Energy Conversion and Management 279 (2023) 116769

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Data-driven soft sensors targeting heat pump systems

Yang Song [a], Davide Rolando [a], Javier Marchante Avellaneda [b], Gerhard Zucker [c], Hatef Madani [a,c,*]

[a] Department of Energy Technology, KTH Royal Institute of Technology, Brinellvägen 68, 100 44 Stockholm, Sweden
[b] Instituto Universitario de Investigación en Ingeniería Energética, Universitat Politècnica de València, Camino de Vera S/n, Ed. 8E Semisótano, 46022, Valencia, Spain
[c] Sustainable Thermal Energy Systems, Austrian Institute of Technology, Giefinggasse 2, A-1210 Vienna, Austria

ARTICLE INFO

Keywords:
Data driven
Heat pumps
Soft sensors
ANN
Regression
Database

ABSTRACT

The development of smart sensors, low cost communication, and computation technologies enables continuous monitoring and accumulation of tremendous amounts of data for heat pump systems. But the measurements, especially for domestic heat pump, usually suffer from incompleteness given technical and/or economic barriers, which prevents database of measurements from being exploited to its full potential. To this end, this work proposes a data-driven soft sensor approach for compensating multiple missing information. The soft sensors are developed based on an ANN model, an integrated multivariate polynomial regression model and empirical model by considering different constrains like data and information availability during model establishing process. All the three models have been validated against the data from a field test installation, and showed good performance for all the compensated variables. Of the three models, the ANN model shows the best performance for all soft sensors, but it has the highest requirement for additional resources to collect training data. While the integrated multivariate polynomial regression model demonstrates excellent accuracy for the majority of soft sensors with manufacturers' subcomponent data which needs no extra cost. Even though empirical model is not as accurate as the other two models, it still performs good accuracy with limited information from performance map. The methods developed in the present study paves the way for available measured data in thousands of installations to be fully utilized for innovative services including but not limited to: improved heat pump control strategies, fault detection and diagnosis, and communication with local energy grids.

1. Introduction

In the recent years, heat pump technology has spread worldwide as an energy efficient technology for the integration of renewable energy [1,2]. According to Energy Technology Perspectives 2020 published by International Energy Agency (IEA), heat pump utilization should be extended in both industry and building sectors under the scenario of sustainable development [3]. Thanks to the development of smart sensors, low cost communication, and computation technologies, modern heat pump units are nowadays equipped with numerous sensors and potentially generate a large amount of data every day that can be stored in databases for further analysis [4,5].

Currently this highly valuable and abundant data with billions of entries every month is merely collected by heat pump manufacturers and is rarely processed or used for any innovative service. Heat pump monitoring systems, especially for domestic heat pumps, are often not designed to provide data that can be directly employed for system analysis. The use of this data is nowadays mostly limited to real-time visualization or for triggering alarms when there is an error during operation. Furthermore, a large amount of data does not necessarily assure that the quality of the data is good enough for system analysis. In real practice, the monitoring data commonly suffers from incompleteness, inconsistency, inaccuracy due to lack of key sensors or meters. For example, it is uneconomical to install mass flow meters or sometimes power meters in domestic heat pumps systems. In addition, most modern heat pump systems use pressure sensors, but the lack of pressure sensors remains an unresolved problem for the thousands of existing systems that started without them. Under this situation, the potential of highly valuable monitoring systems has not been fully unlocked. The collected data, if processed and coupled to other sources of information in an appropriate way, encompasses valuable information about local outdoor and indoor climate, occupancy and activities in the built-environment, people behavior, building characteristics, components

* Corresponding author at: Department of Energy Technology, KTH Royal Institute of Technology, Brinellvägen 68, 100 44 Stockholm, Sweden.
E-mail address: hatef.madani@energy.kth.se (H. Madani).

https://doi.org/10.1016/j.enconman.2023.116769
Received 10 November 2022; Received in revised form 14 January 2023; Accepted 30 January 2023
Available online 4 February 2023

and system efficiencies, and the early symptoms of possible faults or performance degradation in heat pump systems. Therefore, to make the collected measurements ready for system analysis, it is quite essential to conduct a comprehensive database investigation to detect, complement and correct the incomplete, incorrect, inaccurate parts of the data.

To solve the problems related to deficient measurements, soft sensors, also named as virtual sensors [6], inferential sensors [7] are developed to predict unmeasurable, difficult-to-measure, or costly measurements through prediction or estimation models using "cheap or easy to measure" historical data or other resources [8]. During recent years, soft sensors are gaining focus and proliferation in various areas, such as energy systems, pharmaceutical manufacturing, petrochemical, and other industry process [9–11], mainly due to their low cost, quick response, and simple maintenance. Generally, there are two types of models for soft sensors: first-principle models (also called model-driven models) and data-driven models [12]. First-principle models need detailed knowledge of the underlying physical phenomenon to build the governing equations that can describe system behavior. However, in real complex systems the equations are usually very difficult to solve with many unknown or hard-to-estimate parameters. Out of this reason, data-driven models have become popular, since they can mine historical data obtained from normal use or test systems to establish linear or nonlinear models that can describe the relationship between inputs and inferential outputs of the system. Numerous modeling techniques have been successfully applied to data-driven soft sensors, of which the most popular ones are Artificial Neural Networks [13–16], Support Vector Machines [17], Partial Least Squares [18] and Principle Component Analysis combining a regression model [19,20].

The implementation of soft sensors in heat pump systems are also found in last decade studies [21,22]. The commonly utilization of soft sensors in heat pumps is to estimate mass flow rate [23], refrigerant pressures [24], compressor power consumption [25]. Estimating refrigerant mass flow rate is quite essential for system monitoring and fault detection, but it is generally infeasible to install mass flow meters especially for residential heat pump systems due to cost and maintenance problems. Pressures in the refrigerant cycle can reveal the operational performance of heat pump systems. However, pressure sensors are usually expensive. Also, the installation of pressure sensors are challenging since the refrigerant loop needs to be evacuated and recharged. Moreover, pressure sensors may possibly cause refrigerant leakage over time [26]. Compressor power consumption is an essential indicator to evaluate the efficiency of heat pumps, but again power meters are often not installed for economic reasons. Therefore, it is critical to develop soft sensors based on easy-to-measure and low-cost measurements, e.g., temperatures and revolutions per minute (rpm), so that the soft sensors can take place of their hardware counterparts.

However, in most of the existing references, soft sensors can only estimate one single measurement based on other measurements that are also difficult or expensive to obtain. Table 1 summarizes commonly seen soft sensors and associated inputs for heat pumps in literature. As can be seen for the table, most soft sensors not only require easy measurements like discharge temperature, suction temperature, or ambient temperature, but also need not readily available measurements like condensing temperature, evaporating temperature or pressures as inputs. What is more, based on our experience several or even all of expensive or difficult-to-measure parameters are usually missing in the existing heat pump monitoring systems. Thus, it is necessary to develop models that can compensate multiple absent costly measurements combined with more affordable physical sensors.

To bridge the research gaps mentioned above, this paper proposes exploration on the application of data-driven soft sensors in heat pumps through following:

1. Three data-driven models based on different additional resources are introduced to compensate the missing measurements in actual heat pump monitoring systems.
2. Considering the complexity of integrated multivariate polynomial regression model, sensitivity analysis is conducted to figure out the most dominated inputs.
3. Each model's performance in terms of accuracy are compared for every compensated variable.

Table 1
Soft sensors and related inputs in referred literature.

Literature	Soft sensor	Inputs
[23]	Refrigerant mass flow rate	W, T_{dis}, P_{dis}, T_{suc}, P_{suc}, f_{comp}
[27]	Refrigerant mass flow rate	P_{dis}, P_{suc}, T_{suc}, T_{amb}
[28]	Refrigerant mass flow rate	W, T_c, P_c, T_e, P_e, T_{dis}, T_{suc}
[24]	Pressure sensors	T_e, T_c
[25]	Compressor power	T_e, T_c, f_{comp}
[28]	Compressor power	T_e, T_c, T_{suc}, P_{suc}

Table 2
Available and missing variables for system performance analysis.

Available measurements	Missing information
Compressor speed (rpm)	Evaporation temperature/pressure
Water inlet temperature	Condensation temperature/pressure
Water outlet temperature	Mass flow rate of water/brine/refrigerant
Brine inlet temperature	Heating capacity
Brine outlet temperature	Cooling capacity
Compressor inlet temperature	Compressor power consumption
Compressor outlet temperature	COP
Condenser outlet temperature	

The remainder of the paper is structured as follows: Section 2 describes the data analysis from the heat pump operation database which motivated the design of soft sensors, and also for the analysis of field test installation data which is used to build and validate the models. Then Section 3 demonstrates the details of three different models. In Section 4, soft sensors' prediction and validation results are presented. Finally, the main conclusions are given in Section 5.

2. Data analysis

2.1. Database analysis

This work is motivated by the prevalence of incomplete data collection in a database containing operational data from over 4000 domestic heat pump installations across Sweden. The data collected from thousands of domestic heat pumps share similar characteristics since those heat pumps are equipped with almost the same type of sensors. The analysis of monitoring data revealed that mass flow meters, compressor power meters, pressure sensors and temperature sensors for condensation and evaporation are typically not installed and therefore not monitored, mainly due to economic barriers. This prevents the direct evaluation and monitoring of system performance. Table 2 shows the list of variables found in the database, along with a list of missing variables that would allow the direct evaluation of system performance. In order to fully utilize the database for heat pump innovative services, a novel data-driven multi-model approach is proposed for soft sensors that can predict several missing measurements at the same time without the need for installing extra costly sensors or meters.

The missing measurements in the database motivated the development of data driven soft sensors. Once soft sensors are accomplished, they can be applied for thousands of heat pump installations. The common status of missing variables in the big database 2 provides guidance regarding which soft sensors should be developed. However, for the purpose of developing and validating soft sensor models at the same time, field test data is utilized, which is described in Section 2.2.

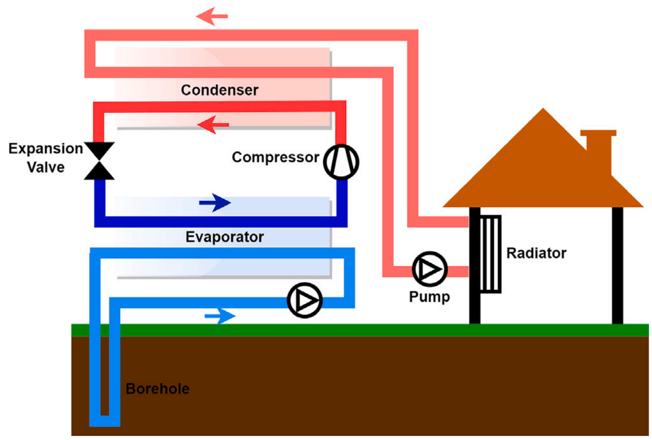

Fig. 1. Schematic configuration of the investigated heat pump system.

2.2. Field test data analysis and preprocess

The models developed in this study have been validated using the measurement data from one field test heat pump installation equipped with several extra sensors, including energy meters and pressure sensors. The heat pump is a classic ground source heat pump, whose schematic configuration is shown as in Fig. 1. So apart from common measurements (i.e.,compressor speed, water inlet and outlet temperatures, brine inlet and outlet temperatures, condenser inlet and outlet temperatures, and evaporator outlet temperatures), extra measurements like water mass flow rate, evaporation and condensation pressures were also collected. The first part of the analysis consisted in data cleaning, including the identification of outliers, anomalies and data gaps. The outliers have been identified following the interquartile range approach, typically used in visualization tools like Box whisker diagram [29]. Besides, the sensor measurements are sometimes discontinuous, and a time based interpolation is introduced to fill the missing values which exist sporadically.

The operational range and distribution of the measurements are quite important for developing models. If the ranges of data are too narrow, the trained models are at high risk of over fitting. Violin plots are adopted to explore the collected data, which can graphically show the statistical features and density shape of numerical data [30].

During model developing process, 16 numeric variables are utilized to train and/or validate data-driven models. The range and density distribution of these 16 variables are shown with violin plots in Fig. 2. The temperature ranges of outdoor, condensation and evaporation are 15 °C, 15 °C, 5 °C respectively. The compressor frequency changes from 20 Hz to 45 Hz. The dataset shows a reasonable spread ensuring the trained model is not over fitting.

3. Methodology

3.1. Framework of data-driven modeling

This work proposes a methodology based on a multi-model approach that aims at compensating the missing values through soft sensors for heat pump monitoring systems, without requiring the installation of additional devices. The whole frame of research methodology is demonstrated in Fig. 3. Three different data-driven models are developed in parallel for soft sensors by considering the availability of different additional resources. An ANN model can be developed when training data collected from installations with added sensors and meters is accessed. With the data and information of heat pump components from manufacturers, an integrated multivariate regression model can be built up. In the worst scenario, if neither training data nor data of component is available, an empirical model is established as an alternative with the heat pump performance map, which is normally easy to obtain for most heat pump systems. The details of three models are introduced in Section 3.2, Section 3.3 and Section 3.4 respectively.

3.2. Multi-inputs and multi-outputs ANN model

ANN is a very popular algorithm inspired by a biological neural system which is composed of interconnected neurons in input, hidden and output layers. Different neurons in different layers are connected by weights and parameters, which can be trained with back propagation algorithm. Recent studies proposed the implementation of ANN for heat pump system analysis. The research areas of ANN analysis in heat pump systems broadly include: (i) prediction of heat pump energy consumption [31], (ii) prediction of the performance of sub-components [32,33], (iii) prediction of properties of refrigerants [34]. Fannou et al. [31] applied ANN to predict the heating capacity and compressor power consumption of a direct expansion geothermal heat pump. In their study, ANN model provided very satisfactory results for the 2 outputs with 6 thermodynamic parameters as inputs, which demonstrated that ANN is a good alternative approach in modeling complex systems. Longo et al. [33] used ANN to predict heat transfer coefficients of refrigerant condensation in a brazed plate heat exchanger. The result showed much better performance than analytical-computation models presented in the recent open literature. Wang et al. [34] developed a ANN models to precisely predict viscosity and thermal conductivity of the HFC/HFO refrigerants. Table 3 summarizes the inputs and outputs of ANN models in the literature above.

In this study, since several missing values are need to be compensated, a multi-inputs and multi-outputs back propagation neural network are established. The schematic diagram of the ANN structure is illustrated in Fig. 4, where the features on input layer are the measurements that are commonly collected in domestic heat pumps, and the outputs are missing values needed to be compensated. The hidden layer plays a role of transmitting and transforming the information coming from input layer to output layer through an activation function. Commonly used activation functions are Sigmoid, Tanh, ReLu, Leaky ReLu, and ELU, etc. In this study, ReLu is chosen as the activation function which is expressed by Eq. (1). Compared to other activation functions, ReLu only activates the neurons if output of linear transformation is larger than 0, which can greatly improve computation efficiency [35].

$$f(x) = \max(0, x) \tag{1}$$

Hyper-parameters play a prominent role on ANN model performance. The process of optimizing hyper-parameters is to find an optimal point where the model has satisfied accuracy and at the same time it is not suffering from over fitting. In this study, Adam is applied as mini-batch stochastic gradient descent optimizer, which is efficient in computation and has low memory requirement [36]. The hyper-parameters to be optimized in this study are number of hidden layers, number of hidden neurons, batch size, epochs, and learning rate. Possible candidates for these hyper-parameters are listed in Table 4. A total of 1260 different hyper-parameter combinations are considered. To search optimized hyper-parameters there are mainly two ways: grid search and random search. In grid search, each possible candidate combination is tried, while in random search, a fixed number of randomly selected hyper-parameters are tried instead. It has been proved that random search is more efficient than grid search for hyper-parameter optimization [37]. In order to improve computation efficiency, RandomizedSearchCV [38] method is adopted to randomly select 250 candidate trials.

The whole dataset is divided into training and testing datasets with proportions of 75% and 25%, respectively. Since variables with different units do not contribute equally to the analysis, the dataset

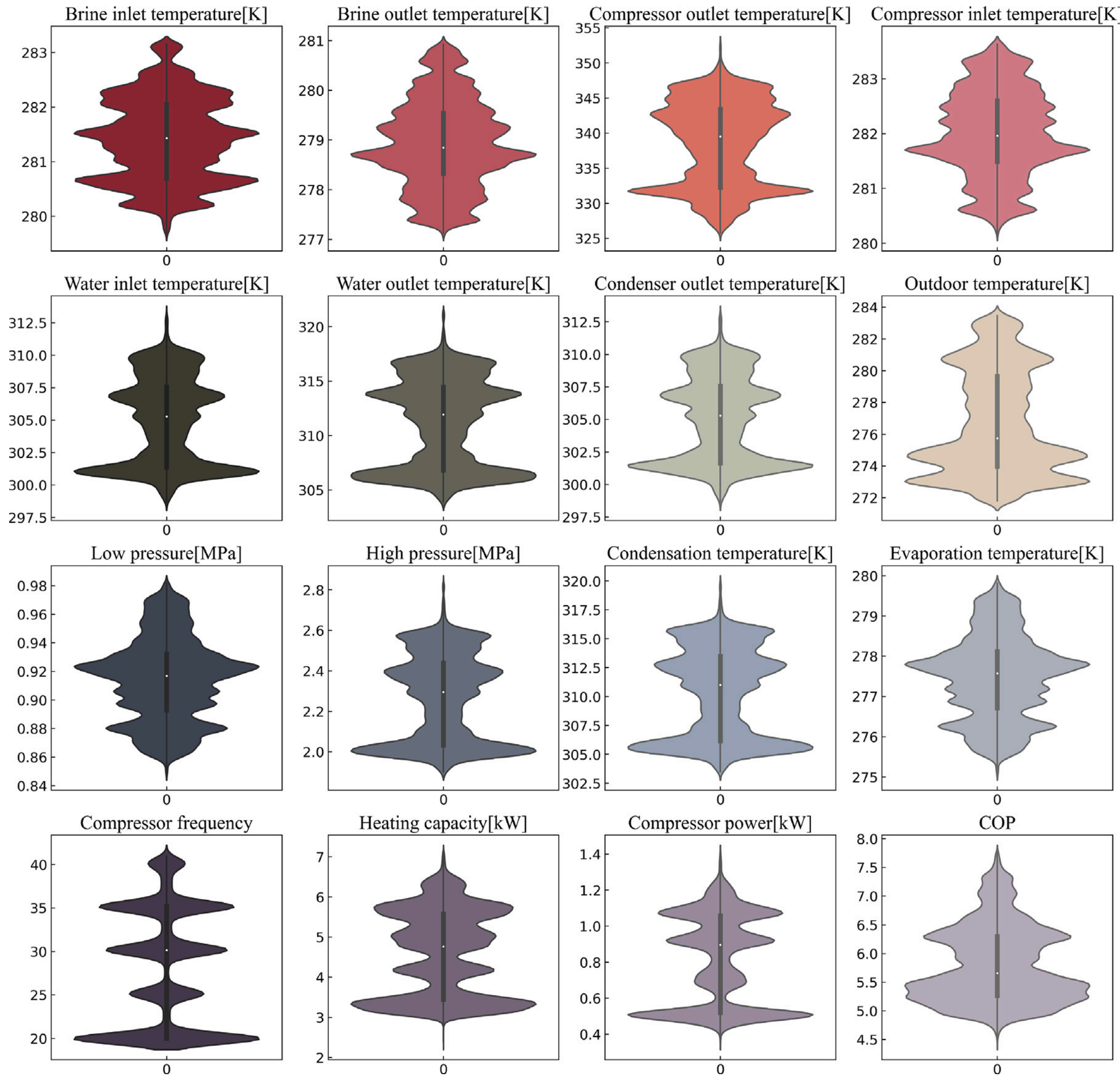

Fig. 2. Range and distribution of the measurements.

Table 3
Inputs and outputs of ANN models in referred literature.

Literature	Inputs	Outputs
[31]	Evaporator inlet temperature and pressure, Evaporator outlet temperature and pressure, Water inlet temperature in condenser, Discharge pressure	Heating capacity Compressor power consumption
[33]	Driving temperature difference, Vapor superheat, Corrugation enlargement ratio, Equivalent Reynolds number, Liquid Prandtl number	Heat transfer factor
[34]	Reduced pressure, Reduced temperature, Mole mass, Acentric factor	Thermal conductivity Viscosity

is normalized in training dataset as Eq. (2), where μ, σ are the mean and standard deviation of variable x respectively. The test dataset is normalized with the same μ and σ from training dataset for the same variable. During training process, five-fold cross-validation is applied to get the hyper-parameters with lowest loss function evaluated by mean square error (MSE), as shown in Eq. (3), where N is the total number of data points, f_i is the value obtained from the model and y_i is the real value for data point i. After the training process, the optimized results is shown in Table 4 and the MSE under this condition is 0.0062.

$$x_{i,normalized} = \frac{x_i - \mu(x)}{\sigma(x)} \tag{2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2 \tag{3}$$

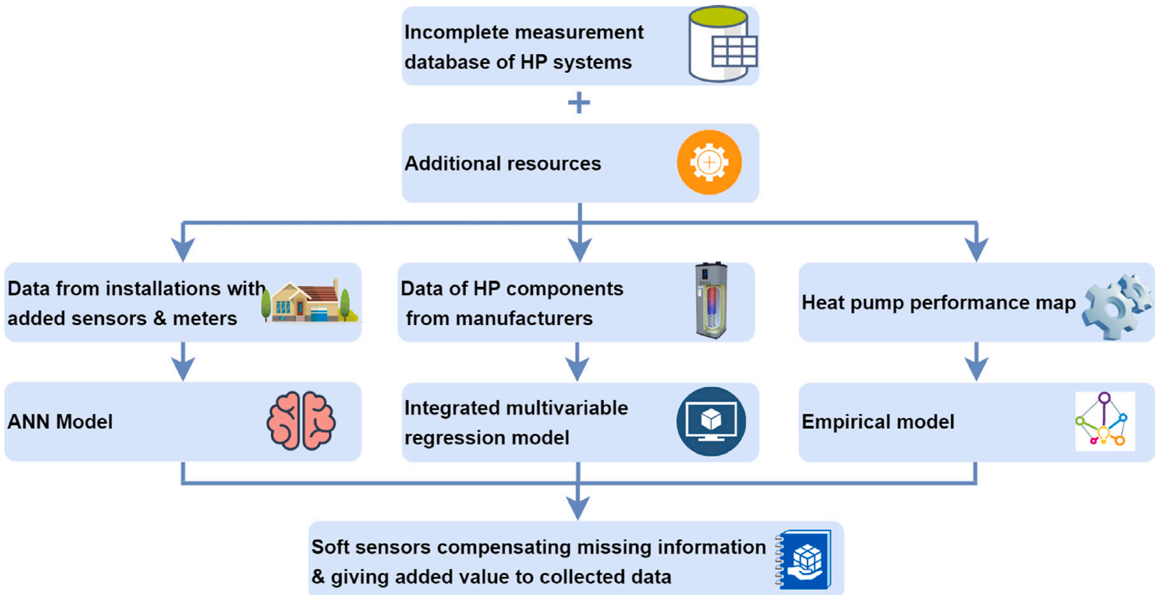

Fig. 3. Frame of research methodology.

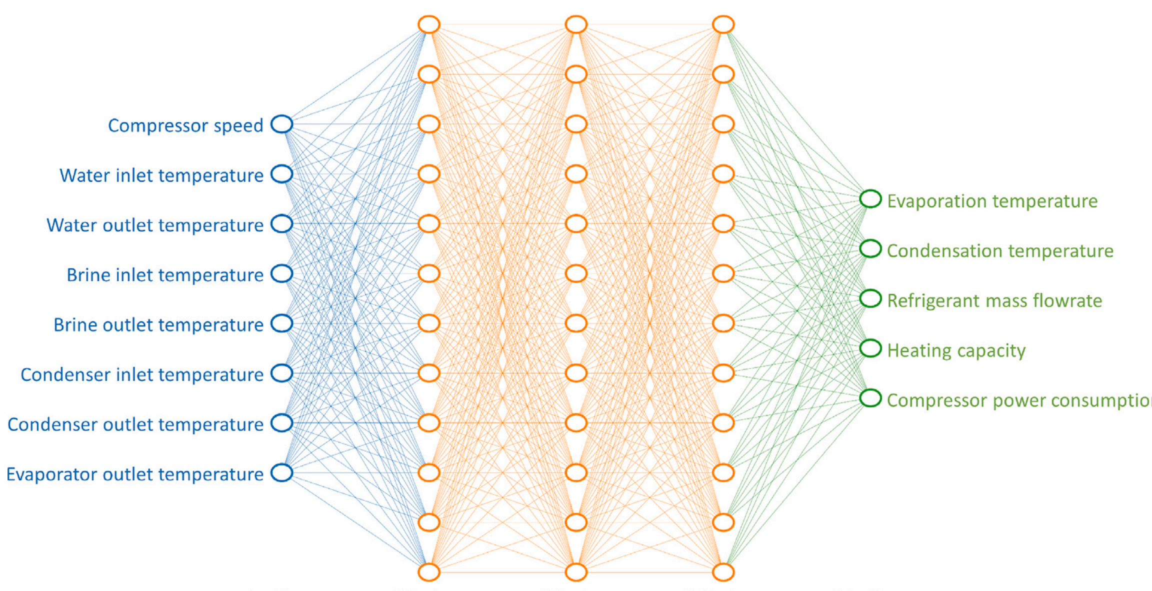

Fig. 4. ANN model structure.

Table 4
Hyper-parameter candidates.

Hpyer-parameters	Candidate values	Optimized result
Number of hidden layers	[1,2,3]	3
Number of hidden neurons	[5, 10, 15, 20, 25, 30,35]	25
Learning rate	[0.0001,0.001,0.01,0.1]	0.0001
Batch size	[10, 20, 40, 60, 80]	20
Epochs	[10, 50, 100]	50

3.3. *Integrated multivariate polynomial regression model*

Considering the circumstance where training data for ANN model are not available due to lack of extra sensors, an alternative integrated multivariate polynomial regression model is designed to solve the compensation task in this study. This model only needs the model type information of sub-components (condenser, evaporator, and compressor) and data from manufacturers' software. Manufacturers' software are mainly designed for customers to select proper products like heat exchangers or compressors. The interface of software is customer-oriented, which means the inputs and outputs are mainly designed based on the majority of customers' need for selecting proper products. From the stand point of model development, the data from manufacturers' software is quite informative and valuable. However, the inputs and outputs of model are usually not the same with software interface. The details of inputs and outputs of software and our models are shown in Tables 5–7. Besides, in the monitoring database, a huge number of missing values are needed to be compensated along timestamps, it is quite time consuming to run the manufacturers' software for every missing value. So the manufacturers' software is usually not directly useful for modeling. For these reasons, a reverse engineering procedure is applied to develop models. Reverse engineering usually involves deconstruct original product and extract design information to generate new products. This method is commonly used in software systems, where the purposes mainly cover: (*i*) determine internal relationships

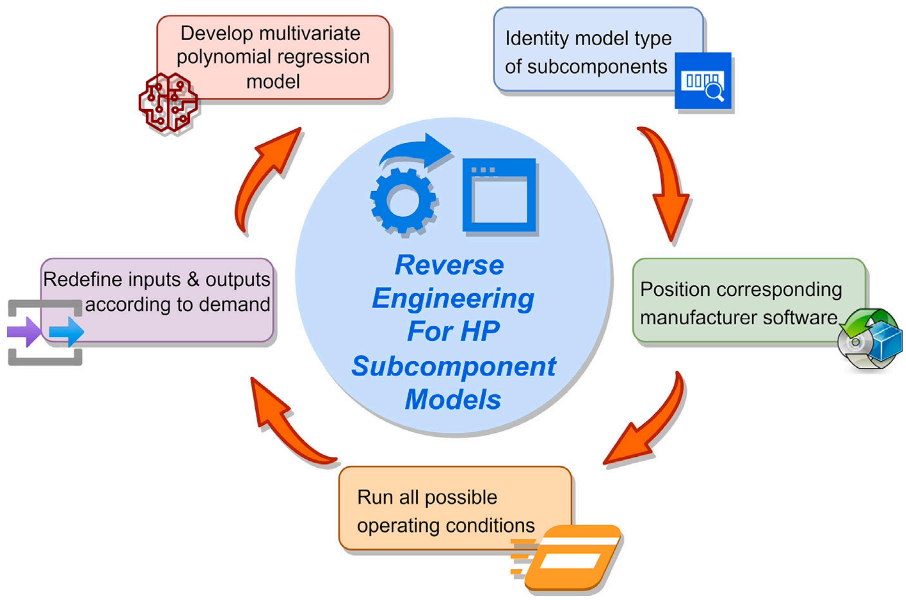

Fig. 5. Multi-variable polynomial regression model develop process for Subcomponents.

among software components and (*ii*) construct novel form or create high-level representations of software [39,40]. In this study, firstly the model types of heat exchangers and compressor are identified. Then manufacturer's software are figured out accordingly. After that all possible operating conditions are ran to collect information and data. According to measurements missing status in real database, inputs and outputs are redefined and finally multivariate polynomial regression models are developed. The model is implemented to compensate all missing information at fast speed. The development process is demonstrated as Fig. 5.

In this paper, multivariate polynomial regression models of subcomponents: evaporator, condenser and compressor are developed separately. The detailed model is described in Sections 3.3.1–3.3.3. Then a heat pump unit integrated model is established by combining the sub-components together, which is described in Section 3.3.4.

3.3.1. Condenser multivariate polynomial regression model

As mentioned before, almost all of the existing heat pump installations monitored in this study are missing pressure sensors in both low and high pressure sides. Therefore, condenser model is designed to obtain condensation temperature or pressure. In this study, the direct output of condenser model is condensation temperature and then the condensation pressure is calculated from CoolProp, which is an open-source thermophysical property library [41]. Based on Eq. (4) to Eq. (8), and the availability of measurements (Table 2), the inputs and output of the model are selected and demonstrated in Table 5. The modeling purpose is to build a multivariate polynomial regression expression matching the output and inputs, as shown in Eq. (9), where b is the regression intercept, c_1 to c_{20} are coefficients of polynomial terms, x_1 to x_5 represents input variables: refrigerant mass flow rate, water inlet temperature, water outlet temperature, condenser inlet temperature, and condenser outlet temperature respectively, T_c is the output of the model, namely condensation temperature. It is worth mentioning when this sub-component model is being developed, the data of all variables is extracted from a manufacturer's heat exchanger software according to specific condenser type provided by heat pump manufacturer. Then the data is used to train and test the polynomial regression model mapping the inputs and output in Eq. (9). But when the completed model is being used as soft sensors for condensation pressure and temperature, refrigerant mass flow rate as one missing

Table 5
Inputs and outputs of software and multivariate polynomial model for condenser.

Variables	Software	Model
Mass flow rate of refrigerant	Output	Input(x_1)
Inlet temperature of water	Input	Input(x_2)
Outlet temperature of water	Input	Input(x_3)
Inlet temperature of refrigerant	Input	Input(x_4)
Outlet temperature of refrigerant	Output	Input(x_5)
Subcooling	Input	–
Heating capacity	Input	–
Condensation temperature	Output	Output(T_c)

measurement, is the output from compressor sub-component model, shown in Section 3.3.4.

$$Q_c = UA \cdot LMTD \tag{4}$$

$$Q_c = f(x_1, h(x_4, P_c), h(x_5, P_c)) \tag{5}$$

$$P_c = f(T_c) \tag{6}$$

$$LMTD = f(x_2, x_3, x_4, x_5) \tag{7}$$

$$UA \cdot f(x_2, x_3, x_4, x_5) = f(x_1, h(x_4, T_c), h(x_5, T_c)) \tag{8}$$

$$\begin{aligned} T_c = b + \sum_{i=1}^{5} x_i c_i + \sum_{i=1}^{5} x_1 x_i c_{5+i} + \sum_{i=2}^{5} x_2 x_i c_{9+i} \\ + \sum_{i=3}^{5} x_3 x_i c_{12+i} + \sum_{i=4}^{5} x_4 x_i c_{14+i} + \sum_{i=5}^{5} x_5 x_i c_{15+i} \end{aligned} \tag{9}$$

To optimize model coefficients, ordinary least squares optimization method [42] is applied. The accuracy performance of the model is measured by relative root mean squared error (RRMSE) [43], defined in Eqs. (10), where f_i is the ith result from models, y_i is the ith actual measurements, and $\bar{y}$ is the average of y_i of all n data points. The smaller this indicator is, the more precise a model is. The RRMSE is 0.1% for the optimized condenser model. The coefficients from c_1 to c_{20} are listed on the top of each subplot in Fig. 6, and the intercept term b is 39.373. To make clear how each coefficient and polynomial term contribute to the final result, an sensitivity analysis is conducted. Every time, only one coefficient is changed by increasing 30%, decreasing 30% and totally dropping out to check how the RRMSE varies. In the

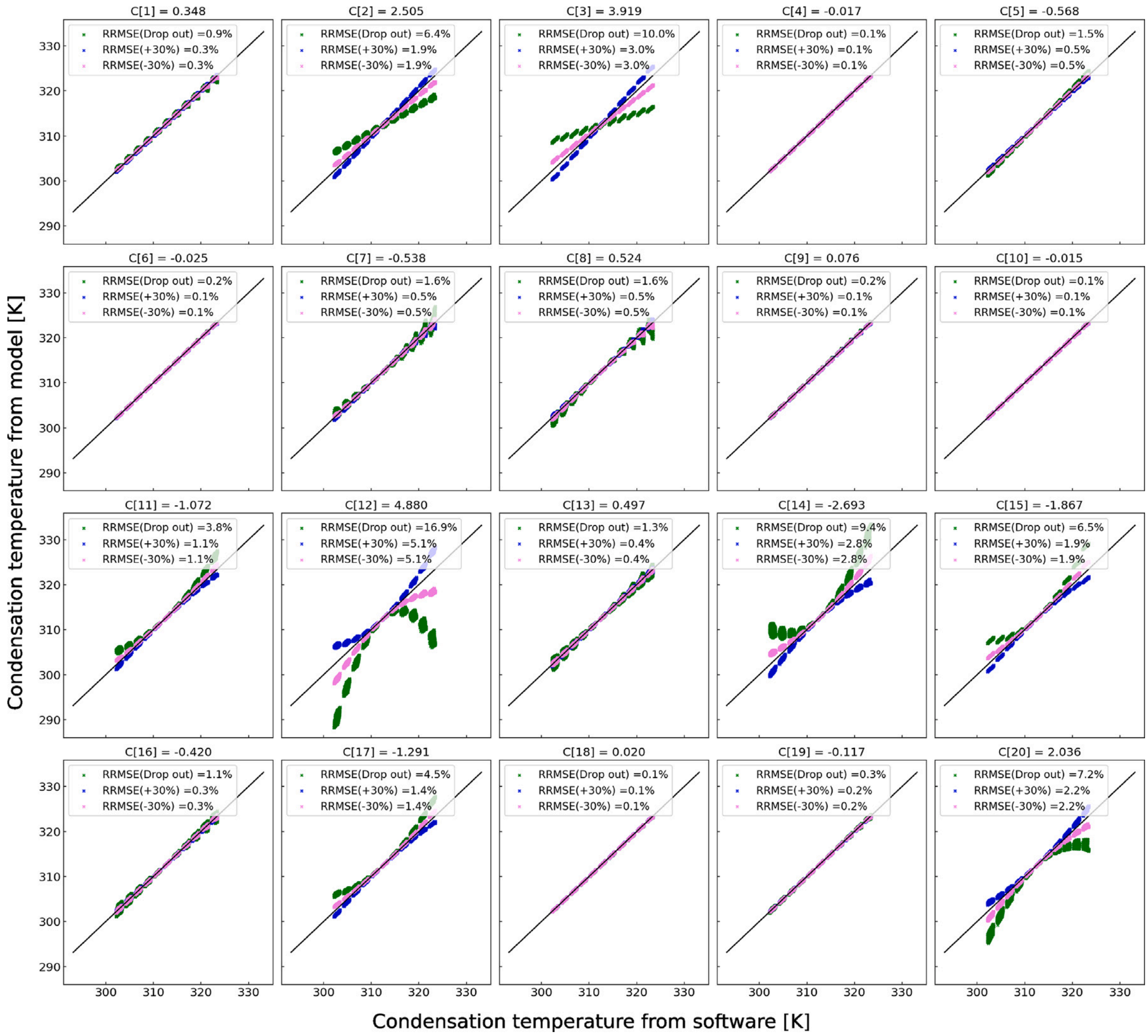

Fig. 6. Sensitivity analysis of condenser model coefficients.

upper left corner of subplots in Fig. 6, new RRMSE are listed when corresponding coefficient is modified. For more intuitive, scatter plots comparing model result against original data in software are exhibited. As can be seen from Fig. 6, the top five prominent coefficients are c_2, c_3, c_{12}, c_{14}, and c_{20}. The corresponding polynomial terms with these five coefficients are: x_2, x_3, x_2x_3, x_2x_5, and x_5^2. This means, water inlet and outlet temperature and condenser outlet temperature are more indispensable and sensible for regression of condensation temperature compared to refrigerant mass flow rate and condenser inlet temperature.

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - f_i)^2}{n}} / \bar{y} \tag{10}$$

3.3.2. Evaporator multivariate polynomial regression model

Similar to condenser model, evaporator model is developed to calculate evaporation temperature or pressure. The selection of inputs to evaporator model is the same way as condenser except for the refrigerant inlet temperature, since this temperature is usually not measured in the monitored data. Out of this reason, inlet enthalpy is selected as an input to evaporator model instead, because the expansion process through the expansion valves are assumed to be isenthalpic. Inputs and output of the model are shown in Table 6. The multivariate polynomial regression model is expressed as Eq. (11), where b is regression constant coefficient, c_1 to c_{20} are coefficients of polynomial terms, x_1 to x_5 represent input variables: suction mass flow rate, evaporator inlet enthalpy, evaporator outlet temperature, brine inlet temperature, and brine outlet temperature respectively. T_e is the out put of the model (evaporation temperature). The data to develop the model is obtained from a manufacturer's heat exchanger software. But when this model has been finished and is implemented in actual database, evaporator enthalpy in and suction mass flow rate are not available from measurements. The information regarding these two variables are provided by outputs from condenser and compressor models respectively, which is introduced in Section 3.3.4.

$$\begin{aligned} T_e = b + \sum_{i=1}^{5} x_i c_i + \sum_{i=1}^{5} x_1 x_i c_{5+i} + \sum_{i=2}^{5} x_2 x_i c_{9+i} \\ + \sum_{i=3}^{5} x_3 x_i c_{12+i} + \sum_{i=4}^{5} x_4 x_i c_{14+i} + \sum_{i=5}^{5} x_5 x_i c_{15+i} \end{aligned} \tag{11}$$

With ordinary least squares optimization method, the coefficients of the model are optimized, with RRMSE being 0.6%. On the top of each subplot in Fig. 7, the coefficients from c_1 to c_{20} are recorded. The intercept term b is −0.397. To further understand how much is the model result sensible to every coefficient and polynomial term, an sensitivity analysis is conducted. Only the coefficient listed on top of the

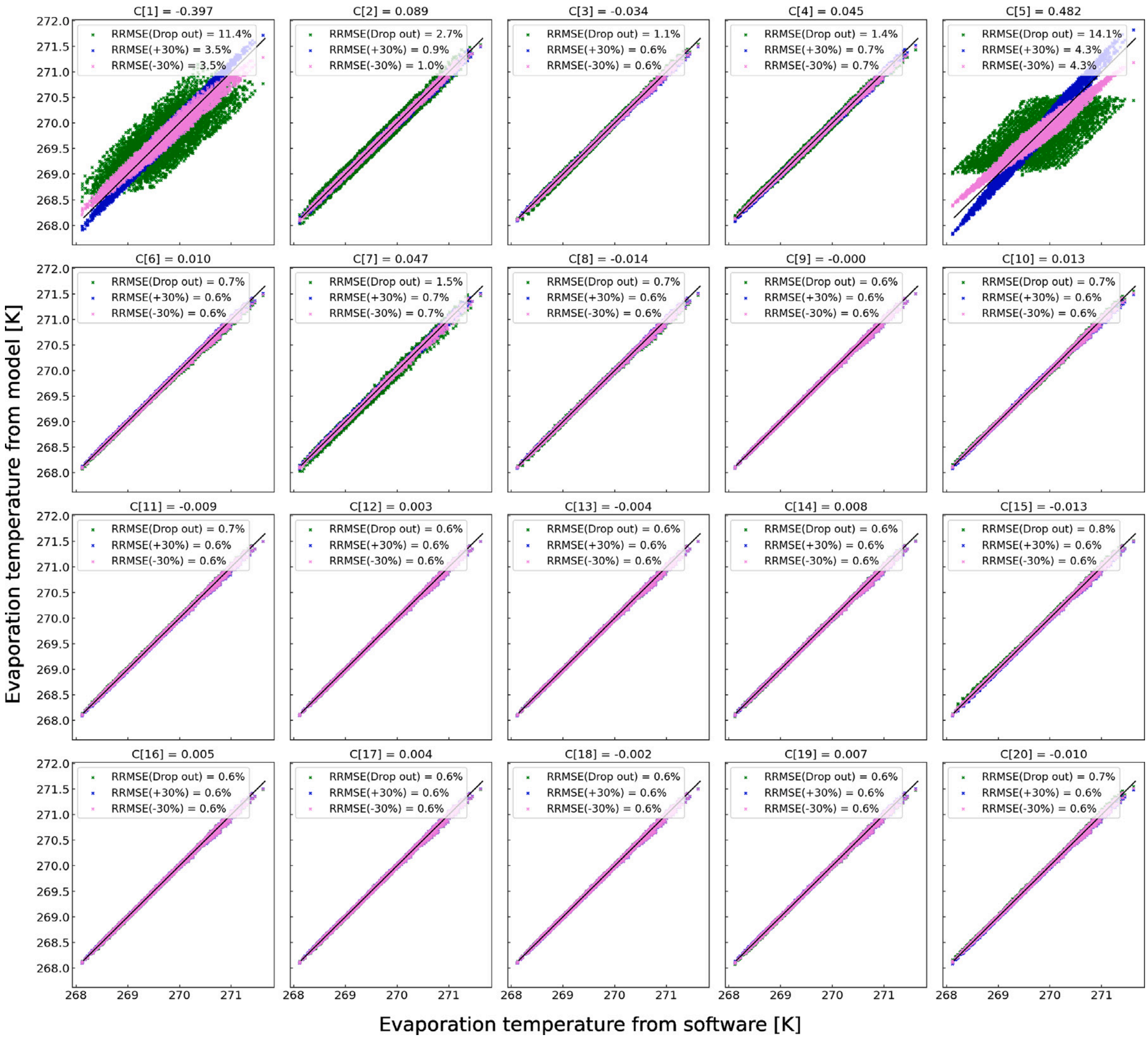

Fig. 7. Sensitivity analysis of evaporator model coefficients.

Table 6
Inputs and outputs of software and multivariate polynomial model for evaporator.

Variables	Software	Model
Mass flow rate of refrigerant	Output	Input(x_1)
Evaporator inlet enthalpy	Output	Input(x_2)
Outlet temperature of refrigerant	Output	Input(x_3)
Inlet temperature of brine	Input	Input(x_4)
Outlet temperature of brine	Input	Input(x_5)
Inlet quality to evaporator	Input	–
Superheating	Input	–
Cooling capacity	Input	–
Evaporation temperature	Output	Output(T_e)

subplot is increased by 30%, decreased by 30% and totally dropped out. The changed RRMSE is cataloged in the upper left corner of subplots in Fig. 7. The figure demonstrates the model is far more sensible to c_1 and c_5 compared to others. These two coefficients belong to x_1 and x_5, representing refrigerant mass flow rate and brine outlet temperature, which are more crucial for the accuracy of evaporator model.

3.3.3. Compressor multivariate polynomial regression model

To our experience, most of the already installed domestic heat pump systems lack power meters and mass flow meters. Two different multivariate polynomial regression models are developed separately for calculating refrigerant mass flow rate and compressor power consumption with different inputs, as shown in Table 7. Compressor frequency, evaporation temperature and superheating are taken as inputs for mass flow rate model according to Eqs. (12)–(13). While for power model, condensation temperature, evaporation temperature and compressor frequency are defined as inputs based on Eq. (14). Then the multivariate polynomial regression models for mass flow rate and power consumption are expressed as Eqs. (15) (16) respectively. In Eq. (15), b is regression constant coefficient. c_1 to c_9 are coefficients of polynomial terms. x_1 to x_3 represent three input variables: compressor frequency, evaporation temperature and suction temperature. $\dot{m}_{ref}$ is the output-compressor suction mass flow rate. In Eq. (16), b is regression constant coefficient. c_1 to c_9 are coefficients of polynomial terms. x_1 to x_3 represent three input variables, namely compressor frequency, evaporation temperature and condensation temperature. W is compressor power consumption, which is the output of the model. The data used for training and test the polynomial regression models come from a compressor manufacturer's software based upon the compressor type in actual heat pump. When the models are put into practice against the real database, evaporation and condensation temperatures are supplied from evaporator and condenser models, shown in Section 3.3.4.

$$\dot{m}_{ref} = \eta \cdot f_{comp} \cdot \frac{V}{v} \tag{12}$$

$$v = f(T_{suc}, P_e) \tag{13}$$

Table 7
Inputs and outputs of software and multivariate polynomial model for compressor.

Variables	Software	Power model	Mass flow rate model
Compressor frequency	Input	Input(x_1)	Input(x_1)
Evaporation temperature	Input	Input(x_2)	Input(x_2)
Condensation temperature	Input	Input(x_3)	–
Superheating	Input	–	Input(x_3)
Power consumption	Output	Output(W)	–
Suction mass flow rate	Output	–	Output(m_{ref})
Subcooling	Input	–	–

$$\begin{aligned} W &= \eta \cdot \frac{\kappa}{(\kappa-1)} \cdot P_c \cdot [(\frac{P_c}{P_e})^{\frac{\kappa-1}{\kappa}} - 1] \cdot V \cdot f_{comp} \\ &= f(P_c, P_e, f_{comp}) \\ &= f(T_c, T_e, f_{comp}) \end{aligned} \tag{14}$$

$$\begin{aligned} \dot{m}_{ref} &= b + \sum_{i=1}^{3} x_i c_i + \sum_{i=1}^{3} x_1 x_i c_{3+i} \\ &\quad + \sum_{i=2}^{3} x_2 x_i c_{5+i} + \sum_{i=3}^{3} x_3 x_i c_{6+i} \end{aligned} \tag{15}$$

$$\begin{aligned} W &= b + \sum_{i=1}^{3} x_i c_i + \sum_{i=1}^{3} x_1 x_i c_{3+i} \\ &\quad + \sum_{i=2}^{3} x_2 x_i c_{5+i} + \sum_{i=3}^{3} x_3 x_i c_{6+i} \end{aligned} \tag{16}$$

Both the mass flow rate model and power consumption model are optimized by ordinary least squares optimization method, and the RRMSE of the models are 0.5% and 0.2% respectively. The coefficients in the models are recorded in Figs. 8 and 9. The intercept term of the mass flow rate model is 10.12 and that of power consumption model is 0.458. Sensitivity analysis for each coefficient and polynomial term are investigated, and the modified RRMSE are registered in the upper left corner of subplots in Figs. 8 and 9 when only the coefficient listed on top of the subplot is increased by 30%, decreased by 30%, or totally dropped. From the analysis, it can be concluded that for mass flow rate multivariate polynomial regression model, compressor frequency is far more prominent compared to evaporation temperature and suction temperature. While for power consumption multivariate polynomial regression model, compressor frequency is the most important factor, and condensation pressure ranks the second. The model is almost not sensible to evaporation temperature since the RRMSE does not change when the coefficients of evaporation temperature changes. In this study, a scroll compressor is applied and the sensitivity analysis results for refrigerant mass flow rate and power consumption are quite similar as [44], where a critical analysis for scroll compressor characterization is conducted. In that study, the compressor power was also quite insensitive to evaporation temperature. For other types of compressors like reciprocating compressors, the result may differ and further investigation is needed.

3.3.4. Integrated heat pump multivariate regression model

In real circumstance, due to missing variable measurements, the subcomponent models cannot work independently to serve as soft sensors. For example, the compressor mass flow rate model cannot work due to the absence of evaporation temperature. So after evaporator, condenser and compressor models have been established, an integrated heat pump model is developed by combining the aforementioned subcomponent models into an iteration loop, which is demonstrated in Fig. 10. In this integrated model, the subcomponent models works collaboratively with output of certain model as inputs to other models.

The iteration loop initiates from a guess refrigerant mass flow rate, which serves as an input to the condenser model firstly. Together with four measurements, namely condenser inlet and outlet temperatures and water inlet and outlet temperatures, the condenser model yields condensation temperature and pressure. Based on condensation pressure and condenser outlet temperature, the enthalpy out of condenser is reckoned. In addition, other inputs to evaporator model are the guess refrigerant mass flow rate and three temperature measurements (evaporator outlet, brine inlet and outlet). As output of evaporator, evaporation temperature together with evaporator outlet temperature and compressor frequency are subsequently sent to compressor model that can calculate refrigerant mass flow rate. This mass flow rate is then compared with the initial guess refrigerant mass flow rate. If relative error of the two mass flow rates is smaller than threshold that is set to be 0.05%, it means the cycle gets a converged refrigerant mass flow rate. What is worthy of being mentioned is that until this point, temperatures and pressures of condensation and evaporation have been successfully compensated based on the accurately guess refrigerant mass flow rate and the refrigerant cycle has already been fixed. Next, compressor power consumption can be obtained from compressor model whose inputs are the compensated temperatures of condensation and evaporation, as well as measured compressor frequency. However, if the relative error between the calculated refrigerant mass flow rate from compressor model and the initial guess value is larger than the threshold, then a new guess mass flow rate is set to be the mean value of current guess mass flow rate and calculated mass flow rate. Then the calculation restarts until the iteration reaches a converged refrigerant mass flow rate.

In summary, this iterated cycle only takes available monitoring data measurements as inputs, and makes up the missing values that are paramount necessary to thoroughly understand about the thermodynamic cycle of the heat pumps.

3.4. Empirical model based on performance map

In reality, it might happen that neither history training data nor subcomponents data can be accessed. Out of this consideration, an empirical model based on performance map working as soft sensors is proposed in this study. Heat pump performance map is usually provided by heat pump manufacturer, which is normally a table containing the following parameters: compressor frequency, water mass flow rate, water inlet temperature, brine mass flow rate, brine inlet temperature, heating power and compressor power.

Fortunately, the heat pump performance is continuous with only smooth trends, so polynomial models are usually efficient functions to describe them. In this sense, compressors, which are the bases for the heat pump performance, are very well studied in such type of empirical models. The standard ANSI/ARI 540-2020 [45] proposes the commonly called AHRI polynomials, two polynomial models to accurately characterize the compressor energy consumption and mass flow rate as a function of the evaporating and condensing temperatures, which is expressed as Eq. (17), where c_1 to c_{10} are coefficients, S and D are the dew point temperatures at suction and discharge respectively, and X can be mass flow rate and power consumption. Unfortunately, both temperatures are only suitable parameters when the main objective is to characterize the compressor as a single component. However, these temperatures are dependent on boundary conditions at the evaporator and condenser side, so polynomials based on the external parameters, i.e., source/sink temperatures and compressor speed, should be able to characterize heat pump unit performance. This is the approach

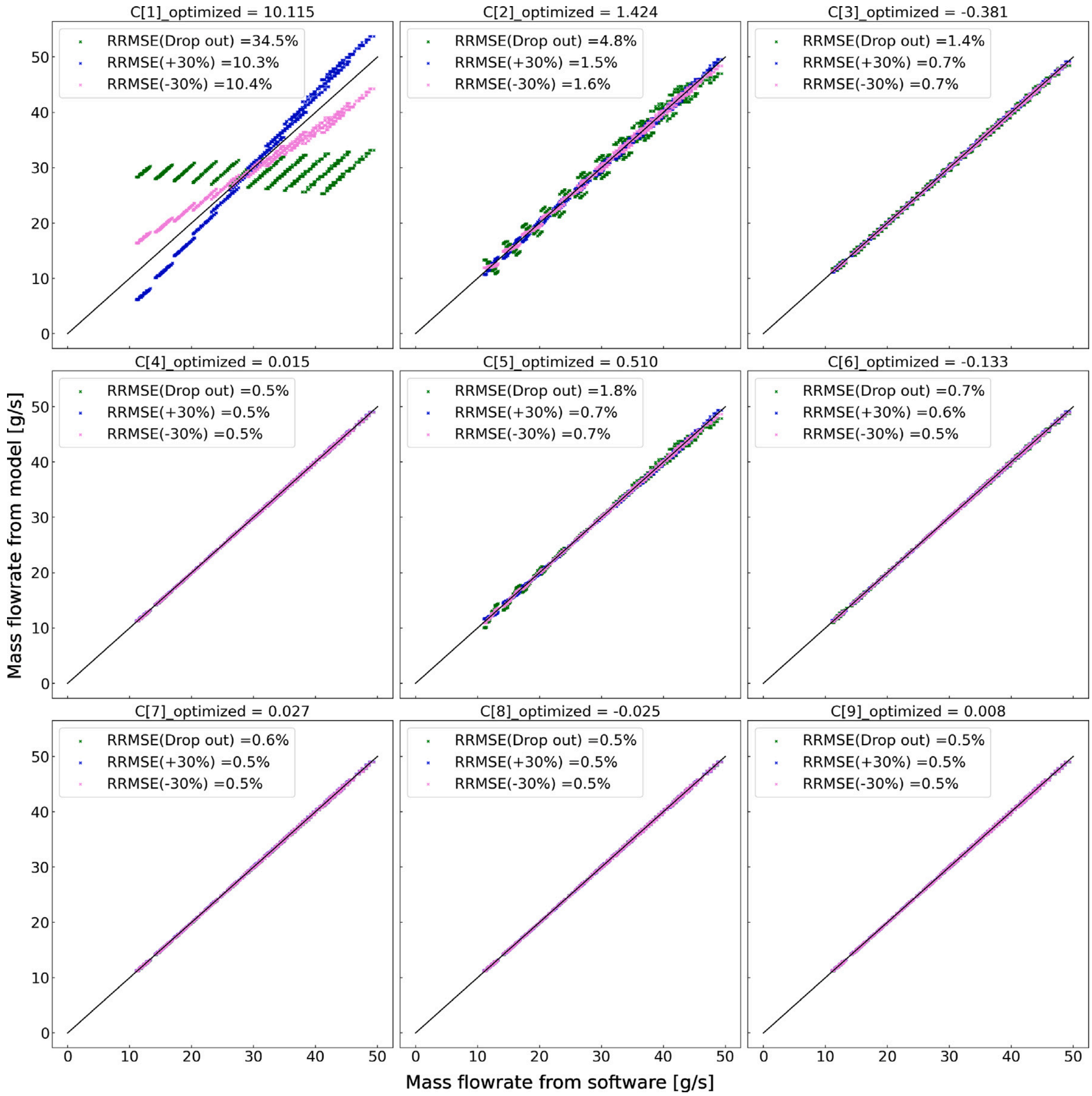

Fig. 8. Sensitivity analysis of mass flow rate estimation model coefficients.

selected for example in [46,47], where the authors supply suitable polynomials in order to characterize the performance – heating, cooling and compressor power – of a brine-to-water heat pump. The equations are Eq. (18), Eq. (19), Eq. (20), where c_i's are coefficients, T_{ei} is brine inlet temperature, T_{ci} is water inlet temperature, dT_e is the brine inlet and outlet temperature difference, dT_c is the water inlet and outlet temperature difference, and f_{comp} is compressor frequency.

Based on the information from heat performance map, the coefficients in Eq. (18) to Eq. (20) can be optimized through ordinary least squares optimization. Therefore, it is possible to predict the unit performance without requiring detailed information about its components or additional measurements.

$$X = c_1 + c_2 S + c_3 D + c_4 S^2 + c_5 SD + c_6 D^2 + c_7 S^3 + c_8 DS^2 + c_9 SD^2 + c_{10} D^3 \tag{17}$$

$$Q_c = f_{comp} \cdot (C_0 + C_1 \cdot T_{ei} + C_2 \cdot T_{ci} + C_3 \cdot T_{ei} \cdot T_{ci} + C_4 \cdot T_{ei}^2 + C_5 \cdot dT_e + C_6 \cdot dT_c + C_7 \cdot T_{ei} \cdot dT_e + C_8 \cdot T_{ci} \cdot dT_c) \tag{18}$$

$$Q_e = f_{comp} \cdot (C_0 + C_1 \cdot T_{ei} + C_2 \cdot T_{ci} + C_3 \cdot T_{ei} \cdot T_{ci} + C_4 \cdot T_{ei}^2 + C_5 \cdot dT_e + C_6 \cdot dT_c + C_7 \cdot T_{ei} \cdot dT_e + C_8 \cdot T_{ci} \cdot dT_c) \tag{19}$$

$$W = f_{comp} \cdot (C_0 + C_1 \cdot T_{ei}^2 + C_2 \cdot T_{ci}^2 + C_3 \cdot dT_e + C_4 \cdot dT_c + C_5 \cdot T_{ei} \cdot dT_e + C_6 \cdot f_{comp} \cdot T_{ei} + C_7 \cdot f_{comp} \cdot T_{ci} + C_8 \cdot f_{comp}^2 \cdot T_{ei} \cdot T_{ci} + C_9 \cdot f_{comp} \cdot dT_e + C_{10} \cdot \frac{1}{f_{comp}}) \tag{20}$$

Up to this point, temperatures and pressures of evaporation and condensation are still missing and need to be compensated. Two correlations regarding condenser and evaporator temperatures based on external variables are listed in Eqs. (21) and (22), where c_i's are coefficients, T_{co} is water outlet temperature, T_{eo} is brine outlet temperature and f_{comp} is compressor frequency. These correlations were obtained in the same framework and unit analyzed in [46,47]. They use the same approach of correlating the performance as a function of external variables but selecting condensing and evaporating temperatures as response variables.

$$T_c = T_{co} + c_0 + c_1 \cdot f_{comp} + c_2 \cdot (T_{co} - T_{eo}) \tag{21}$$

$$T_e = T_{eo} + c_0 + c_1 \cdot f_{comp} + c_2 \cdot (T_{co} - T_{eo}) \tag{22}$$

To optimize the coefficients in Eqs. (21) and (22), a calculation procedure is conducted as following:

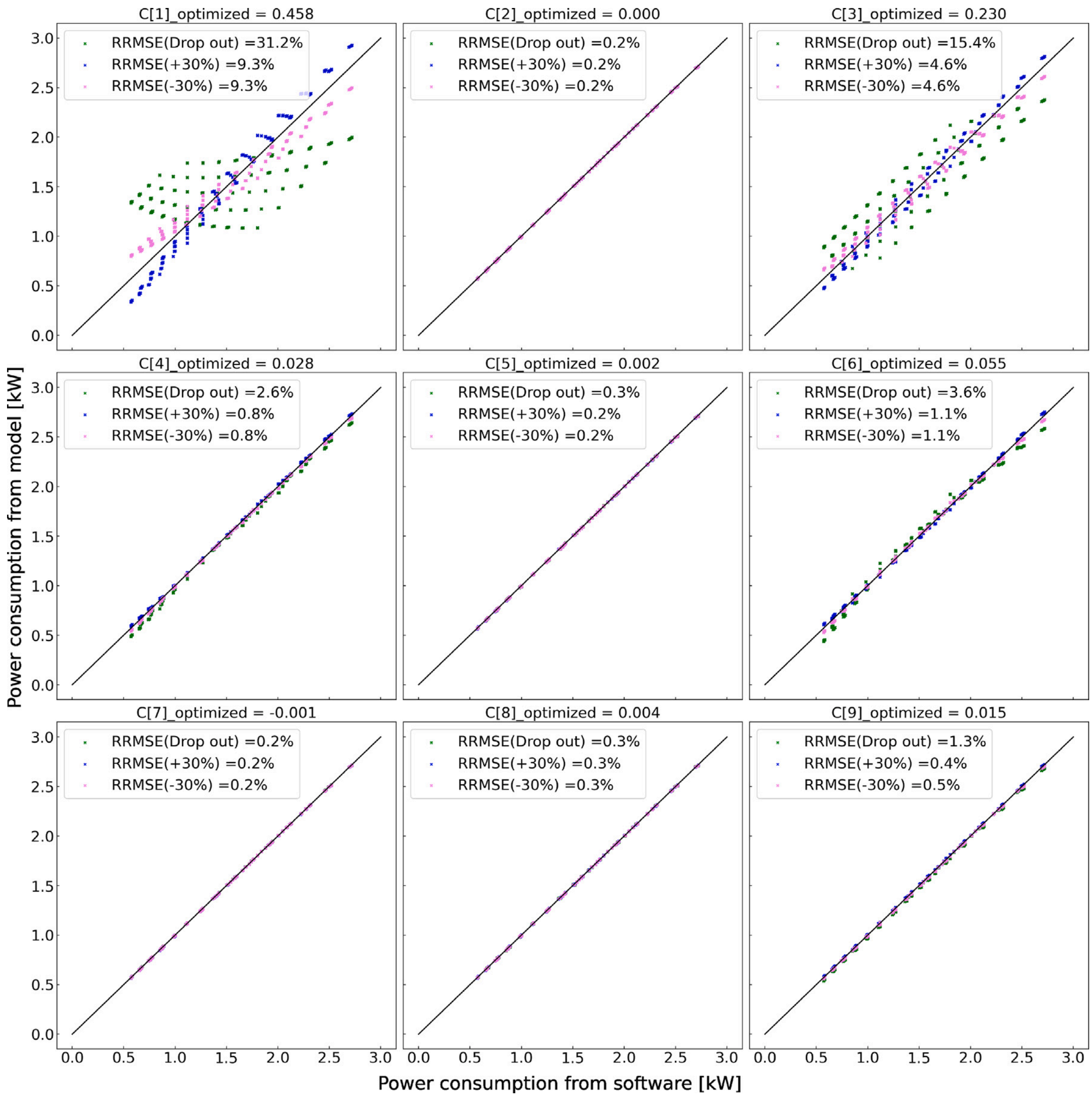

Fig. 9. Sensitivity analysis of compressor power consumption model coefficients.

1. Optimize the coefficients in Eq. (18) to Eq. (20) based on performance map.
2. Apply the optimized correlations in step 1 to compensate heating capacity, cooling capacity and compressor power for the real measurements.
3. Create pressure functions based on CoolProp:

$$P_c = f(T_c) \tag{23}$$

$$P_e = f(T_e) \tag{24}$$

T_c and T_e are expressed in correlations Eqs. (21) and (22) with 6 unknown coefficients.

4. Create enthalpy functions based on CoolProp:

$$h_{dis} = f(P_c, T_{dis}) \tag{25}$$

$$h_{liq} = f(P_c, T_{liq}) \tag{26}$$

$$h_{suc} = f(P_e, T_{suc}) \tag{27}$$

T_{dis}, T_{liq}, and T_{suc} are discharge temperature, liquid line temperature and suction temperature from real measurements. P_c and P_e are from step 3.

5. Create refrigerant mass flow rate functions based on energy balance:

$$\dot{m}_1 = Q_c/(h_{dis} - h_{liq}) \tag{28}$$

$$\dot{m}_2 = Q_e/(h_{suc} - h_{liq}) \tag{29}$$

$$\dot{m}_3 = W \cdot \eta/(h_{dis} - h_{suc}) \tag{30}$$

Q_c, Q_e and W are the calculated results from step 2. h_{dis}, h_{liq}, and h_{suc} are from step 4. η is compressor performance efficient.

6. Minimize the objective function:

$$MSE(\dot{m}_1, \dot{m}_2) + MSE(\dot{m}_1, \dot{m}_3) + MSE(\dot{m}_2, \dot{m}_3) \tag{31}$$

to solve the 6 unknown coefficients in step 3, and η in step 5.

To minimize the objective function in step 6, Sequential Least Squares Programming (SLSQP) is utilized. SLSQP is an effective optimization algorithm to minimize a function with several variables combining bounds, equality and inequality constraints [48]. In this study, condensation temperature is higher than evaporation temperature should be set as the constraint. After optimization, the coefficients in Eq. (21) are: $c_0 = 0.277$, $c_1 = 0.173$, $c_2 = -0.208$. The coefficients in Eq. (22) are: $c_0 = -0.270$, $c_1 = -0.117$, $c_2 = 0.094$. $\eta = 0.93$ in step 5.

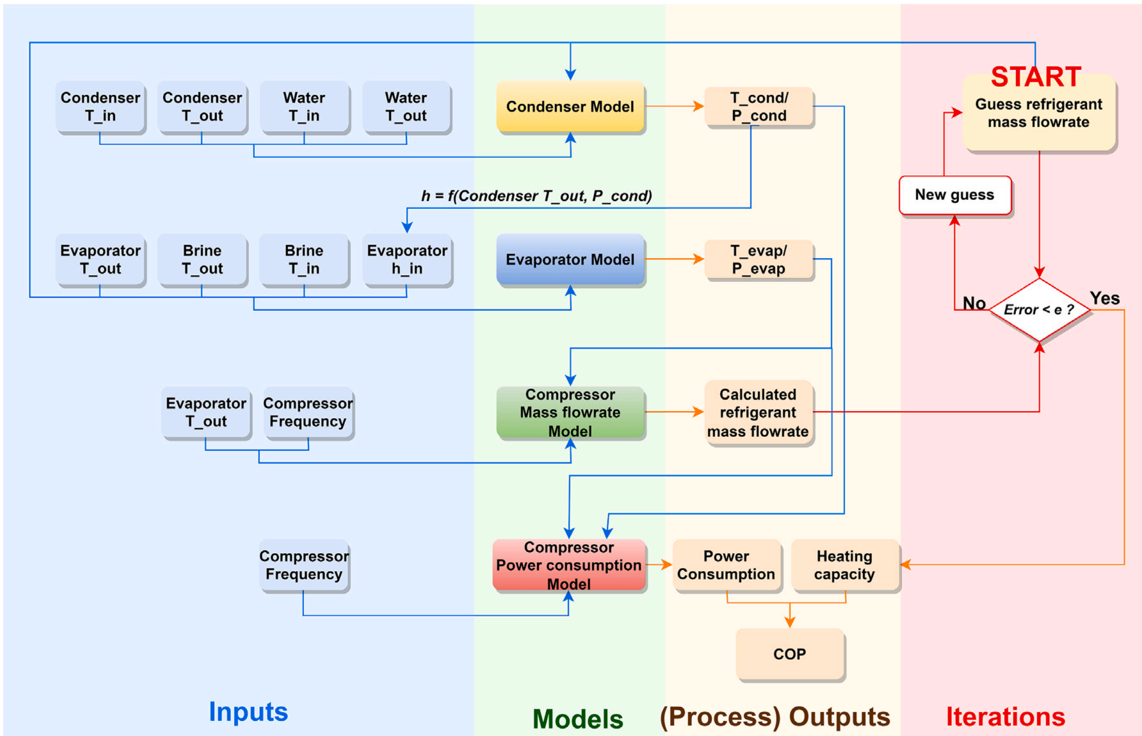

Fig. 10. Integrated heat pump model.

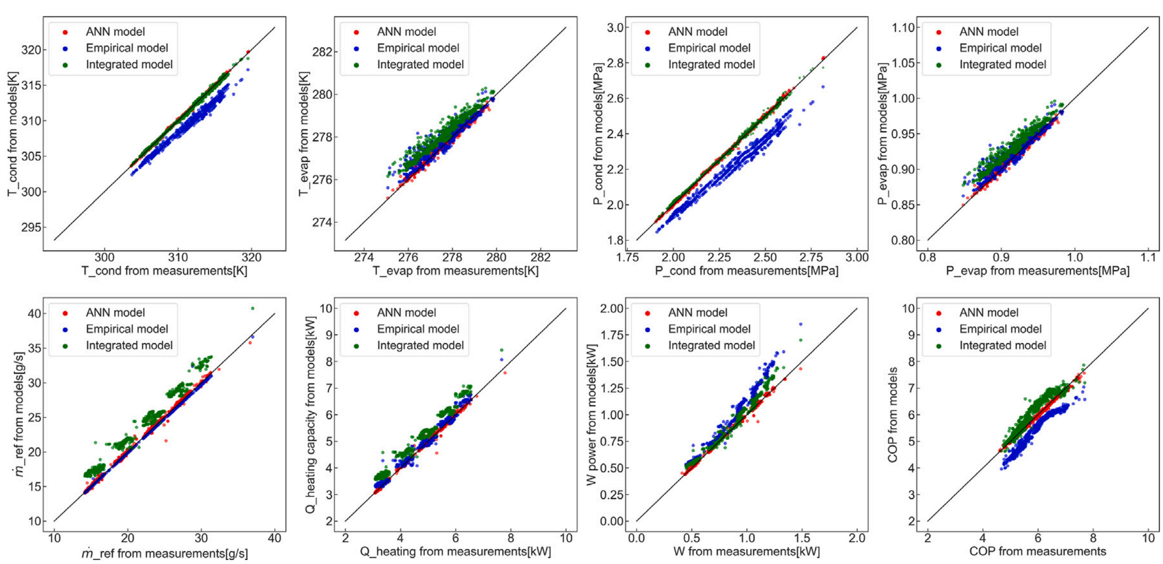

Fig. 11. Soft sensors against real sensors.

4. Results and discussion

4.1. Relative error

Fig. 11 depicts how the results from ANN model, integrated multivariate polynomial regression model(named as integrated model for convenience afterwards) and empirical model deviate from real measurements. COP is calculated based on heating capacity and compressor power consumption. To evaluate the accuracy of soft sensors with quantification, absolute error (AE) (for temperature soft sensors) and relative error (RE) (for the other soft sensors) are applied to measure the deviation between model result from real time monitoring data for every point. The expression of AE and RE is shown as Eqs. (32) and (33), where f_i is the value obtained from the model and y_i is the real value for data point i. The error distribution of different variables from the three models is displayed in Fig. 12. Concrete error distribution information regarding lower quantile, median value, and upper quantile are summarized in Table 8. The result shows that all the three models provides satisfied RE result, with lower and upper quantile of RE for all variables almost between $\pm 15\%$. Comparing the three models, ANN model has the most narrow spread of RE distribution for all variables than other two models. Besides, the median values of RE for ANN model is always close 0, suggesting that ANN model has a fairly accurate estimation for the compensated variables. By contrast, integrated model gives a positive median RE for all the variables which indicates it slightly overestimates all the variables. The empirical model

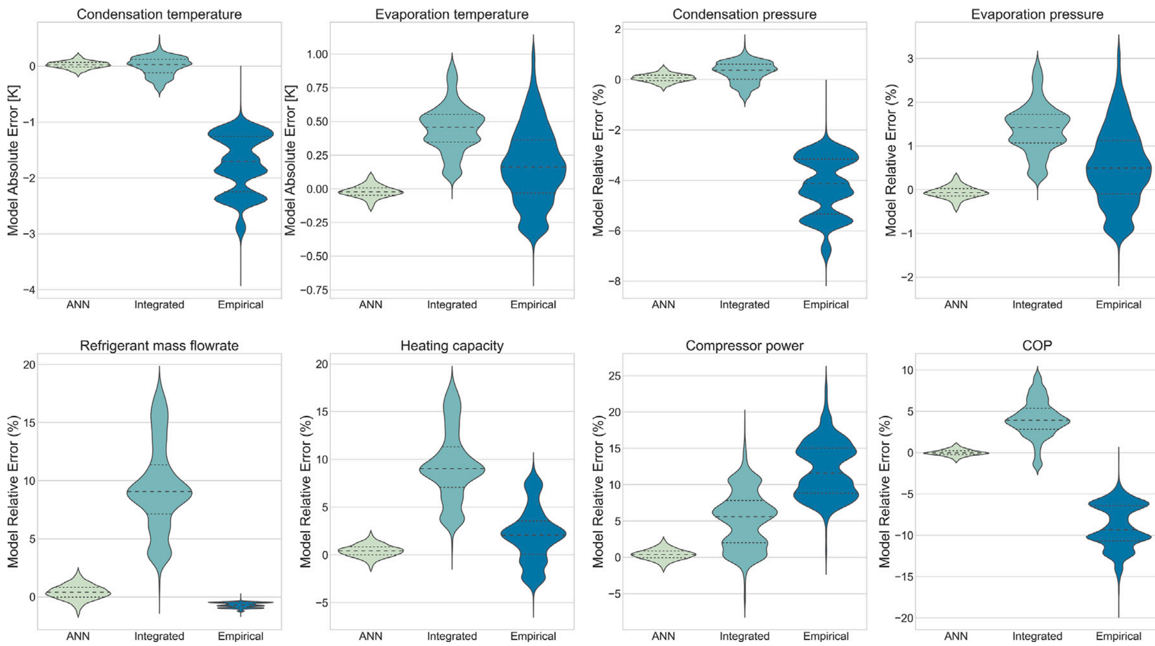

Fig. 12. Distribution of model error.

Table 8
Summary of model error.

Model Type	Quantiles	AE statistics (K)		RE statistics (%)				
		T_c	T_e	P_c	P_e	$\dot{m}_{ref}$	Q_c	W
	Upper quantile	0.07	0.01	0.21	0.11	0.70	0.67	0.98
ANN model	Median	0.03	−0.02	0.09	0.01	0.18	0.17	0.38
	Lower quantile	−0.02	−0.05	−0.01	−0.15	−0.34	−0.33	−0.17
	Upper quantile	0.12	0.55	0.61	1.72	11.41	11.36	7.77
Integrated model	Median	0.03	0.46	0.36	1.42	9.05	9.00	5.59
	Lower quantile	−0.12	0.35	0.01	1.07	7.03	6.95	1.98
	Upper quantile	−1.25	0.36	−3.15	1.14	−0.48	3.55	15.03
Empirical model	Median	−1.71	0.16	−4.12	0.50	−0.69	2.04	11.47
	Lower quantile	−2.25	−0.03	−5.33	−0.07	−0.91	−0.06	8.84

Table 9
Summary of model regression performance.

Model type	$RMSE$ (K)		$RRMSE$ (%)				
	T_c	T_e	P_c	P_e	$\dot{m}_{ref}$	Q_c	W
ANN model	0.09	0.09	0.25	0.29	1.14	1.13	1.37
Integrated model	0.18	0.53	0.51	1.65	9.29	9.25	6.29
Empirical model	1.83	0.36	4.58	1.13	0.89	3.29	14.25

has the widest spread of RE distribution for almost all variables but refrigerant mass flow rate. It is understandable that the empirical model has relatively low accuracy in general since it is developed based on limited information from performance map.

$$AE = f_i - y_i \tag{32}$$

$$RE = \frac{f_i - y_i}{y_i} \tag{33}$$

4.2. Overall regression performance

The distribution of RE only reveals the accuracy of soft sensors for every data point. To evaluate overall performance of the soft sensors, RMSE (for temperature soft sensors) and RRMSE (for the rest) are applied. Normally, the regression accuracy of a model is rated as excellent when RRMSE <10%, good when 10% <RRMSE <20%, fair when 20% <RRMSE <30%, and poor when 30% <RRMSE [49]. Table 9 summarizes the overall accuracy of all the data driven models in terms of RRMSE. It is clearly shown in Table 9 that ANN model has excellent performance with RRMSE for all soft sensors below 10%. Even though integrated model is not as precise as ANN model, its overall regression accuracy can still be rated as excellent. For the empirical model, the compressor power soft sensor has lowest accuracy, rated as good. The rest of the soft sensors from empirical model have RRMSE below 10% and can be ranked as excellent.

5. Conclusion

This study propose a multi-model approach for soft sensors in heat pump systems. ANN model, Integrated multivariate polynomial regression model, and empirical model are developed considering the availability of different additional information. Comparing these three models, ANN model can most accurately compensate the missing measurements for heat pump monitoring data, but it strongly depends on training data containing all indispensable parameters. That would mean additional sensors need to be installed at least in one installation to collect the required training data. Considering the limitation from ANN model, an integrated multivariate polynomial regression model is developed based on data from component software utilizing reverse engineering method. This model can compensate the missing variables with high accuracy. During the submodel development process, sensitivity analysis is conducted for different coefficients, which indicates the importance of the coefficients and variables quantitatively. This analysis can lead to development of simplified models by reducing it to the most relevant variables as the input parameters. Given the possibility that there is no corresponding manufacturer's software, the integrated multivariate polynomial regression model cannot be developed due to lack of data. Thus, an empirical model is proposed based on heat pump performance map that is available in most cases. Because the information that a performance map contains is limited, the soft sensors' performance is not as good as the other two models. Despite this, the empirical model can still have relatively good performance according to RRMSE metric. It is noteworthy that the empirical model unexpectedly has a relatively high accuracy for refrigerant mass flow rate estimation.

The three data-driven models for soft sensors provide multi-choices depending on the type of supplementary sources. Choosing any of these three methods can accomplish compensating missing measurements according to the availability of additional information. If all the required information is available, two or all three models can be developed in parallel so that they can do cross validation to provide more reliable results. Besides, the latency of developed soft sensors is negligible. To be more specific, the longest response time pertains to the integrated multivariate polynomial regression model, estimated to approximately 0.1s per set of results.

This study not only demonstrates the great potential of soft sensors to substitute several costly physical sensors, but also paves the way for exploiting the existing monitored data which is incomplete and lacks key measured parameters to provide innovative services to the end-users and manufacturers. Firstly, since soft sensors can provide accurate estimation of refrigerant mass flow rate, condensation pressures and temperatures, as well as evaporation pressures and temperatures, they can be used to monitor heat pump operational state in case of deviation from normal conditions to identify possible faults or performance degradation. Secondly, the data driven soft sensors can predict power consumption with high accuracy, which can be utilized to build smart control methods to establish communication with the local electric grid for electricity load management. Besides, the soft sensors developed in this study can be utilized as the basis for development of digital twins and consequently more advanced energy management of the heating systems.

CRediT authorship contribution statement

Yang Song: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Davide Rolando:** Conceptualization, Methodology, Software, Resources, Data curation, Writing – review & editing, Supervision, Project administration. **Javier Marchante Avellaneda:** Methodology, Writing – original draft, Writing – review & editing. **Gerhard Zucker:** Writing – review & editing, Supervision, Funding acquisition. **Hatef Madani:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Acknowledgments

Yang Song would like to thank Austrian Institute of Technology [grant number T6283] and the Chinese Scholarship Council (CSC) for providing the joint PhD funding.

References

[1] Luo J, Luo Z, Xie J, Xia D, Huang W, Shao H, et al. Investigation of shallow geothermal potentials for different types of ground source heat pump systems (GSHP) of Wuhan city in China. Renew Energy 2018;118:230–44.
[2] Rybach L, Sanner B. Geothermal heat pump development: trends and achievements in Europe. In: Perspectives for geothermal energy in Europe. World Scientific; 2017, p. 215–53.
[3] IEA. Energy technology perspectives. 2020, https://www.iea.org/reports/energy-technology-perspectives-2020. [Accessed 2020].
[4] Wei Z, Ren F, Yue B, Ding Y, Zheng C, Li B, et al. Data-driven application on the optimization of a heat pump system for district heating load supply: A validation based on onsite test. Energy Convers Manage 2022;266:115851.
[5] Lun YV, Tung SD. Heat pump configuration. In: Heat pumps for sustainable heating and cooling. Springer; 2020, p. 125–37.
[6] Stephant J, Charara A, Meizel D. Virtual sensor: Application to vehicle sideslip angle and transversal forces. IEEE Trans Ind Electron 2004;51(2):278–89.
[7] Jordaan E, Kordon A, Chiang L, Smits G. Robust inferential sensors based on ensemble of predictors generated by genetic programming. In: International conference on parallel problem solving from nature. Springer; 2004, p. 522–31.
[8] Khatibisepehr S, Huang B, Khare S. Design of inferential sensors in the process industry: A review of Bayesian methods. J Process Control 2013;23(10):1575–96.
[9] Nakhaei M, Akrami M, Gheibi M, Daniel Urbina Coronado P, Hajiaghaei-Keshteli M, Mahlknecht J. A novel framework for technical performance evaluation of water distribution networks based on the water-energy nexus concept. Energy Convers Manage 2022;273:116422.
[10] Zambonin G, Altinier F, Corso L, Sessolo M, Beghi A, Susto GA. Soft sensors for estimating laundry weight in household heat pump tumble dryers. In: 2018 IEEE 14th international conference on automation science and engineering. 2018, p. 774–9. http://dx.doi.org/10.1109/COASE.2018.8560522.
[11] Yuan X, Gu Y, Wang Y, Yang C, Gui W. A deep supervised learning framework for data-driven soft sensor modeling of industrial processes. IEEE Trans Neural Netw Learn Syst 2020;31(11):4737–46. http://dx.doi.org/10.1109/TNNLS.2019.2957366.
[12] Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. Comput Chem Eng 2009;33(4):795–814.
[13] Varlamis I, Sardianos C, Chronis C, Dimitrakopoulos G, Himeur Y, Alsalemi A, et al. Smart fusion of sensor data and human feedback for personalized energy-saving recommendations. Appl Energy 2022;305:117775.
[14] Ke W, Huang D, Yang F, Jiang Y. Soft sensor development and applications based on LSTM in deep neural networks. In: 2017 IEEE symposium series on computational intelligence. IEEE; 2017, p. 1–6.
[15] Fan Y, Tao B, Zheng Y, Jang S-S. A data-driven soft sensor based on multilayer perceptron neural network with a double LASSO approach. IEEE Trans Instrum Meas 2019;69(7):3972–9.
[16] Wang K, Shang C, Liu L, Jiang Y, Huang D, Yang F. Dynamic soft sensor development based on convolutional neural networks. Ind Eng Chem Res 2019;58(26):11521–31.
[17] Herceg S, Andrijić ŽU, Bolf N. Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models. Chem Eng Res Des 2019;149:95–103.
[18] Zheng J, Song Z. Semisupervised learning for probabilistic partial least squares regression model and soft sensor application. J Process Control 2018;64:123–31.
[19] Xibilia MG, Latino M, Marinković Z, Atanasković A, Donato N. Soft sensors based on deep neural networks for applications in security and safety. IEEE Trans Instrum Meas 2020;69(10):7869–76. http://dx.doi.org/10.1109/TIM.2020.2984465.
[20] Cheng T, Harrou F, Sun Y, Leiknes T. Monitoring influent measurements at water resource recovery facility using data-driven soft sensor approach. IEEE Sens J 2018;19(1):342–52.
[21] Kim W. Fault detection and diagnosis for air conditioners and heat pumps based on virtual sensors. Purdue e-Pubs; 2013.
[22] Ploennigs J, Ahmed A, Hensel B, Stack P, Menzel K. Virtual sensors for estimation of energy consumption and thermal comfort in buildings with underfloor heating. Adv Eng Inform 2011;25(4):688–98.
[23] Kim W, Braun JE. Development and evaluation of virtual refrigerant mass flow sensors for fault detection and diagnostics. Int J Refrig 2016;63:184–98.

[24] Li H, Braun JE. Virtual refrigerant pressure sensors for use in monitoring and fault diagnosis of vapor-compression equipment. HVAC&R Res 2009;15(3):597–616.
[25] Kim W, Braun JE. Virtual refrigerant mass flow and power sensors for variable-speed compressors. Purdue e-Pubs; 2012.
[26] Bellanco I, Fuentes E, Vallès M, Salom J. A review of the fault behavior of heat pumps and measurements, detection and diagnosis methods including virtual sensors. J Build Eng 2021;39:102254.
[27] Li H, Braun JE. Decoupling features and virtual sensors for diagnosis of faults in vapor compression air conditioners. Int J Refrig 2007;30(3):546–64.
[28] Kim W, Braun JE. Development, implementation, and evaluation of a fault detection and diagnostics system based on integrated virtual sensors and fault impact models. Energy Build 2020;228:110368.
[29] Larsen RD. Box-and-whisker plots. J Chem Educ 1985;62(4):302.
[30] Hintze JL, Nelson RD. Violin plots: a box plot-density trace synergism. Amer Statist 1998;52(2):181–4.
[31] Fannou J-LC, Rousseau C, Lamarche L, Kajl S. Modeling of a direct expansion geothermal heat pump using artificial neural networks. Energy Build 2014;81:381–90.
[32] Gang W, Wang J. Predictive ANN models of ground heat exchanger for the control of hybrid ground source heat pump systems. Appl Energy 2013;112:1146–53.
[33] Longo GA, Righetti G, Zilio C, Ortombina L, Zigliotto M, Brown JS. Application of an Artificial Neural Network (ANN) for predicting low-GWP refrigerant condensation heat transfer inside herringbone-type Brazed Plate Heat Exchangers (BPHE). Int J Heat Mass Transfer 2020;156:119824.
[34] Wang X, Li Y, Yan Y, Wright E, Gao N, Chen G. Prediction on the viscosity and thermal conductivity of hfc/hfo refrigerants with artificial neural network models. Int J Refrig 2020;119:316–25.
[35] Gupta D. Fundamentals of deep learning – Activation functions and when to use them? 2020, https://www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/. [Accessed 30 January 2020].
[36] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.
[37] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13(2).
[38] Paper D, Paper D. Scikit-learn classifier tuning from simple training sets. In: Hands-on scikit-learn for machine learning applications: Data science fundamentals with python. Springer; 2020, p. 137–63.
[39] Canfora G, Di Penta M. New frontiers of reverse engineering. In: Future of software engineering. IEEE; 2007, p. 326–41.
[40] Chikofsky EJ, Cross JH. Reverse engineering and design recovery: A taxonomy. IEEE Softw 1990;7(1):13–7.
[41] Bell IH, Wronski J, Quoilin S, Lemort V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. Ind Eng Chem Res 2014;53(6):2498–508.
[42] Dismuke C, Lindrooth R. Ordinary least squares. In: Methods and designs for outcomes research, vol. 93. American Society of Health-System Pharmacists Bethesda, MD; 2006, p. 93–104.
[43] Despotovic M, Nedic V, Despotovic D, Cvetanovic S. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. Renew Sustain Energy Rev 2016;56:246–60.
[44] Marchante-Avellaneda J, Corberan JM, Navarro-Peris E, Shrestha SS. A critical analysis of the AHRI polynomials for scroll compressor characterization. Appl Therm Eng 2022;119432.
[45] ANSI/AHRI. 2020 Standard for performance rating of positive displacement refrigerant compressors. 2020, https://www.ahrinet.org/search-standards. [Accessed 2020].
[46] Corberán JM, Cazorla-Marín A, Marchante-Avellaneda J, Montagud C. Dual source heat pump, a high efficiency and cost-effective alternative for heating, cooling and DHW production. Int J Low-Carbon Technol 2018;13(2):161–76.
[47] Corberán JM, Cazorla-Marín A, Marchante-Avellaneda J, Montagud-Montalvá C, Masip Sanchis X. Modelling and energy analysis of a dual source heat pump system in an office building. In: International conference on sustainable energy technologies. Università di Bologna; 2017.
[48] Kraft D. A software package for sequential quadratic programming. In: Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt. 1988.
[49] Li M-F, Tang X-P, Wu W, Liu H-B. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Convers Manage 2013;70:139–48.