Energy Proceedings
Vol 41, 2024
ISSN 2004-2965

Development and validation of data-driven soft sensors for heat pumps

Yang Song[1], Davide Rolando[1], Javier Marchante Avellaneda[2], Gerhard Zucker[3], Hatef Madani[1,3*]
1 KTH Royal Institute of Technology, Stockholm, 10044, Sweden
2 Universitat Politcnica de Valncia, Valencia, 46022, Spain
3 Austrian Institute of Technology, Vienna, A-1210, Austria
(*Corresponding Author: hatef.madani@energy.kth.se)

ABSTRACT

Modern heat pump systems often come equipped with sensors, enabling the collection of substantial operational data. However, many residential heat pumps installed in preceding decades lack pressure sensors, energy meters, or mass flow meters, primarily due to financial limitations. As a result of these incomplete measurements, the direct analysis of the heat pump system's performance or the leveraging of the amassed data for inventive applications like prognosticating energy consumption, detecting and diagnosing faults, and implementing intelligent control becomes challenging.

In existing literature, the focus of soft sensors in heat pump systems has been on estimating a single parameter. This approach, however, overlooks the reality that multiple parameters are often missing due to the lack of all-encompassing physical meters and sensors. Furthermore, current soft sensor models are typically developed using inputs such as compressor power consumption, pressures, evaporation, and condensation temperatures. These inputs, unfortunately, tend to be inaccessible within existing heat pump monitoring installations.

In practice, it is a challenge to compensate for several critical measurements, encompassing mass flow rate, pressures, power consumption, and heating capacity, by using only commonly available sensors such as secondary loop temperatures and compressor frequency are available. Currently, there is a notable gap in research concerning this practical issue.

To address the problems associated with inadequate measurements, this study presents the development and validation of soft sensors based on a data-driven approach, which can compensate for the parameters often unavailable with data collected from a limited number of commonly used sensors. Each component model employs a multivariate polynomial regression that calculates the evaporation temperature, condensation temperature, mass flow rate, and compressor power consumption, respectively. Subsequently, we present an integrated heat pump model that combines these component models into a comprehensive heat pump model.

Finally, we validate the data-driven model against field test installations, demonstrating its accuracy with a relative root mean squared error (RRMSE) ranging from 10% to 20%.

Keywords: heat pump, data-driven, soft sensor, empirical model, polynomial regression

NONMENCLATURE

Abbreviations	
RMSE	Root mean squared error
RRMSE	Relative root mean squared error
LMTD	Logarithmic mean temperature difference
Symbols	
Q_c	Heating capacity in condenser
P_c	Pressure in condenser
T_c	Temperature in condenser
T_e	Temperature in evaporator
$\dot{m}_{ref}$	Mass flow rate of refrigerant
W	Compressor power consumption

1. INTRODUCTION

Heat pumps can reduce emissions in the heating and cooling sector and will play an inevitable role in the transition to low-carbon energy [1]. According to International Energy Agency (IEA) forecast, the number

This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

of heat pumps installed worldwide will more than triple by 2030 compared to 2020 [2]. Modern heat pump systems generate a large amount of data on daily basis and proper use of the information collected in the measurements facilitates the development of different innovative services, like prediction of energy consumption, fault detection and diagnosis, and smart control. However, the collected data is usually incomplete due to economic or technical barriers. To be more concrete, there are often no pressure sensors, energy meters, or mass flow meters in existing domestic heat pumps installed in the last decades due to relatively high costs. This prevents the utilization of the monitoring data for direct analysis of heat pump performance. Therefore, it is quite important to compensate for the missing information by coupling it with other resources.

To address the problems associated with inadequate measurements, this paper presents the development and validation of soft sensors based on a data-driven approach, which can compensate for the parameters often unavailable with data collected from a limited number of commonly used sensors. Previous studies mainly focused on the development of one certain kind of soft sensor applied in a heat pump system. A refrigerant mass flow rate soft sensor is developed for the purpose of fault detection and diagnostics in [3], and the inputs include compressor frequency, power consumption, temperatures, and pressures at suction and discharge points. Another research by the same authors in[4] demonstrates a mass flow rate soft sensor by taking suction and ambient temperatures, high and low pressures on refrigerant loop as inputs. Besides, there is also research targeting compressor power soft sensors. In [5], the compressor power is estimated based on the measurements of suction temperature and pressure, evaporation, and condensation temperatures. Pressure soft sensors are also investigated with evaporation temperature and condensation temperature as inputs[6].

However, in existing references, soft sensors in heat pump systems are always designed to estimate one single parameter, while in reality, several parameters are normally missing because of the lack of multiple physical meters and sensors. Moreover, the current soft sensor models are usually developed based on some inputs like compressor power consumption, pressures, evaporation, and condensation temperatures that are generally missing in existing heat pump monitoring systems. In practice, multiple missing measurements (mass flow rate, pressures, power consumption, heating capacity) need to be compensated only based on commonly used sensors like secondary loop temperatures and compressor frequency. While the research in terms of this practical issue is still blank.

To fill the above-mentioned research gap, this paper proposes to explore the application of data-driven soft sensors in heat pump systems through the following steps:

- Development of data-driven soft sensors based on manufacturer's data, not requiring additional installation costs.
- Development of models based on secondary loop temperature data (i.e., brine inlet and outlet temperatures, water inlet and outlet temperatures) and compressor frequency to compensate for condensation and evaporation temperatures and pressures, refrigerant mass flow rate, and compressor power.
- Validation of the performance of the data-driven soft sensor model against field measurements.

The paper is structured as follows: Section 2 describes the methodology, including the data analysis from the heat pump operation database that motivated this work. Apart from this, data from a field test installation is analyzed and pre-processed and the data is used to build and validate the model. Then the model development process is introduced in detail. In section 3, the results of the soft sensors are demonstrated. Finally in Section 4 main conclusions are drawn.

2. METHODOLOGY

2.1 Database analysis

This analysis of the monitoring data from a large number of existing heat pump units that were installed decades ago in Sweden showed that compressor frequency and temperatures of suction, discharge, water inlet, water outlet, brine inlet, and brine outlet are commonly measured. This is mainly due to the fact that temperature sensors are inexpensive, easy to install, and easy to maintain. However, the measurements are typically unavailable from mass flow meters, compressor power meters, and pressure sensors, since these meters and sensors are often not installed mainly due to the extra costs involved.

The situation of missing measurement data in the database motivated the development of data-driven soft sensors that can be applied to the system analysis of thousands of existing heat pump installations.

2.2 Field test data analysis and preprocessing

In order to validate the accuracy of the model developed in this study, one field test heat pump unit is specially equipped with some additional sensors, such as energy meters and pressure sensors.

In the data set, some data points deviate significantly from the rest of the dataset or the expected behavior, which are called outliers. There are many reasons for the presence of outliers in a data set, including errors in the measurement, recording, or sampling process; abnormal but true data; or incorrectly reported data. Before model development, outliers are detected and cleaned based on the interquartile range method [7].

Since the variables with different scales have an unequal influence on the regression result, the dataset is normalized with the min-max method. This approach is based on a linear transformation, where variables are rescaled to zero and one. The normalization calculation equation is expressed as Eq.(1), where $x_{i,normalized}$ represents for the normalized feature values, x_i is the original value, x_{min} and x_{max} are the minimum and maximum values of feature x respectively [8].

$$x_{i,normalized} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

2.3 Model development

The data used for developing the models is firstly obtained from the software of component manufacturers, which is usually available for design and dimensioning purposes. Then the extracted data is used to optimize the model coefficients for the components: condenser, evaporator, and compressor. The details of each component model are described in sections 2.3.1, 2.3.2, and 2.3.3.

<u>2.3.1 Condenser multivariate polynomial regression model</u>

As the condensation temperature and pressure are usually not measured in existing domestic heat pumps, the condenser model is developed to calculate these two variables. The direct output of the condenser model is condensation temperature, based on which the condensation pressure is calculated from the thermophysical properties software CoolProp[9]. According to heat transfer calculations in heat exchangers and the available measurements in heat pump systems, Eq.(2) to Eq.(6), the inputs and output of the model are originally selected as shown in Table 1. In previous research [10], a sensitivity analysis is conducted and the predominated inputs are figured out. The result shows that with only water inlet and outlet temperatures as inputs, the model can estimate the condensation temperature accurately. Fewer input variables are required for the model means a smaller number of sensors is needed. Thus, the model can be applicable to more heat pump installations with low requirements for sensor installations. Out of this consideration, the condenser model is expressed as a multivariate polynomial regression that can match the output and inputs, as shown in Eq.(7), where b is the regression intercept, a_1 to a_5 are coefficients of polynomial terms, y_1, and y_2 indicate water inlet temperature and water outlet temperature respectively. The coefficients of the model are optimized using the data obtained from the manufacturer's heat exchanger software through the ordinary least squares optimization method [11].

$$Q_c = UA \cdot LMTD \tag{2}$$

$$Q_c = f\big(x_1, h(x_4, P_c), h(x_5, P_c)\big) \tag{3}$$

$$P_c = f(T_c) \tag{4}$$

$$LMTD = f(x_2, x_3, x_4, x_5) \tag{5}$$

$$UA \cdot f(x_2, x_3, x_4, x_5) = f\big(x_1, h(x_4, T_c), h(x_5, T_c)\big) \tag{6}$$

Table 1. Inputs and outputs of software and multivariate polynomial model for condenser

Variables	Software	Original Model	Simplified Model
Mass flow rate of refrigerant	Output	Input (x_1)	-
Inlet temperature of water	Input	Input (x_2)	Input (y_1)
Outlet temperature of water	Input	Input (x_3)	Input (y_2)
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)	Output (T_c)

$$T_c = b + a_1 y_1 + a_2 y_2 + a_3 y_1^2 + a_4 y_1 y_2 + a_5 y_2^2 \tag{7}$$

<u>2.3.2 Evaporator multivariate polynomial regression model</u>

Similarly, the evaporator model is designed to compensate for evaporation temperature or pressure. The input and output selection process is the same as for the condenser, except for refrigerant inlet temperature. Because this temperature is usually not monitored in practice, instead the inlet enthalpy to the evaporator is selected as input based on the assumption of the expansion process to be isenthalpic. The inputs and output of the model are originally selected as shown in Table 2. Also, to simplify the inputs to the model, a sensitivity analysis is performed, and finally, brine inlet and outlet temperatures and mass flow rate are three more essential variables to estimate evaporation temperature. Then the evaporator model is expressed as a multivariate polynomial regression matching the output and inputs, as shown in Eq.(8), where b is the regression intercept, a_1 to a_9 are coefficients of polynomial terms, y_1 to y_3 denote mass flow rate of refrigerant, brine inlet temperature and brine outlet temperature respectively. The coefficients of the optimized model are performed by ordinary least squares using data obtained from the manufacturer's heat exchanger software. After the model has been accomplished and is applied in real monitoring data, the mass flow rate of refrigerant is not available from measurement, since it is also a missing measurement to be compensated. The information of this variable is provided by the compressor model that will be introduced afterward in Section 2.3.3.

Table 2. Inputs and outputs of software and multivariate polynomial model for evaporator

Variables	Software	Original Model	Simplified Model
Mass flow rate of refrigerant	Output	Input (x_1)	Input (y_1)
Evaporator inlet enthalpy	Output	Input (x_2)	-
Outlet temperature of refrigerant	Output	Input (x_3)	-
Inlet temperature of brine	Input	Input (x_4)	Input (y_2)
Outlet temperature of brine	Input	Input (x_5)	Input (y_3)
Inlet quality to evaporator	Input	-	-
Superheating	Input	-	-
Cooling capacity	Input	-	-
Evaporation temperature	Output	Output (T_e)	Output (T_e)

$$T_e = b + \sum_{i=1}^{3} y_i\, a_i + \sum_{i=1}^{3} y_1\, y_i a_{3+i} + \sum_{i=2}^{3} y_2\, y_i a_{5+i} + \sum_{i=3}^{3} y_3\, y_i a_{6+i} \quad (8)$$

2.3.3 Compressor multivariate polynomial regression model

From the investigation of real heat pump monitoring data, the majority of the currently in use domestic heat pump systems are not equipped with power meters and mass flow meters. Therefore, two different polynomial models are developed respectively for refrigerant mass flow rate and compressor power consumption. The inputs of the refrigerant mass flow rate model are firstly defined according to Eq.(9) and Eq.(10). Then a sensitivity analysis is conducted for the input variables in a former study[10], and the result shows that evaporation temperature and compressor frequency play a more important role in calculating refrigerant mass flow rate, so these two variables are chosen as the inputs for the simplified model. The expression of the mass flow rate model is demonstrated in Eq.(11), where b is the regression constant, a_1 to a_2 are regression coefficients. For the power consumption model, there are multiple polynomial models available in the literature[12], [13]. In this study, the model is expressed as Eq.(12). The inputs and outputs of compressor models are listed in Table 3. During the model development phase, the optimization of the coefficients is based on the data from the compressor manufacturer's software. After the models have been finished and are being implemented in the real database from the heat pump, the evaporation temperature and condensation temperature are provided by the evaporator and condenser model, which has been illustrated in Section 2.3.1 and Section 2.3.2.

$$\dot{m}_{ref} = \eta \cdot f_{comp} \cdot \frac{V}{v} \quad (9)$$

$$v = f\left(T_{suc}(T_e, T_{super}), P_e(T_e)\right) \quad (10)$$

$$\dot{m} = b + a_1 f + a_2 T_e + a_3 f^2 + a_4 f T_e + a_5 T_e^2 \quad (11)$$

$$W = (a_2 f^2 + a_2 f + a_3) \cdot (a_4 + a_5 T_e + a_6 T_c + a_7 T_e^2 + a_8 T_e T_c + a_9 T_c^2) \quad (12)$$

Table 3. Inputs and outputs of software and multivariate polynomial model for compressor

Variables	Software	Power Model	Original mass flow rate model	Simplified mass flow rate model
Compressor frequency	Input	Input(f)	Input (f)	Input (f)
Evaporation temperature	Input	Input (T_e)	Input (T_e)	Input (T_e)
Condensation temperature	Input	Input (T_c)	-	-
Superheating	Input	-	Input (T_{super})	-
Power consumption	Output	Output (W)	-	-
Suction mass flow rate	Output	-	Output ($\dot{m}_{ref}$)	Output ($\dot{m}_{ref}$)

Subcooling	Input	-	-	-

2.3.4 Integrated heat pump multivariate polynomial regression model

The component models cannot estimate missing measurements independently because certain inputs to the models still missing in real measurements. For example, the compressor power consumption model cannot work because of the lack of evaporation and condensation temperatures. Therefore, after the component models have been finished, an integrated model is proposed by assembling the sub-models into an iteration loop, as seen in Figure 1. In the integrated model, the sub-models collaborate with each other.

The iterative loop starts with the guessed refrigerant mass flow rate, which is used as input to the evaporator model. Together with two measured values, the water inlet and outlet temperatures, the evaporator model derives the evaporation temperature and pressure. As the output of the evaporator, the evaporating temperature and compressor frequency are then sent to a compressor model that calculates the refrigerant mass flow rate. This mass flow rate is then compared to the initially guessed mass flow rate of the refrigerant. If the relative error between these two mass flow rates is less than a threshold set to 0.05%, a converged refrigerant mass flow rate is obtained for the cycle. After that, water inlet and outlet temperatures are taken as inputs to the condenser model so that the condensation temperature and pressure are compensated. By this current point, the condensing and evaporating temperatures and pressures have been successfully compensated based on an accurate guess of the refrigerant mass flow rate. Next, the power consumption of the compressor can be obtained from the compressor model with the previously calculated condensing and evaporating temperatures as well as the measured compressor frequency as inputs. In most cases, the relative error between the refrigerant mass flow rate calculated from the compressor model and the initial guess value is greater than the threshold

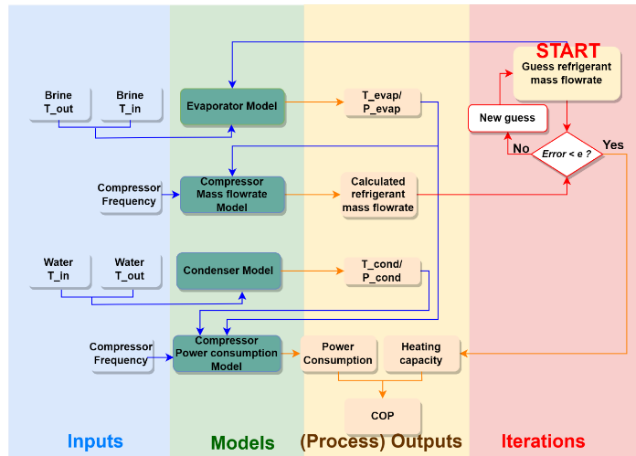

Fig. 1. Flowchart of integrated multivariate regression procedure for heat pump modeling

value, therefore a new guess mass flow rate is set. The new guess mass flow rate is calculated as the average of the current guess mass flow rate and the one currently calculated by the compressor model. Then the calculation is iterated until the refrigerant mass flow rate difference between the guess value and compressor model result is below the threshold.

In summary, this iterative loop only requires the temperatures of the secondary loop as well as the compressor frequency as inputs, which are easy and inexpensive variables to measure in domestic heat

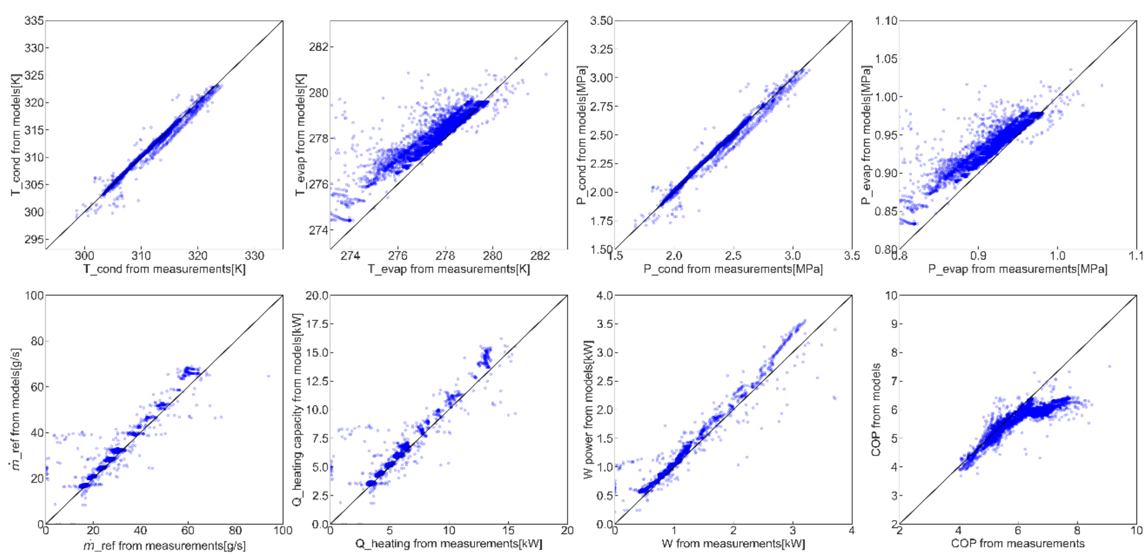

Fig. 2. Result of data-driven soft sensor models against real sensors

pumps. The output values of the cycle, i.e., pressures, refrigerant mass flow rate, compressor power consumption, etc., are more expensive measurements. The compensated values are essential for a thorough understanding of the thermodynamic cycle of a heat pump.

3. RESULT AND DISCUSSION

Figure 2 shows the deviation of the results of the data-driven models from the actual measurements. COP is calculated based on heating capacity and compressor power consumption. To evaluate the overall performance of the soft sensors, root mean squared error (RMSE, Eq.(13)), and relative root mean squared error (RRMSE, Eq.(14)) are applied to the individual compensated variables, where y_i is the ith actual measurement, f_i is the ith result from the model, and $\bar{y}$ is the average value of total data points. The temperatures are evaluated by RMSE while the rest variables are rated by RRMSE. The result of the model regression performance is summarized in Table 4. According to reference [14], a model can be rated as excellent if the RRMSE is less than 10%, and rated as good when the RRMSE is between 10% to 20%. Based on the combined analysis of Table 4 and Figure 2, the following conclusions can be drawn: (1) The model estimates the condensation temperature and pressure quite accurately, and the regression performance can be rated as excellent for condensation pressure. (2) There is a slight offset of the model result for evaporation temperature and pressure, and the overall estimation performance for this measurement is excellent. (3) The model performance for refrigerant mass flow rate, heating capacity, and power consumption are rated as good with the RRMSE moderately above 10%.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - f_i)^2}{n}} \tag{13}$$

$$RRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i - f_i)^2}{n}}}{\bar{y}} \tag{14}$$

Table 4. Summary of model regression performance

Regression metrics	T_c	T_e	P_c	P_e	$\dot{m}_{ref}$	Q_c	W
RMSE (K)	0.37	0.64	-	-	-	-	-
RRMSE(%)	-	-	1.00	2.14	10.28	10.44	12.59

4. CONCLUSIONS

This work proposes a simplified data-driven model for soft sensors based on a previous study [10] with only four secondary loop temperatures and frequency as inputs to compensate for multiple expensive or uneasy-to-measure measurements. The results show that the proposed data-driven model provides accurate estimations for the compensated measurements.

This work not only demonstrates the possibility of soft sensors to replace several relatively expensive meters and sensors but also opens the door for the heat pump incomplete database to be utilized to its full potential. For example, based on the good performance of virtual mass flow meters and pressure sensors, advanced fault detection and diagnosis, performance degradation services can be achieved. Possible refrigerant leaks, fouling, or frost in heat exchangers can be accurately detected by mass flow rate or pressure changes. Besides, nowadays, each heat pump installed in each building is an isolated island with no communication with other stakeholders. Since the power consumption, heating, and cooling capacity of heat pumps are accurately compensated by data-driven soft sensor models, they can be leveraged for intelligent energy control, and advanced energy management by establishing communications among heat pumps, local grid, district heating, and cooling networks. Future works will include a thorough sensitivity analysis to evaluate the benefits of the proposed approach for fault detection and performance degradation applications.

ACKNOWLEDGEMENT

The Austrian Institute of Technology and the Chinese Scholarship Council (CSC) are kindly acknowledged for co-funding the industrial Ph.D. project “Data-driven lab for building energy systems”. The Swedish Energy Agency is also kindly acknowledged for the support of this work through the Swedish Research Project 51501-1 “Digitalization and IoT technologies for Heat Pump systems” within the IEA HPT Annex 56 “Internet of Things for Heat Pumps”.

DECLARATION OF INTEREST STATEMENT

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. All authors read and approved the final manuscript.

REFERENCE

[1] M. Wahl, T. Droscher, J. Sprey, and A. Moser, “Modelling of Heat Pump Load Profiles for Grid Expansion Planning,” in 2018 53rd International Universities Power Engineering Conference (UPEC), Sep. 2018, pp. 1–6. doi: 10.1109/UPEC.2018.8541958.

[2]“Heat Pumps – Analysis,” IEA. https://www.iea.org/reports/heat-pumps (accessed Nov. 20, 2022).

[3] W. Kim and J. E. Braun, “Development and evaluation of virtual refrigerant mass flow sensors for fault detection and diagnostics,” International Journal of Refrigeration, vol. 63, pp. 184–198, Mar. 2016, doi: 10.1016/j.ijrefrig.2015.11.005.

[4] H. Li and J. E. Braun, “Decoupling features and virtual sensors for diagnosis of faults in vapor compression air conditioners,” International Journal of Refrigeration, vol. 30, no. 3, pp. 546–564, May 2007, doi: 10.1016/j.ijrefrig.2006.07.024.

[5] W. Kim and J. E. Braun, “Development, implementation, and evaluation of a fault detection and diagnostics system based on integrated virtual sensors and fault impact models,” Energy and Buildings, vol. 228, p. 110368, Dec. 2020, doi: 10.1016/j.enbuild.2020.110368.

[6] H. Li and J. E. Braun, “Virtual Refrigerant Pressure Sensors for Use in Monitoring and Fault Diagnosis of Vapor-Compression Equipment,” HVAC&R Research, vol. 15, no. 3, pp. 597–616, May 2009, doi: 10.1080/10789669.2009.10390853.

[7] R. D. Larsen, “Box-and-whisker plots,” Journal of Chemical Education, vol. 62, no. 4, p. 4, 1985.

[8] D. Singh and B. Singh, “Feature wise normalization: An effective way of normalizing data,” Pattern Recognition, vol. 122, p. 108307, Feb. 2022, doi: 10.1016/j.patcog.2021.108307.

[9] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort, “Pure and Pseudo-pure Fluid Thermophysical Property Evaluation and the Open-Source Thermophysical Property Library CoolProp,” Ind. Eng. Chem. Res., vol. 53, no. 6, pp. 2498–2508, Feb. 2014, doi: 10.1021/ie4033999.

[10] Y. Song, D. Rolando, J. Marchante Avellaneda, G. Zucker, and H. Madani, “Data-driven soft sensors targeting heat pump systems,” Energy Conversion and Management, vol. 279, p. 116769, Mar. 2023, doi: 10.1016/j.enconman.2023.116769.

[11] E. C. G. Chumney and K. N. Simpson, Methods and Designs for Outcomes Research. ASHP, 2006.

[12] W. Kim and J. E. Braun, “Virtual Refrigerant Mass Flow and Power Sensors for Variable-Speed Compressors,” p. 9, 2012.

[13] “AHRI 540-2020 (SI/I-P): Performance Rating of Positive Displacement Refrigerant Compressors and Compressor Units | AHRI.” https://www.ahrinet.org/search-standards/ahri-540-2020-sii-p-performance-rating-positive-displacement-refrigerant-compressors-and-compressor (accessed Nov. 25, 2022).

[14] P. D. Jamieson, J. R. Porter, and D. R. Wilson, “A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand,” Field Crops Research, vol. 27, no. 4, pp. 337–350, Nov. 1991, doi: 10.1016/0378-4290(91)90040-3.