurons in adjacent layers are fully connected. The MLP is suitable for tasks with a fixed input dimension and is commonly used for regression or classification problems. Typical activation functions include nonlinear functions such as the Rectified Linear Unit (ReLU) or the Hyperbolic Tangent (Tanh). The training phase is carried out via backpropagation in combination with optimization algorithms such as stochastic gradient descent or Adam. MLPs require the definition of numerous hyperparameters, such as the number of neurons, layers, and the learning rate.

Convolutional Neural Networks [14] were originally developed for image processing applications but are increasingly used for time series analysis. They also belong to the class of feedforward networks. The core component is the convolutional layer, in which local features are extracted from the input data. Pooling layers aggregate neighboring information, thereby reducing data dimensionality.

Recurrent Neural Networks [14] are specialized for processing sequential data. In contrast to feedforward networks, they include feedback connections within the network structure, allowing previous states to influence the computation of current outputs. This makes them particularly suitable for modeling dynamic systems with temporal dependencies. However, simple RNNs suffer from the so-called vanishing gradient problem, in which information from earlier time steps is gradually lost during training. In practice, variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are often used, as they can recognize and retain relevant patterns over longer periods.

A Generative Adversarial Network [14] consists of two neural networks trained in opposition: a generator and a discriminator. The generator produces synthetic data, while the discriminator attempts to distinguish between real and generated data. Training is considered successful when the discriminator can no longer reliably differentiate between real and artificially generated data. This learning principle has proven particularly effective for generating realistic data in image, speech, and text domains. In the present paper, a GAN is investigated to assess whether its generator is suitable for parameter prediction.

Table 1 summarizes the main differences between MLP, CNN, RNN, and GAN based on the preceding discussion. The comparison includes the network type, typical application areas, and key characteristics.

Table 1. Comparison of network architectures

Architecture	Network Type	Typical Task	Key Characteristics
MLP	Feedforward	Regression, Classification	Fully connected layers
CNN	Feedforward	Image processing, Time series analysis	Convolutional layers for feature extraction
RNN	Feedback	Sequences, Time-dependent data	Modeling temporal dependencies
GAN	Feedforward	Data generation (image, text, speech)	Two mutually connected networks

The identification models are based on synthetically generated simulation data. For model development, the dataset is divided into training, validation, and test subsets (see Figure 6). The datasets are disjoint, ensuring that no overlap occurs between the individual groups. The partitioning is carried out with respect to the variable system parameters — pipe length L, pipe inner diameter D, and buffer storage volume V. A detailed description of the data segmentation procedure is provided in Sections 2.6 and 2.7. The datasets are utilized in two phases of the modeling process. In Phase 1, the hyperparameters of the neural network are optimized. Training is performed using the training dataset, while the validation dataset is used for independent performance monitoring. In total, 30 randomly selected combinations of hyperparameters are tested. After completing the optimization, the combination yielding the lowest validation error is selected. In Phase 2, the neural network is retrained using the optimal hyperparameters determined in Phase 1 to generate the final model. Subsequently, the model performance is evaluated using the previously unseen test dataset. Due to stochastic effects during the training process (e.g., random initialization of network weights or the ordering of training samples), both training and evaluation are repeated ten times. Performing ten repetitions ensures statistically reliable averaging while keeping computational effort within reasonable limits. Model performance is assessed using the Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{1}$$

Here, N denotes the number of test samples, y_i represents the reference value from the simulation, and $\hat{y}_i$ denotes the model prediction. The MAE indicates the average deviation of the predicted values from the reference values.

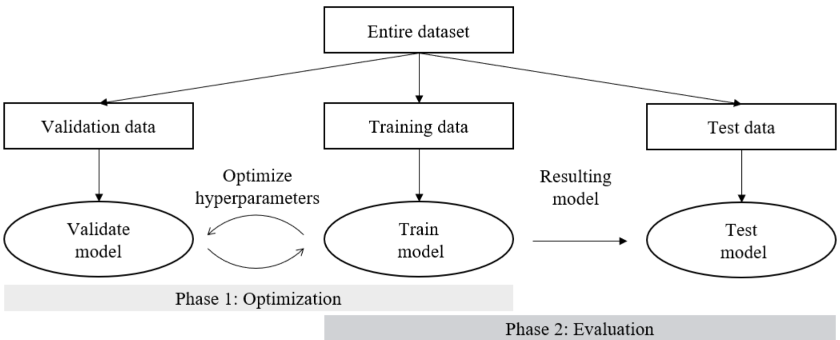

Figure 6. Dataset split (own illustration based on [16])

2.6. Study 1: One-Dimensional Analysis

In the first study, the ability of the four selected neural network architectures to predict individual system parameters in isolation is investigated. For this purpose, only one parameter is varied at a time, while the other two remain constant. The value ranges and step sizes are defined as follows: the pipe length L varies between 1 m and 30 m with a step size of 1 m, the pipe inner diameter D between 16 mm and 33 mm with a step size of 0.59 mm, and the buffer storage volume V between 50 dm^3 and 400 dm^3 with a step size of 12.07 dm^3. These step sizes yield 30 parameter configurations for each case. For each configuration, a simulation is carried out to generate the corresponding synthetic time series of sensor signals. Figure 7 illustrates the dataset partitioning for the three investigated parameters. Grey represents the training data, orange denotes the validation data, and blue indicates the test data. Each box in the figure corresponds to one simulated parameter configuration. The upper subplot shows the variation of the pipe length L, while the other parameters remain constant, with a pipe inner diameter of $D = 33$ mm and a buffer storage volume of $V =$

200 dm^3. The validation dataset includes the lengths $\{5, 11, 17, 23, 29\}$ m, whereas the test dataset comprises $\{2, 8, 14, 20, 26\}$ m. This distribution is uniform across the entire parameter range, ensuring complete coverage. The same principle is applied analogously to the pipe inner diameter D (middle subplot) and the buffer storage volume V (lower subplot). Accordingly, for each parameter, 20 training, 5 validation, and 5 test time series are available.

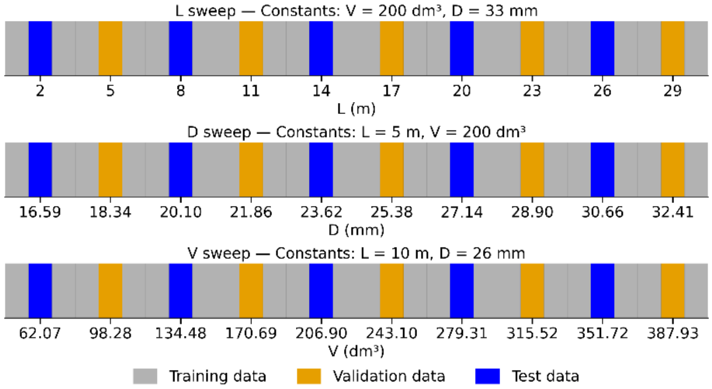

Figure 7. Dataset composition for one-dimensional analysis

2.7. Study 2: Multidimensional Analysis

In the second study, the ability of the four neural network architectures to predict all three system parameters simultaneously is investigated. For this purpose, the pipe length L, pipe inner diameter D, and buffer storage volume V are varied jointly. The parameter ranges are identical to those in Study 1. To reduce the computational effort, the number of parameter steps for D and V was reduced. The buffer storage volume V is now varied in increments of 50 dm^3, while the pipe inner diameter D takes the discrete values 16 mm, 26 mm, and 33 mm. The pipe length L remains unchanged with a step size of 1 m. This results in a total of 720 different parameter configurations. For each configuration, a simulation is performed to generate the corresponding synthetic time series of sensor signals. The datasets are subsequently divided into training, validation, and test sets. Approximately 80 % of the data are used for training, and 10 % each for validation and testing. Figure 8 illustrates the distribution of these datasets for the three diameters $D = 16$ mm, $D = 26$ mm, and $D = 33$ mm. The x-axis represents the pipe length L, and the y-axis represents the buffer storage volume V. As in Study 1, the validation and test datasets are uniformly distributed across the parameter space to ensure representative coverage. The test data shown in Figure 8 are hereafter referred to as test dataset A, which is used to assess the general model performance. To evaluate the models' interpolation capability, an additional dataset — test dataset B — is introduced. It consists of 60 randomly selected points within the parameter space, generated through uniformly distributed random sampling. An example parameter combination from dataset B is $L = 1.32$ m, $D = 19.85$ mm, and $V = 229.23$ dm^3. While dataset A includes only discrete grid values (e.g., $L = 1$ m or 2 m), dataset B also contains intermediate values such as $L = 1.32$ m. The same applies to D and V. Both test datasets are evaluated using the same trained models, enabling a direct comparison of their results.

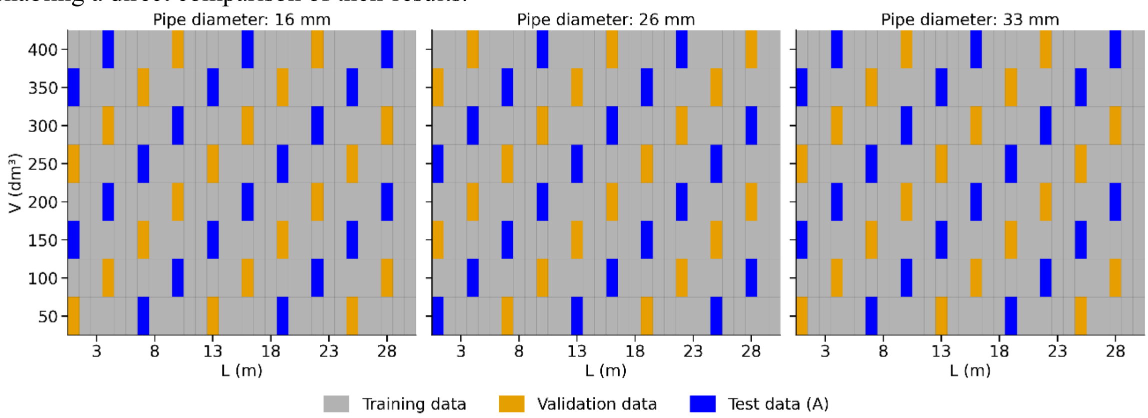

Figure 8. Dataset composition for multidimensional analysis

3. Results and Discussion

3.1. Study 1: One-Dimensional Analysis

In the first study, the prediction accuracy of the four neural network architectures was investigated, with each model predicting only one system parameter at a time. Table 2 summarizes the results obtained. The left column lists the respective target parameter (L, D, or V), including its unit and value range. The value range is provided to relate the results to the corresponding physical scale. The columns to the right present the mean absolute errors (MAE) of the four model architectures — MLP, CNN, RNN, and GAN. The reported error values represent the average MAE over all test data of the respective study. This allows for a concise comparison of the general predictive accuracy of the models without resolving the results for individual test points in detail. The results in Table 2 reveal distinct differences in prediction accuracy among the four investigated model architectures. The GAN model exhibits the highest mean error values for all three parameters and thus achieves the lowest accuracy. For the buffer storage volume V, the average deviation between the predicted and actual value amounts to 18.99 dm^3. The RNN and CNN architectures achieve comparable overall results. The RNN provides the best performance for estimating pipe length L (MAE = 0.21 m) and pipe inner diameter D (MAE = 0.14 mm), while for buffer storage volume V, the CNN achieves the highest accuracy with a mean error of 2.96 dm^3. The MLP shows an intermediate performance in all cases, lying between the RNN and GAN models. The observed differences in prediction accuracy can be explained by the architecture-specific characteristics of the models. The underlying task represents a regression problem with temporally correlated input signals. The Multi-Layer Perceptron processes these signals as static feature vectors and therefore does not account for temporal dependencies. Consequently, it can capture overall relationships between input and target variables but is unable to model dynamic patterns or short-term variations within the time series. In contrast, Convolutional Neural Networks can detect local structures within the signals. Through their convolutional layers, they can identify typical patterns such as temperature increases, speed fluctuations, or periodic signal behaviors and exploit these for parameter prediction. Recurrent Neural Networks extend this concept by additionally modeling temporal dependencies between consecutive data points. This enables them to capture long-term relationships across the entire time series. This capability explains the higher accuracy of RNN and CNN models compared to MLPs.

Table 2. MAE values of the neural networks for one-dimensional analysis based on the test data

Parameter	**Unit**	**Range**	**MAE test dataset**			
			MLP	CNN	RNN	GAN
L	m	1-30	0,45	0,31	**0,21**	1,17
D	mm	16-33	0,97	0,16	**0,14**	1,20
V	dm^3	50-400	4,68	**2,96**	4,01	18,99

3.2. Study 2: Multidimensional Analysis

In the second study, the reliability of the four neural network architectures in simultaneously predicting all three system parameters is examined. The obtained results are summarized in Table 3. The reported mean absolute errors (MAE) refer to the two test datasets, A and B. While test dataset A contains only parameter combinations located on the grid points, test dataset B includes randomly selected intermediate points within the parameter space. The results for test dataset A indicate that the CNN and RNN architectures again achieve the highest prediction accuracy. For the pipe length L, the CNN yields the best results with a mean error of 0.24 m, whereas the RNN achieves the lowest deviations for the pipe inner diameter D (MAE = 0.29 mm) and the buffer storage volume V (MAE = 3.19 dm^3). The MLP produces the largest errors in all cases, while the GAN performs slightly better than the MLP. For test dataset B, an increase in error values can be observed for all models. The largest increase occurs for the RNN, whereas the CNN achieves the lowest mean errors across all parameters (L = 0.81 m; D = 0.74 mm; V = 5.05 dm^3). The results obtained from test dataset A corroborate the findings from Study 1. While MLPs process input signals in a purely static manner, CNNs and RNNs can more effectively capture dynamic relationships by accounting for local and temporal structures, respectively. The decrease in accuracy observed for test dataset B can be attributed to data-related factors. The randomly sampled intermediate points correspond to regions of the parameter space that were insufficiently represented during training. As a result, the models need to perform stronger interpolation in these areas, which increases prediction uncertainty.

Table 3. MAE values of the neural network for multidimensional analysis on the test data

Parameter	**Unit**	**Range**	**MAE test dataset A**				**MAE test dataset B**			
			MLP	CNN	RNN	GAN	MLP	CNN	RNN	GAN
L	m	1-30	0,6	**0,24**	0,41	0,46	1,37	**0,81**	4,10	1,23
D	mm	16-33	0,79	0,48	**0,29**	0,61	1,23	**0,74**	2,12	0,81
V	dm³	50-400	6,45	4,10	**3,19**	4,45	9,49	**5,05**	9,41	5,40

4. Summary and Outlook

This paper presents a method for identifying technical parameters in building energy systems. The focus lies on estimating the pipe length, pipe inner diameter, and buffer storage volume of a heat pump system. These parameters influence the dynamic behavior of the overall system and are therefore important for control design. In practical applications, however, the actual values of these parameters are often unknown, which complicates the adaptation and optimization of control strategies. To address this issue, the paper investigates to what extent the system parameters can be reconstructed from time-resolved sensor data. Four different architectures of artificial neural networks (MLP, CNN, RNN, and GAN) were applied. The models were trained on synthetically generated simulation data of a heat pump system that represents all relevant thermal and hydraulic components.

Two studies were conducted in this work. In the first study, the models' ability to identify individual system parameters was evaluated. The CNN and RNN achieved the lowest mean deviations, as they were able to capture local and temporal structures in the input signals, respectively. The MLP exhibited lower accuracy due to its static architecture, while the GAN produced the highest error values. In the second study, the simultaneous estimation of all three parameters was investigated. Two test datasets were used: Dataset A comprised parameter combinations located at predefined grid points, whereas Dataset B contained randomly selected intermediate points within the parameter space. The results for Dataset A show that both the CNN and RNN achieved the highest estimation accuracy. When evaluated on Dataset B, the error values increased because these regions were not represented during training. Among all architectures, the CNN demonstrated the most robust performance on Dataset B.

Previous studies have primarily focused on identifying parameters on the building side, whereas the present work emphasizes the technical components on the generation side. The results demonstrate that neural networks can extract implicit system information from time-series data and utilize it for parameter identification. Notably, the models can estimate multiple system parameters simultaneously, eliminating the need for separate training runs for individual targets. Despite these promising results, several limitations of the proposed approach and its scalability toward real-world applications must be acknowledged. The presented investigation is based exclusively on synthetically generated data obtained from a simulation model. While this enables controlled experiments and systematic parameter variation, simulated data represent an idealized system behavior. Real-world heat pump installations are affected by measurement noise, sensor offsets, and unmodeled disturbances, which may lead to deviations from the dynamics observed in simulation. As a result, a performance gap between simulation-based evaluation and real-world deployment cannot be excluded. In addition, the study focuses on a specific hydraulic system configuration with a predefined topology. In practice, however, heat pump systems often exhibit heterogeneous hydraulic layouts and deviations from documented plumbing designs due to installation-specific adaptations or undocumented modifications. Such variations can alter the thermal and hydraulic behavior of the system and introduce ambiguities in the inverse parameter estimation problem. Moreover, interactions between the generation side and different building characteristics were only represented in a simplified manner, which may limit the transferability of the trained models across buildings with varying thermal properties. Addressing these challenges will require extending the approach to a broader range of system configurations and validating it using real measurement data. Furthermore, this paper considered only point estimators as identification models. In this case, the models yield a single predicted value for each system parameter without providing any measure of uncertainty. Future research should therefore explore methods that incorporate uncertainty quantification, such as Bayesian Neural Networks (BNNs) or simulation-based inference (SBI). These approaches could enable more robust and reliable decision-making in practical applications. Finally, the impact of estimation errors in L, D, and V on control performance and heat pump efficiency is not addressed in this study. Future work should investigate the parameter accuracy required to consistently attain a given efficiency improvement.

Acknowledgements

The author thanks the Robert Bosch GmbH for supporting this industrial PhD project.

References

[1] International Energy Agency (IEA), Heat Pumps, Paris, 2023, https://www.iea.org/energy-system/buildings/heat-pumps
[2] Ruhnau, O., Hirth, L., Praktiknjo, A., Heating with wind: Economics of heat pumps and variable renewables, Energy Economics 92 (2020), 104967.
[3] International Energy Agency (IEA), Heating, Paris, 2023, https://www.iea.org/energy-system/buildings/heating
[4] Villegas Mier, O., Dittmann, A., Herzberg, W., Ruf, H., Lorenz, E., Schmidt, M., Gasper, R., Predictive Control of a Real Residential Heating System with Short-Term Solar Power Forecast, Energies 16 (2023), Art. 6980
[5] Maasoumy, M., Razmara, M., Shahbakhti, M., Sangiovanni-Vincentelli, A., Handling model uncertainty in model predictive control for energy efficient buildings, Energy and Buildings 77 (2014), 377–392
[6] Shin, H., Park, C.-S., Parameter estimation for building energy models using GRcGAN, Building Simulation 16 (2023), 629–639
[7] Wang, D., Zheng, W., Li, S., Chen, Y., Lin, X., Wang, Z., Impact analysis of uncertainty in thermal resistor-capacitor models on model predictive control performance, Energy and Buildings 328 (2025), Art. 115112
[8] Bengea, S. et al., Parameter estimation of a building system model and impact of estimation error on closed-loop performance, Proc. 50th IEEE Conf. on Decision and Control and European Control Conference, Orlando, FL, 2011, 5137–5143
[9] Ko, Y.-D., Park, C.-S., Transfer learning based inverse modeling to identify unknown building properties, Proc. Building Simulation 2021: 17th Conference of IBPSA, Bruges, 2021, Vol. 17, 1896–1903
[10] Lettermann, L., Jurado, A., Betz, T., Wörgötter, F., Herzog, S., Tutorial: a beginner's guide to building a representative model of dynamical systems using the adjoint method, Communications Physics 7 (2024), Art. 128
[11] Furtwengler, J., Constantin, A., Keuper, J., Neuronale Netze zur Identifikation relevanter Parameterwerte in Wärmepumpensystemen, Proc. DKV-Tagung 2025: Deutscher Kälte- und Klimatechnischer Verein e. V., Magdeburg, 2025, Vol. 52
[12] Shirani, A., Merzkirch, A., Roesler, J., Leyer, S., Scholzen, F., Maas, S., Experimental and analytical evaluation of exhaust air heat pumps in ventilation-based heating systems, Journal of Building Engineering 44 (2021), Art. 102638
[13] LeCun, Y. A., Bottou, L., Orr, G. B., Müller, K.-R., Efficient BackProp, in: G. Montavon, G. B. Orr, K.-R. Müller (Hrsg.), Neural Networks: Tricks of the Trade: Second Edition, Springer, Berlin, 2012
[14] Sarker, I. H., Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions, SN Computer Science 2 (2021), Art. 420
[15] Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., Gandomi, A. H., Deep learning modelling techniques: current progress, applications, advantages, and challenges, Artificial Intelligence Review 56 (2023), 13521–13617
[16] Choo, K., Greplova, E., Fischer, M., Neupert, T., Machine Learning kompakt: Ein Einstieg für Studierende der Naturwissenschaften, Springer Vieweg, Heidelberg, 2020