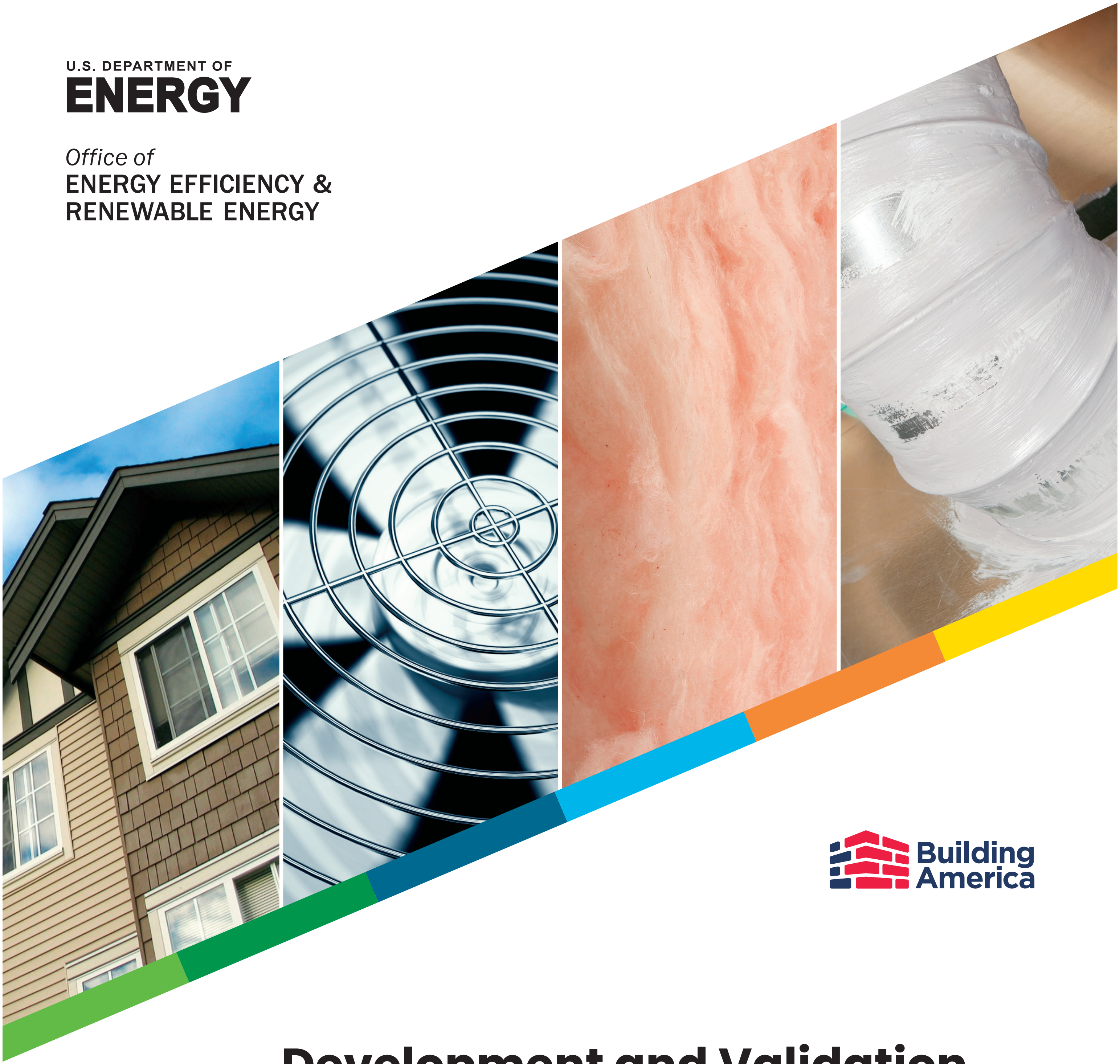
U.S. DEPARTMENT OF
ENERGY

Office of
ENERGY EFFICIENCY &
RENEWABLE ENERGY



Building America

Development and Validation of Home Comfort System for Total Performance Deficiency/Fault Detection and Optimal Comfort Control

May 2024

Disclaimer

EERE Award Number DE-EE0008697
Award Recipient: University of Oklahoma

Available electronically at Office of Scientific and
Technical Information website (www.osti.gov)

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062

OSTI www.osti.gov
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

NTIS www.ntis.gov
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Development and Validation of Home Comfort System for Total Performance Deficiency/ Fault Detection and Optimal Comfort Control

Prepared for:
U.S. Department of Energy Building America Program
Office of Energy Efficiency and Renewable Energy

Prepared by:
Yilin Jiang,[1] Kevwe Andrew Ejenakevwe,[1] Junke Wang,[1] Li Song (Principal Investigator),[1] Choon Yik Tang,[2] Gang Wang,[3] and Michael Brambley[4]

[1] School of Aerospace and Mechanical Engineering, University of Oklahoma
[2] School of Electrical and Computer Engineering, University of Oklahoma
[3] University of Miami
[4] Pacific Northwest National Laboratory

University of Oklahoma
660 Parrington Oval
Norman, OK 73019

May 2024

Suggested Citation
Jiang, Yilin, Kevwe Andrew Ejenakevwe, Junke Wang, Li Song, Choon Yik Tang, Gang Wang, and Michael Brambley. 2024. *Development and Validation of Home Comfort System for Total Performance Deficiency/ Fault Detection and Optimal Comfort Control*. Norman, OK. DOE/GO-102024-5884. www.nrel.gov/docs/fy24osti/85496.pdf.

This material is based upon work supported by the Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Building Technologies Office under Award Number EE0008697.

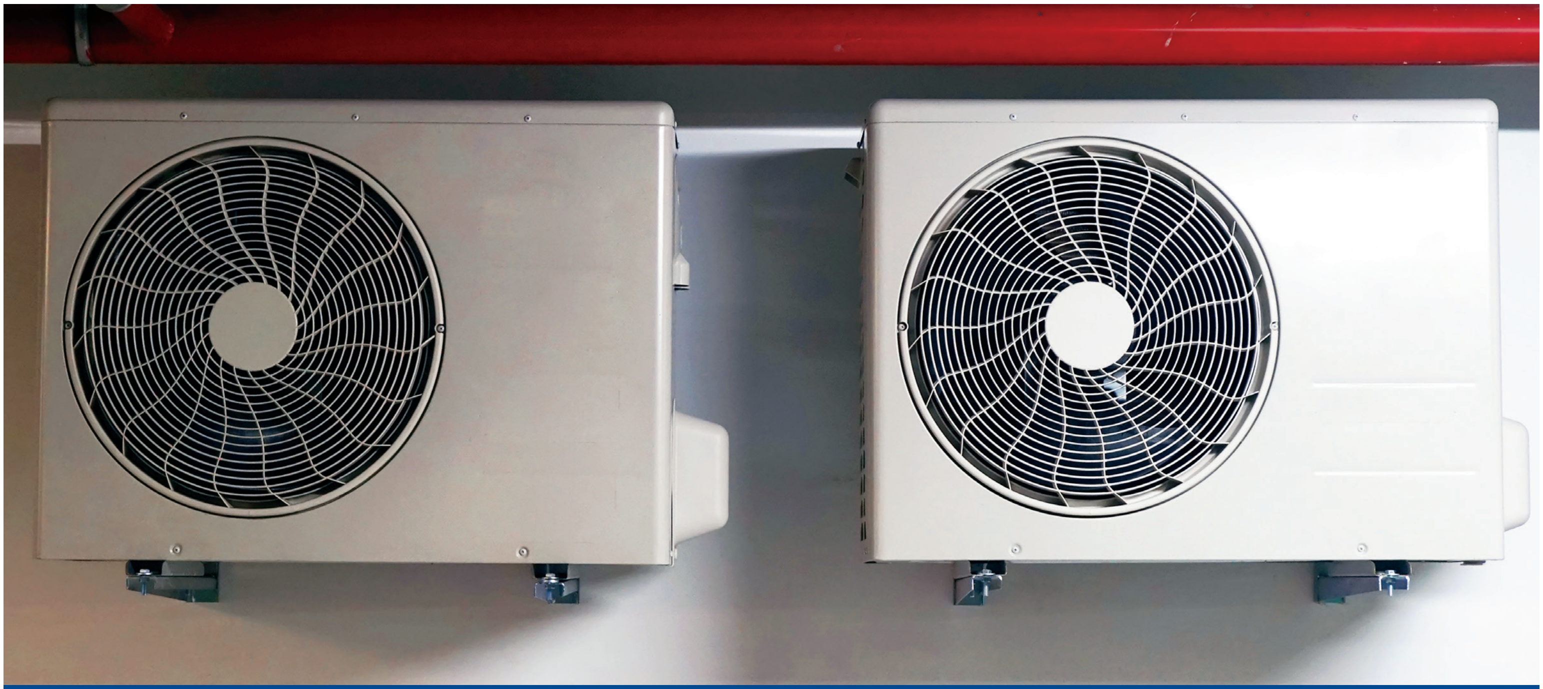
The work presented in this EERE Building America report does not represent performance of any product relative to regulated minimum efficiency requirements.

The laboratory and/or field sites used for this work are not certified rating test facilities. The conditions and methods under which products were characterized for this work differ from standard rating conditions, as described.

Because the methods and conditions differ, the reported results are not comparable to rated product performance and should only be used to estimate performance under the measured conditions.

Foreword

The U.S. Department of Energy (DOE) Building America Program has spurred innovations in building efficiency, durability, and affordability for more than 25 years. Elevating a clean energy economy and skilled workforce, this world-class research program partners with industry to leverage cutting-edge science and deployment opportunities to reduce home energy use and help mitigate climate change.



In cooperation with the Building America Program, the University of Oklahoma Team is one of many **Building America teams** working to drive innovations that address the challenges identified in the Program's **Research-to-Market Plan**.

This report, *Development and Validation of Home Comfort System for Total Performance Deficiency/Fault Detection and Optimal Comfort Control*, explores the potential of using smart thermostat data for optimal control and performance degradation detection for home ACs.

As the technical monitor of the Building America research, the National Renewable Energy Laboratory encourages feedback and dialogue on the research findings in this report as well as others. Send any comments and questions to building.america@ee.doe.gov.

Acknowledgments

The work presented in this report was funded by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Building Technologies Office.

The research was conducted by the University of Oklahoma Building America team, with guidance and support from Lucas Phillips and Eric Werling of DOE Building America, who we thank for their guidance and support of this project, including many communications and discussions.

The authors also thank the following people for their contributions to this effort:

- Mr. Sina Shahandeh, former vice president of data science at ecobee
- Ms. Jessica King, director of energy efficiency and customer programs at Oklahoma Gas & Electric.

Photo Credits

Cover, from left to right: Photos from iStock 182149008, 178447161, 184944590, 467972591

Page ii: Photo from Getty 539128158

Page v: Photo from Getty 486929964

List of Acronyms

AC	air conditioning
AFDD	automated fault detection and diagnosis
AFR	airflow reduction
CUSUM	cumulative sum
DOE	U.S. Department of Energy
EPT	expanded performance table
FDD	fault detection and diagnosis
IoT	Internet of Things
IP	indoor power
MPC	model predictive control
PNNL	Pacific Northwest National Laboratory
RA	return air
RCL	refrigerant charge level
SA	supply air
SADP	supply air dew-point
SAT	supply air temperature
TOU	time of use

List of Variables/Symbols

E_c	cooling effort
Q_{int}	internal heat gains
Q_{sol}	space air temperature increase representing solar impact on home
Q_{sys}	HVAC output
R_{vw}	variable thermal resistance dependent on wind speed
a_1–a_6, b_1–b_6, a'_1, a'_2, a'_3, b'_1, b'_2, c_1, c_2	empirical/regression coefficients
c_p	specific heat capacity at constant pressure
$\hat{i}$	predicted specific enthalpy
k_1–k_6	mathematical constants
$\hat{p}$	predicted indoor power
$\dot{q}$	heat flow rate or heat gains
q'_{vw}	heat transfer rate due to wind effect
u_i	internal activity schedules
u_s	HVAC system on/off signal
$\bar{\delta}$	deviation threshold
$\bar{\tau}$	duration threshold
τ_1, τ_2, τ_3	time constants
$\varnothing$	relative humidity
Δ	change/difference
A, B, d, X, Y, Φ	matrices
C	thermal capacity
E	HVAC system power use
F	cycling frequency
G	global horizontal irradiation
P	utility rate or power
Q	airflow rate
R	thermal resistance
S	solar irradiance
T	temperature
W	wind speed
i	measured specific enthalpy
k	discrete time
p	measured indoor power
t	time
ß	vector of unknown parameters
δ	deviation
ρ	density
τ	duration
ω	humidity ratio

List of Subscripts

dp	dew point
est	estimated
ewb	entering wet bulb
i	internal
ie	interior envelope
in	indoor
lb	lower bound
mea55	measured with SADP at 55°F
meas	measured
o	outdoor
odb	outdoor dry bulb
pred	predicted
predEPT	predictive model trained with EPT data
ra	return air
sa	supply air
sol	solar
sys	system
ub	lower bound
ve	virtual envelope

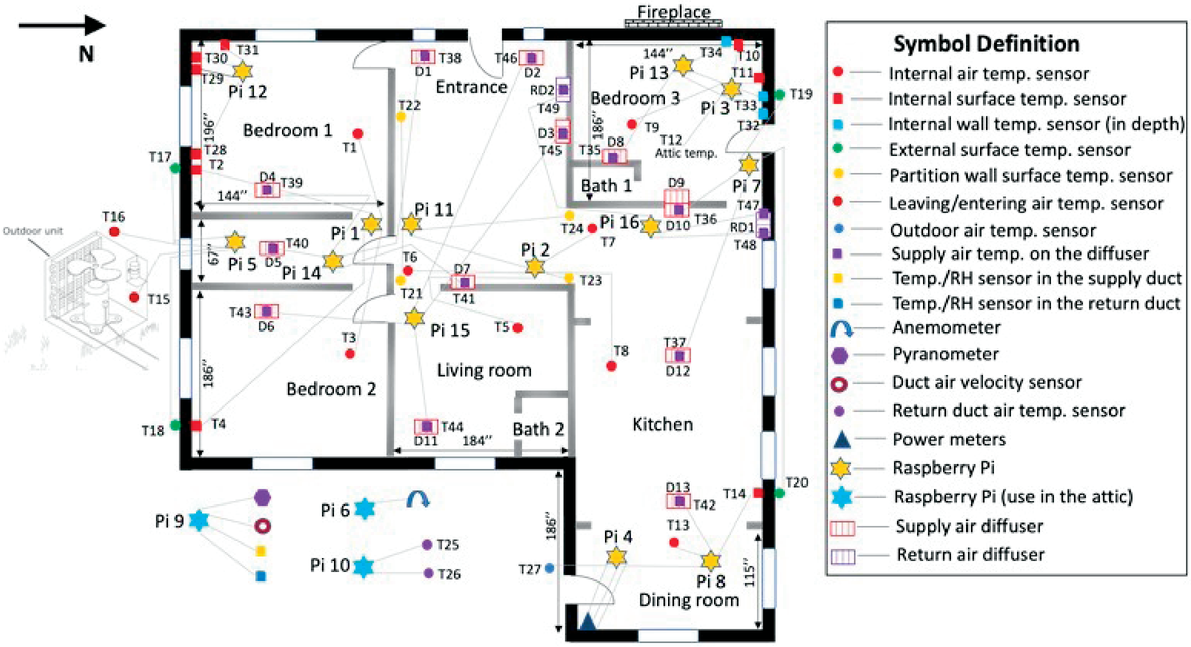
EXECUTIVE SUMMARY

> As stated in the Building America Research to Market Plan (Werling 2015), homes present 27% to 42% energy savings opportunities by using advanced monitoring of residential loads and fault detection and diagnosis (FDD). Moreover, heating savings of 5% to 15% were observed from simply setting back home thermostat set points by 10°F to 15°F for 8 hours per day in the fall and winter (DOE 2021). There is substantial energy- and cost-saving potential in advancing home air-conditioning (AC) systems for automated fault detection and optimal operation.

Meanwhile, although the smart thermostat industry has increasingly provided data specific to advanced AC controls and energy management in homes, there is a lack of a systematic framework that can connect data on comfort, occupancy, weather, energy use, and time-of-use (TOU) electricity pricing to generate meaningful information for advanced home comfort system diagnosis and optimal control (Rotondo et al. 2016).

In this project, we developed and tested a learning-based home thermal model that facilitates the operation of a model predictive control (MPC)-based optimization agent and an automated fault detection and diagnosis (AFDD) agent. The home thermal model was constructed using a two-node resistor-capacitor model. Moreover, two accompanying parameter identification methods were introduced, least-squares and optimization. Based on the home thermal model, the MPC-based optimization agent was developed to optimize residential HVAC operation. Using two FDD methods, the AFDD agent was constructed to detect and diagnose two prevalent residential AC faults, airflow reduction and refrigerant undercharge. The home thermal model, along with the MPC-based optimization agent and AFDD agent, were tested at the Norman Test House, Miami Test House, Pacific Northwest National Laboratory (PNNL) Test House A, and PNNL Test House B. Finally, they were also field tested in nine demonstration homes with real occupants.

Photo from Getty 1124228019



Oklahoma house floor plan with locations of the sensors. *Figure by the authors*

In conclusion, the home thermal model and its two parameter identification methods were successfully verified. Results from the Norman Test House and Miami Test House show that the model can accurately predict 12-hours-ahead space air temperature, achieving 1.96°F error at 95% confidence, which surpasses the success criterion of 2°F error at 90% confidence defined in the statement of project objectives. However, an error of 2.6°F at 90% confidence was observed at PNNL Test House A, and an invalid model was obtained at PNNL Test House B. A closer data examination revealed that erroneous sensor measurements in the PNNL datasets were the root cause of these issues. Therefore, sensor/data quality control is necessary before performing model identification.

In addition, both the MPC-based optimization agent and AFDD agent were successfully developed using data collected from the four test houses and implemented at the nine demonstration homes. An online, cloud-based data management platform was also created to facilitate data collection from smart thermostats and smart meters and enable remote control of AC units. The MPC-based optimization agent was found to possess the capability of making optimal AC operation decisions that leverage information from various sources such as the users' preferred comfort temperature, weather and weather forecasts, home thermal condition, and utility TOU rate. Field tests at the demonstration homes have shown that with MPC, up to 51% and 62% cost savings can be obtained on hot and mild summer days, respectively.

For the AFDD agent, two methods were developed and tested to detect and diagnose two prevalent faults in residential AC units: incorrect refrigerant charge level (RCL) and airflow reduction (AFR). Method 1 compares actual enthalpy changes across an evaporator with baseline enthalpy changes and uses their differences to signal a possible fault. While the baseline enthalpy changes can be obtained from manufacturer data for installation fault detection (Phase 1) and from operational data for degradation fault detection (Phase 2), the actual enthalpy changes are obtained from data collected by a smart thermostat and a node sensor placed in a diffuser, along with engineering assumptions. The method was validated to be effective in detecting and diagnosing RCL and AFR faults when the fault severity reached 30%. However, when both faults occur simultaneously, Method 2, which uses two fault indices—enthalpy changes and indoor fan power—is required. The method first uses indoor fan power measurements to diagnose AFR and then uses enthalpy changes to diagnose overcharge, undercharge, and occurrent faults with AFR. Through laboratory and field tests, the method was shown to be capable of catching less severe (around 15%) AFR faults. Overall, the AFDD agent was found to be effective.

Table of Contents

1 Introduction 1
2 Laboratory Test Home Instrumentation 5
  2.1 Lab House in Norman, Oklahoma 5
  2.2 Test House in Miami, Florida 8
  2.3 Test Houses in Richland, Washington 11
3 Home Thermal Model and Parameter Identification Development 15
  3.1 Home Thermal Model 15
  3.2 Parameter Identification Methods 19
4 Home Thermal Model Verification 24
  4.1 Model Verification Using Data From the Norman Test House 24
  4.2 Model Verification Using Data From the Miami Test House 47
  4.3 Model Verification Using Data From the PNNL Test Houses 53
5 Development and Verification of Optimal Operation Algorithm 60
  5.1 MPC-Based Optimization Methodology 60
  5.2 MPC-Based Optimization Agent Validation in Norman Test House 62
6 Development and Verification of FDD Algorithm 64
  6.1 Faults Considered 64
  6.2 FDD Methodology 64
  6.3 Algorithm Development 65
  6.4 Determination of Fault Threshold Through Uncertainty Propagation Analysis . 70
  6.5 Validation of the Proposed FDD Algorithm in the Norman Test House 72
  6.6 Validation of the Effectiveness of Fault Detection Using the Miami House 78
  6.7 Validation of the Effectiveness of Fault Detection Using PNNL House B 83
7 Field Test 87
  7.1 Introduction of Participating Test Houses 87
  7.2 Implemented MPC-Based Optimization Agent and Its Test Results 93
  7.3 Implemented AFDD Agent and Its Test Results 105
8 Technology Transfer 116
  8.1 Technology Dissemination 116

8.2 Technology Commercialization 117
9 Conclusions 119
References 122
Appendix 127
Appendix A. Lists of the Pi and Sensors Installed in the Norman Test Home 127
Appendix B. Model Parameter Estimation 130
Appendix C. Weather Conditions 132
Appendix D. News Coverage of the Project for Technology Dissemination 134

List of Figures

Figure 2.1. Oklahoma test home and data acquisition device location and information .. 7
Figure 2.2. Oklahoma house floor plan with locations of the sensors 8
Figure 2.3. Florida test home and data acquisition device location and information 10
Figure 2.4. Miami house floor plan 11
Figure 2.5. Two views of PNNL's identical lab homes viewed from (a) the northeast and (b) the northwest. The second sits in the background. 12
Figure 2.6. Lab homes floor plan with locations of heating/cooling registers, pyranometer, and thermocouples measuring air temperatures 12
Figure 2.7. Washington lab homes, heating/cooling equipment, and data acquisition systems 13
Figure 3.1. One virtual envelope 16
Figure 3.2. Home heat loss rate vs. outdoor air temperature for three different wind speeds 17
Figure 3.3. Schematic diagram of the 2R2C network 18
Figure 3.4. A schematic diagram describing the parameter estimation 22
Figure 4.1. Validation using data from June 18–July 1, 2020 (total of 14 days when AC is on) 26
Figure 4.2. Validation using data from July 2–9, 17–20, and 24–26, 2020 (total of 15 days when AC is on) 27
Figure 4.3. Simulation results on August 1–10 and 12–16, 2020 (total of 15 days) 28
Figure 4.4. Data from August 1–10 and 12–16, 2020 (total of 15 days) for training and validation 29
Figure 4.5. Simulation result using data from May 22–June 3, 2020 (total of 13 days). The red markers indicate the mean error and the error at 95% deviation. 31
Figure 4.6. Validation results on May 28–June 11, 2020 (total of 15 days) 33
Figure 4.7. Validation results on June 18–July 1, 2020 (total of 14 days) 34
Figure 4.8. Validation results on July 2–9, 17–20, and 24–26, 2020 (total of 15 days) .. 35
Figure 4.9. Validation results on August 1–16, 2020 (total of 15 days) 36
Figure 4.10. Simulation results using data from August 1–16, 2020 (total of 15 days). The red markers indicate the mean error and the error at 95% deviation. 38
Figure 4.11. Validation results on May 22–June 3, 2020 (total of 13 days) 40
Figure 4.12. Validation using data from May 28–June 11, 2020 (total of 15 days) 41
Figure 4.13. Validation using data from June 18–July 1, 2020 (total of 14 days) 42
Figure 4.14. Validation using data from July 2–9, 17–20, and 24–26, 2020 (total of 15 days) 43
Figure 4.15. Comparison of measured and predicted indoor air temperatures 45
Figure 4.16. Comparison of the distributions of the absolute errors. The red markers indicate the mean error and the error at a 95% confidence interval. 46

Figure 4.17. Simulation results using data from November 25–30, 2020 (total of 6 days) 49
Figure 4.18. Validation results on December 1–8, 2020 (total of 8 days) 50
Figure 4.19. Simulation results using data from December 1–8, 2020 (total of 8 days) 52
Figure 4.20. Validation results using data from November 25–30, 2020 (total of 6 days) 53
Figure 4.21. Simulation results using data from July 3–8, 2020 (total of 6 days) 55
Figure 4.22. Validation results using data from July 9–14, 2020 (total of 6 days) 56
Figure 4.23. Indoor air temperature issue at PNNL Home A; data are from July 3–4.... 57
Figure 4.24. Simulation results using data from July 3–8, 2020 (total of 6 days)........ 58
Figure 4.25. Temperature measurement issue at PNNL Home B; data are from July 3–July 4, 2020. 59
Figure 5.1. Schematic of MPC-based optimization agent........ 60
Figure 5.2. Initial MPC agent test results on October 11, 2021 62
Figure 5.3. Additional MPC agent test results on November 11, 2021 63
Figure 6.1. Level of studies on typical faults in residential HVAC systems........ 64
Figure 6.2. Schematic representation of proposed AFDD 65
Figure 6.3. AFDD algorithm for IoT-based method 70
Figure 6.4. Norman Test House and data acquisition devices: (a) outdoor unit, (b) smart thermostat, (c) indoor unit, (d) RA hobo logger, and (e) SA hobo logger. 73
Figure 6.5. Comparison of measured and estimated SA dew points........ 75
Figure 6.6. Validation of model trained with rated data 75
Figure 6.7. Detection of low indoor airflow fault........ 77
Figure 6.8. Energy consumption increase due to low indoor airflow fault........ 77
Figure 6.9. Detection of low indoor airflow using indoor fan power 78
Figure 6.10. Validation of model trained with rated data for the Miami house 79
Figure 6.11. Test for the Miami house based on Method I and using actual SADP measurements 80
Figure 6.12. Test for the Miami House based on Method 1 and using estimated SADP81
Figure 6.13. 12% indoor airflow reduction detected using Method 2 for the Miami house 82
Figure 6.14. 21% indoor airflow reduction detected with Method 2 for the Miami house 83
Figure 6.15. Comparison between estimated and measured enthalpy changes for PNNL House B (dry climate)........ 84
Figure 6.16. Test for airflow rate reduction fault in PNNL House B (dry climate) 85
Figure 6.17. Test for 30% undercharge fault in PNNL House B (dry climate) 86
Figure 6.18. Measured fan power and predicted fan power for normal operation (dry climate)........ 87
Figure 6.19. Fan power analysis for low flow rate fault in PNNL House B (dry climate) 87

Figure 7.1. Information about the eleven (11) test houses 88
Figure 7.2. Installed smart thermostat, node sensor (surface temperature), and smart meter in one test home... 91
Figure 7.3. Set point distribution in eight different test houses 92
Figure 7.4. MPC controller schematic 93
Figure 7.5. Data pipeline of MPC agent 96
Figure 7.6. Real-time MPC test results and weather conditions over 7 consecutive days 98
Figure 7.7. Total cost vs. temperature difference between outdoor and indoor when MPC was used and not used 100
Figure 7.8. Weather clustering results in Norman and Miami from July to October..... 102
Figure 7.9. Cost savings summary for each house on hot summer days for normal operation days and MPC days 103
Figure 7.10. Cost savings summary for each house on mild summer days for normal operation days and MPC days 103
Figure 7.11. Real-time monitoring of enthalpy change across the evaporator for four test homes for Phase 1 107
Figure 7.12. Detection of installation issues for two of the selected homes using Method 2 108
Figure 7.13. Real-time monitoring of indoor power usage for all four test homes for Phase 1 110
Figure 7.14. Alerts sent for all four test homes 111
Figure 7.15. Real-time monitoring of enthalpy change across the evaporator for four test homes for Phase 2 112
Figure 7.16. Detection of low charge fault for OU lab house 113
Figure 7.17. Alert sent for detected low charge fault in OU lab house 113
Figure 7.18. Real-time monitoring of indoor power usage for all four test homes for Phase 2 115
Figure 7.19. Indoor power measurements for Home J showing airflow rate degradation and subsequent filter replacement 115
Figure C.1. Weather data on August 9, 2020 132
Figure C.2. Weather data on August 19, 2020 132
Figure C.3. Weather data from August 1–7, 2020 133
Figure D.1. Newspaper report about the project in August 2019................................ 134
Figure D.2. Newspaper report about the project in December 2022 135

List of Tables

Table 4.1. Estimated Parameters 25
Table 4.2. Absolute Error Comparison Using Thermocouple Temperature for Different Months 29
Table 4.3. Training Results of Optimization Method Using May Data 30
Table 4.4. Absolute Error Comparison Using Thermocouple Temperature for Different Months for the Optimization Method 32
Table 4.5. Training Results of Optimization Method Using August Data 37
Table 4.6. Absolute Error Comparison Using Thermocouple Temperature for Different Months for the Optimization Method 39
Table 4.7. Estimated Parameters for the Two Identified Models 44
Table 4.8. Absolute Error Comparison for the Two Identified Models 47
Table 4.9. Training Results of Least-Squares Method Using November Data 47
Table 4.10. Absolute Error Comparison Using Thermocouple Temperature 50
Table 4.11. Training Results of Least-Squares Method Using December Data 51
Table 4.12. Absolute Error Comparison Using Thermocouple Temperature for the Least-Squares Method 53
Table 4.13. Training Results of Optimization Method Using PNNL Home A Data 54
Table 4.14. Absolute Error Comparison Using Thermocouple Temperature at PNNL Home A 56
Table 4.15. Training Results of Optimization Method Using PNNL Home B Data 58
Table 6.1. Uncertainties in Sensor Measurements 72
Table 6.2. Data From EPT of Test System 74
Table 6.3. Parameters of Cooling Capacity Model 74
Table 6.4. EPT Data for Miami Test House 78
Table 7.1. Test House Information and TOU Enrollment 89
Table 7.2. Cost Savings and Energy Consumption Summary for Houses on Hot Summer Days 104
Table 7.3. Cost Savings and Energy Consumption Summary for Houses on Mild Summer Days 105
Table 7.4. Cost Savings and Energy Consumption Summary for Houses on Cold Summer Days 105
Table 7.5. Specifications of Tested Homes 106
Table A.1. Specifications of All Measurements for the Data Acquisition System 127
Table A.2. Specifications of Location and Function of Sensors with Connected Pi for the Data Acquisition System 129

1 Introduction

As stated in the Building America Research to Market Plan (Werling 2015), homes present 27% to 42% energy savings opportunities by using advanced monitoring of residential loads and fault detection and diagnosis (FDD). Moreover, heating savings of 5% to 15% were observed from simply setting back home thermostat set points by 10°F to 15°F for 8 hours per day in the fall and winter (DOE 2021). There is substantial energy- and cost-saving potential in advancing home air-conditioning (AC) systems for automated fault detection and optimal operation. *The savings potential is in line with the mission of the Building Technologies Office's Residential Buildings Integration Program, i.e., reduce the energy used for space conditioning and water heating in single-family homes by 40% in 2025, compared to its 2010 level*. However, only a few automated controls, sensors, diagnostics, and fault correction systems exist for residential comfort systems (DOE 2016). The tools available for residential HVAC in the market include Sensi Predict, which requires 10 additional sensors to be installed alongside the HVAC (Emerson 2021), Comfort Alert diagnostics, which use data from Copeland scroll compressors (Emerson 2017), and ComfortGuard, which also requires 10 additional sensors to be installed (Cericola 2015). Although innovative and shown to be efficient, these products incur additional costs, such as for installation and additional sensors, and as with ComfortGuard, a monthly or annual monitoring fee.

With the advent of smart thermostats, which record and store data such as space temperature, relative humidity, and HVAC on/off times, residential AC FDD without installation of additional sensors has recently received attention. Although the smart thermostat industry has increasingly provided data specific to advanced HVAC controls and energy management in homes, *there is a lack of a systematic framework that can connect data on comfort, occupancy, weather, energy use, time-of-use (TOU) electricity pricing, design of the home and its systems, and code compliance benchmarks to generate meaningful information for advanced home comfort system diagnosis and optimal control.* Home thermal loads, which bridge weather impacts with a home's unique thermal properties and internal gains, provide ground truthing for energy use by residential HVAC systems. A self-learning home thermal model that calculates thermal loads and automatically learns the unique thermal properties of a home is key to model-based optimal control and performance degradation detection without the need for intervention by homeowners.

Our proposed home comfort system for total performance deficiency/fault detection and optimal control ("SYSTEM" hereafter) is aimed at constructing such a framework. The SYSTEM uses *a computationally efficient, self-learning, validated home thermal model*, along with *performance deficiency/fault detection and optimal control*. The overall goal of the project is to commercialize an affordable smart SYSTEM that fills the gap in the market for a product that provides a systematic framework for data connection. The connected data will generate meaningful information for advanced home comfort system diagnosis and optimal control, and consequently help achieve or exceed the Building Technologies Office savings target. The SYSTEM will revolutionize current smart thermostats by adopting a home thermal model and a set of performance deficiency/fault detection and optimization rules built upon relevant data. The project will also include necessary testing and validation of the proposed technology. Specifically, our project objectives are to:

- **Objective 1**: Validate the hypothesis that the home thermal model can be applied to accurately capture home thermal properties and predict space temperature dynamics.
- **Objective 2**: Validate the hypothesis that data collected from smart thermostats and smart meters can be applied to detect both the deficiencies in system design and construction, and the faults during residency.
- **Objective 3**: Validate the hypothesis that real-time optimization can be achieved to balance space temperature and energy costs based on occupants' preferences, home thermal properties, weather forecasts, occupancy schedules, and TOU energy pricing.
- **Objective 4**: Demonstrate the cost and performance benefits of the technology at homes with different ages, sizes, and household incomes.
- **Objective 5**: Disseminate the technology through a public domain and/or on a website so that potential users, developers, and vendors can download pseudocode, publications, and presentations, and identify a minimum of one vendor as the technology licensee to commercialize the technology at the end of this funded project.

The following research questions are designed to address the project objectives.

- **Question 1**: Can the thermal model effectively predict energy use of an HVAC system (extracted heat) with less than 15% mean absolute error at 90% confidence?
- **Question 2**: Can the thermal model effectively predict the space air temperature within 2°F error at 90% confidence?
- **Question 3**: Can the MPC-based optimization agent be executed in real time with given weather forecasts and parameters identified automatically through data training?
- **Question 4**: To what extent (severity of AC faults) can the SYSTEM detect faults using data collected through smart thermostats?

In this project, the University of Oklahoma project team first instrumented four laboratory test houses for laboratory-scale technology development and validation. To cover the impact from different climates, three locations were chosen: the Norman Test House, Miami Test House, and Pacific Northwest National Laboratory (PNNL) Test House. Meanwhile, the project team developed and tested a learning-based second-order home thermal model that facilitates the operation of a model predictive control (MPC)-based optimization agent and an automated fault detection and diagnosis (AFDD) agent and successfully answered the four research questions. The second-order home thermal model, along with its two parameter identification schemes—the least-squares method and optimization method—were successfully developed and tested. Results from the Norman Test House and Miami Test House show that the model can accurately predict 12-hours-ahead space air temperature, achieving 1.96°F error at 95% confidence, which surpasses the success criterion of 2°F error at 90% confidence defined in the statement of project objectives. However, an error of 2.6°F at 90% confidence was observed at PNNL Test House A, and an invalid model was obtained at PNNL Test House B. A closer data examination revealed that erroneous sensor measurements in the PNNL datasets were the root cause of these issues. Therefore, sensor/data quality control is necessary before performing model identification.

In addition, both the MPC-based optimization agent and AFDD agent were successfully developed using data collected from the four test houses. The MPC-based optimization agent has been found to possess the capability of making optimal AC operation decisions that leverage information from various sources such as users' preferred comfort temperature, weather and weather forecasts, home thermal condition, and utility TOU rate. For the AFDD agent, two methods have been developed and tested to detect and diagnose two prevalent faults in residential AC units: incorrect refrigerant charge level (RCL) and airflow reduction (AFR). Method 1 compares actual enthalpy changes across an evaporator with baseline enthalpy changes and uses their differences to signal a possible fault. While the baseline enthalpy changes can be obtained from manufacturer data for installation fault detection (Phase 1) and from operational data for degradation fault detection (Phase 2), the actual enthalpy changes are obtained from data collected by a smart thermostat and a node sensor placed in a diffuser, along with engineering assumptions. The method was validated to be effective in detecting and diagnosing RCL and AFR faults when the fault severity reached 30%. However, when both faults occur simultaneously, Method 2, which uses two fault indices—enthalpy changes and indoor fan power—is required. The method first uses indoor fan power measurements to diagnose AFR and then uses enthalpy changes to diagnose overcharge, undercharge, and occurrent faults with AFR. Through the test in the Norman Lab House, Method 2 was shown to be capable of catching less severe (around 15%) AFR faults.

An online, cloud-based data management platform was created to facilitate data collection from smart thermostats and smart meters and enable remote control of AC units to prepare for field testing in 10 demonstration houses. The field tests were conducted in the third year of the project. The results at the demonstration homes have shown that with MPC, up to 51% and 62% savings in cost—compared with normal space air temperature control without the demand respond operation—can be obtained on hot and mild summer days, respectively. The AFDD method was shown to be capable of catching installation faults—mismatching of indoor and outdoor units—in two houses, along with a dirty filter fault. The AFDD agent that was purely constructed using the data collected from smart thermostats was found to be effective for identifying a single occurrence of two prevalent faults—flow rate reduction and undercharge—with 30% severity; the AFDD agent with additional data collected by smart meters was found to be effective at detecting an AFR fault with 15% severity. Thus, residential AC performance efficiency can be improved and cost savings can be obtained for homeowners, with no additional hardware needed, by using the data readily available through smart thermostats and cloud-based computing and the validated technologies in this project, including the learning-based thermal model, the optimization agent, and AFDD agent.

This report is organized as follows. Section 2 introduces the four laboratory test facilities. In Section 3, we explain the physics-based home thermal model and two parameter identification methods. This modeling approach and the parameter identification methods are verified using the four lab test facilities in Section 4. In Section 5, we introduce a developed MPC-based optimization agent for optimal AC operation and partially validate its feasibility of implementation using a lab-installed smart thermostat. Similarly, we introduce development and verification of an FDD agent in Section 6. In Section 7, we include field test results from running the agents in participating demonstration homes. In Section 8, we provide a summary of technology transfer and commercialization activities. Lastly, Section 9 provides our conclusions.

In terms of the project objectives and research questions, Objective 1 and Questions 1 and 2 on effectiveness of the home thermal model are specifically addressed in Sections 4.1, 4.2, and 4.3. Objective 2 and Question 4 concerning viability of the fault detection methods are answered in Sections 6.5, 6.6, and 6.7. Objective 3 and Question 3 on performance of the MPC-based optimization agent are discussed in Sections 5.2 and 7.2. Objective 4 concerning costs and benefits of the technology is examined in Sections 7.1, 7.2, and 7.3. Finally, Objective 5 on dissemination and commercialization of the technology is discussed in Sections 8.1 and 8.2.

2 Laboratory Test Home Instrumentation

In the technology development phase, we used four lab houses: the Norman Test House, Miami Test House, PNNL Test House A, and PNNL Test House B. The Norman Test House was the main test bed, where the project team conducted experiments to collect a massive amount of operational data for the model development. The test houses at PNNL and in Miami, Florida, are intended only for model validation. In the field testing phase, we recruited 11 participating houses for technology verification. This section includes the introduction of the four laboratory test houses. The field testing houses will be introduced in Section 7.

2.1 Lab House in Norman, Oklahoma

The lab house located in Norman, Oklahoma, is shown in Figure 2.1 (a). It is a single-family, one-story home with a floor area of 1,658 ft^2, built in 1940. The home is equipped with 3.5 tons (42,000 Btu/h) of cooling capacity and 1,400 cfm of airflow rate and includes three bedrooms and one living room. The thermostat is in the living room. A data acquisition system was installed in the house, which measured the entering and leaving air temperature from the outdoor unit of the HVAC system, indoor and outdoor air temperature, indoor wall surface temperature, supply and return air temperatures from the air duct, air temperatures from the supply and return air diffusers, wind speed, normal direct irradiation, return airflow rate, power consumption for the indoor and outdoor unit, and the total power use for the entire house. These data were measured using T-type thermocouples, a velocity sensor, an anemometer, a pyranometer, and power meters, respectively, as shown in Figure 2.1 (b)–(j). The data were logged using the connected Raspberry Pi and its associated thermocouple hat, as shown in Figure 2.1 (k) and (l). All the thermocouples were calibrated according to the ASTM standard E220 (ASTM 2019). An outdoor weather station, shown in Figure 2.1 (m), was set up for outdoor temperature, wind, and total solar measurements at thirty-second intervals, which were compared with data downloaded from Mesonet (Brock et al. 1995; McPherson et al. 2007) at five-minute intervals.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

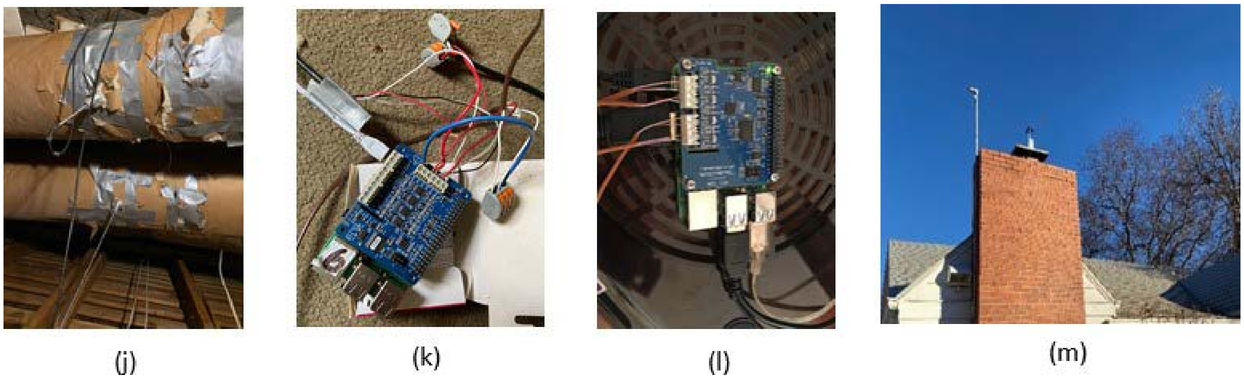


Figure 2.1. Oklahoma test home and data acquisition device location and information

(a) Outside view of the front of the test home; (b) Indoor unit with measurements of the relative humidity and temperature from the supply and return air duct; (c) Outdoor unit with measurements of the entering and leaving air temperatures; (d) Thermocouples for indoor air temperature measurements; (e) Thermocouples for interior wall surface temperature measurements; (f) Thermocouples for exterior wall surface temperature measurements; (g) Thermocouple for air temperature measurements from one supply diffuser; (h) Thermocouples for air temperature measurements from two return diffusers; (i) Power meters for the indoor unit and outdoor unit of the HVAC system and total power for the house; (j) Flow rate sensor; (k) and (l) Raspberry Pi and its associated thermocouple hats; and (m) Outdoor weather station.

All photos in report by the authors, unless noted otherwise.

To further illustrate the sensors and their measurements and locations in the house, a floor plan was drawn and is shown in Figure 2.2. Specifications of all measurements and locations of the sensors with their Pi for the data acquisition system are shown in Appendix A (Tables A.1 and A.2). Because the combination of the indoor air temperature T7 and interior wall surface temperature T11 showed most consistent results, T7 and T11 were used in the study. Moreover, weather data from the data acquisition system were compared with data downloaded from Mesonet (Brock et al. 1995; McPherson et al. 2007) at five-minute intervals. The comparison shows that Mesonet data provided similar results. Considering that Mesonet data quality is routinely maintained by the National Weather Center, Mesonet data were used in the study.

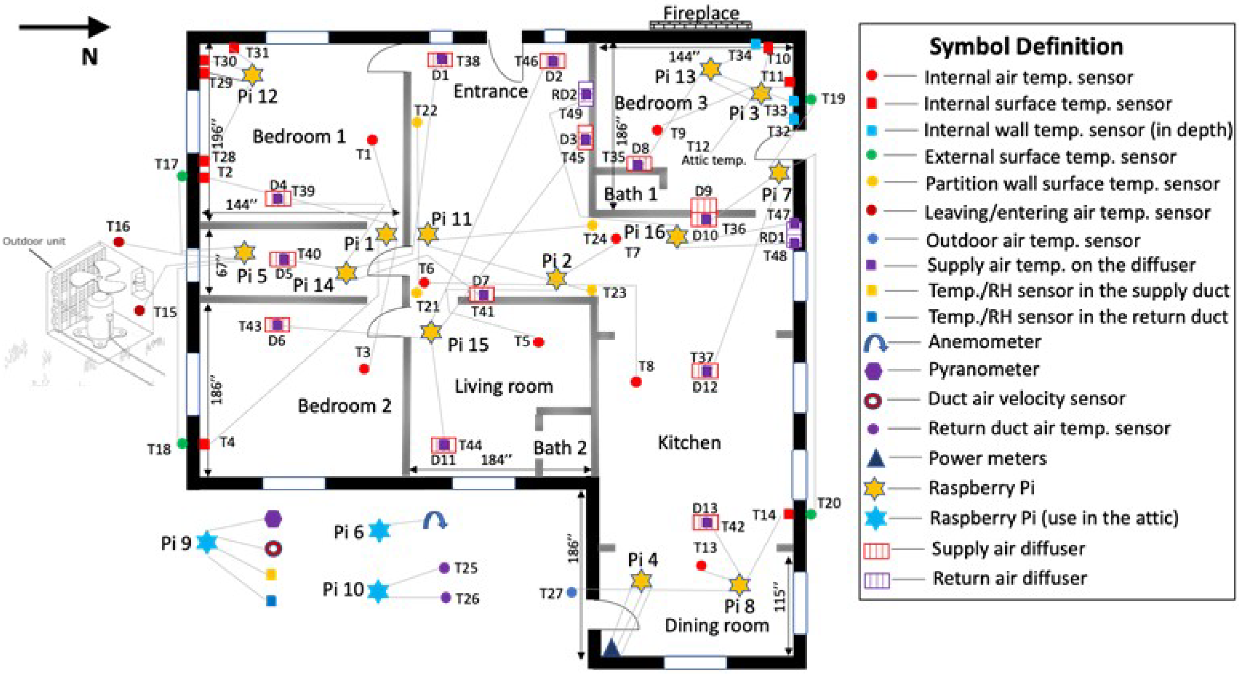
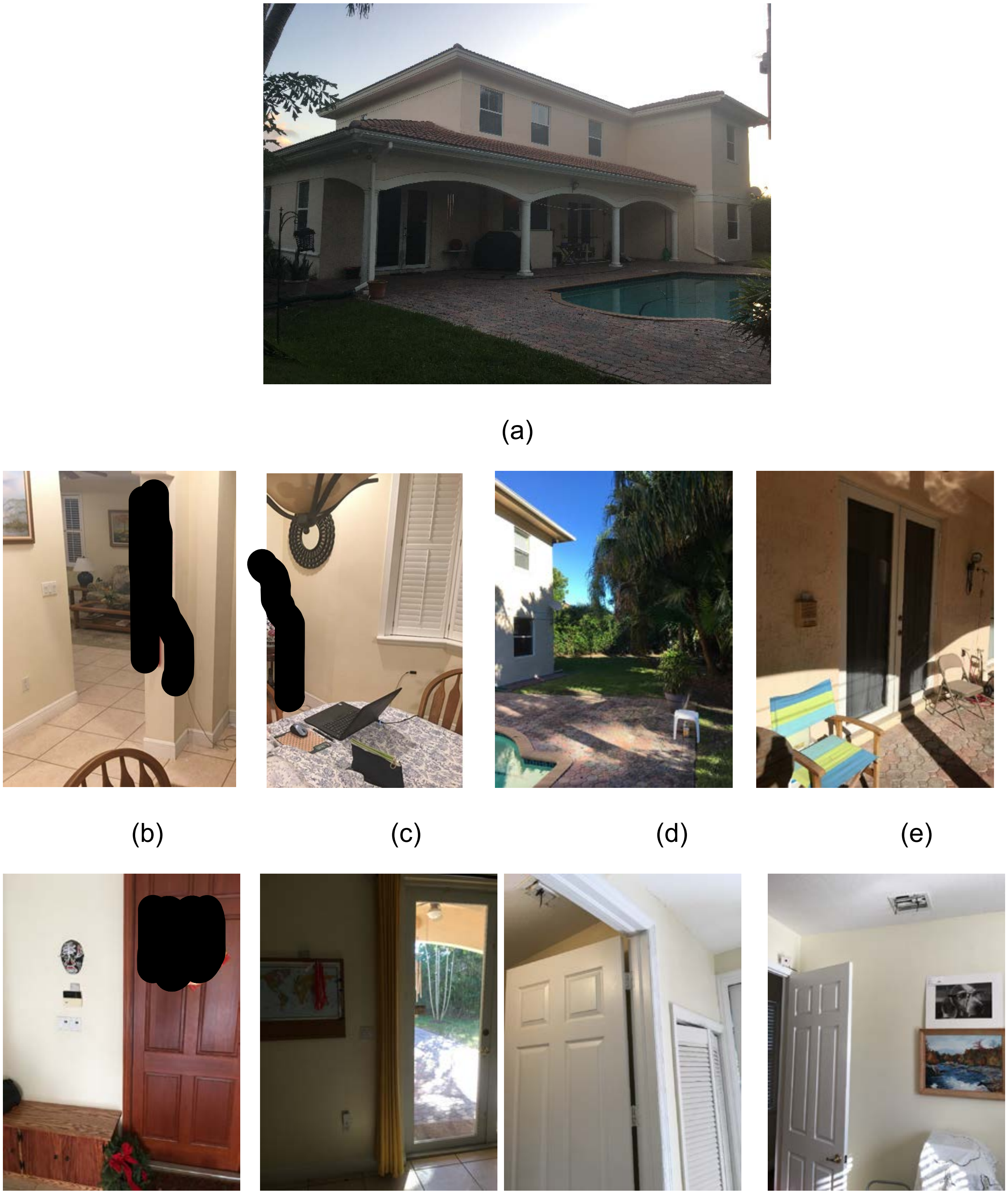


Figure 2.2. Oklahoma house floor plan with locations of the sensors

2.2 Test House in Miami, Florida

The test house in Miami, Florida, is shown in Figure 2.3(a). It is a single-family, two-story home with a floor area of 2,757 ft^2, built in 2002. The first floor is equipped with a 4-ton AC with an airflow rate of 1,600 CFM. The second floor is equipped with a 3-ton AC with an airflow rate of 1,113 CFM. The home includes three bedrooms on the second floor and one bedroom, one living room, one family room, a dining room, and a kitchen on the first floor. A thermostat is in the hallway on each floor. Hobo data loggers were installed in the house, which measured the temperature and humidity of the indoor air, outdoor air, supply air from one diffuser, and return air inside the AC cabinet, as well as the interior surface temperature of an interior wall and a north-side exterior wall in the dining room. A power meter was installed to measure the power of the indoor unit and outdoor unit of the two AC systems. The airflow rate was measured from all diffusers and applied to obtain the airflow rate of the two indoor units using a one-time measurement. An irradiation sensor was installed outside one south side window on the second floor to measure the solar irradiation. The locations of the sensors and power meter are shown in Figure 2.3(b)–(m). The measurement interval is set at one minute.



[redacted]

[redacted]

[redacted]

(a)

(b) (c) (d) (e)

(f) (g) (h) (i)

[redacted]

(j) (k) (l) (m)

(n) (o)

Figure 2.3. Florida test home and data acquisition device location and information

(a) Back view of the test home; (b) Interior wall with temperature measurements; (c) Exterior wall with temperature measurements; (d) outdoor air temperature measurements far away from the house; (e) outdoor air temperature measurements close to the pool; (f) air temperature measurement in the living room (entrance); (g) temperature measurement in the family room; (h) temperature measurement at one supply diffuser on the first floor; (i) temperature measurement at one supply diffuser on the second floor; (j) temperature measurement in the return cabinet on the first floor; (k) temperature measurement in the return cabinet on the second floor; (l) outdoor air temperature in front; and (m) power meters for the indoor unit and outdoor unit of the two AC systems; (n) temperature measurement in a bedroom on the second floor; (o) temperature measurement in the master bedroom on the second floor.

To further illustrate the sensors and their measurements and locations in the house, floor plans were drawn and are shown in Figure 2.4. Because the combination of the indoor air temperature from Front OA was found to show more consistent results, it was used in the study.

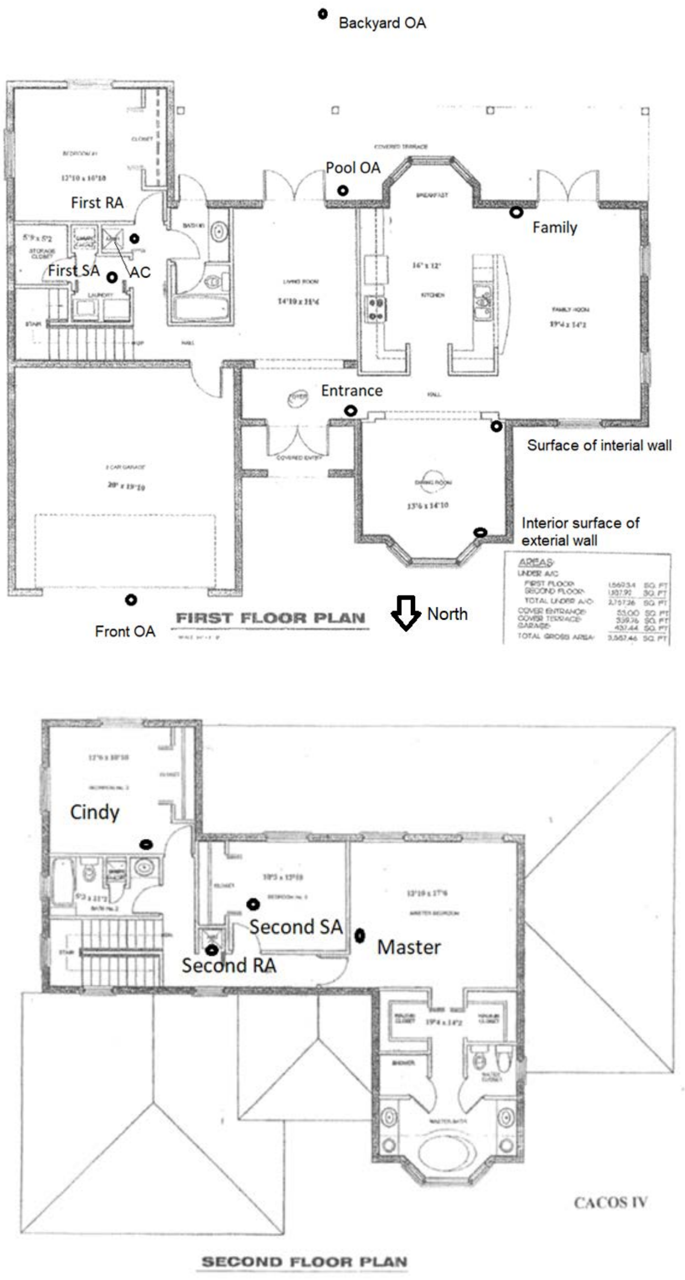


Figure 2.4. Miami house floor plan

2.3 Test Houses in Richland, Washington

The PNNL laboratory homes located in Richland, Washington, are shown in Figure 2.5(a) and Figure 2.5(b) with the floor plan in Figure 2.6. The homes are identically constructed manufactured single-family homes with floor areas of 1,493 ft^2 on a single floor over a crawl space, assembled in place circa 2011. Each home has three bedrooms, two baths, a dining room, and a living room (see floor plan in Figure 2). The floors are insulated to R-22 and finished with carpet and vinyl flooring. The walls are insulated to R-11 and the ceiling to R-22. Wood siding covers the exterior walls, and windows represent 195.7 ft^2 of the total exterior wall area.

(a) (b)

Figure 2.5. Two views of PNNL's identical lab homes viewed from (a) the northeast and (b) the northwest. The second sits in the background.

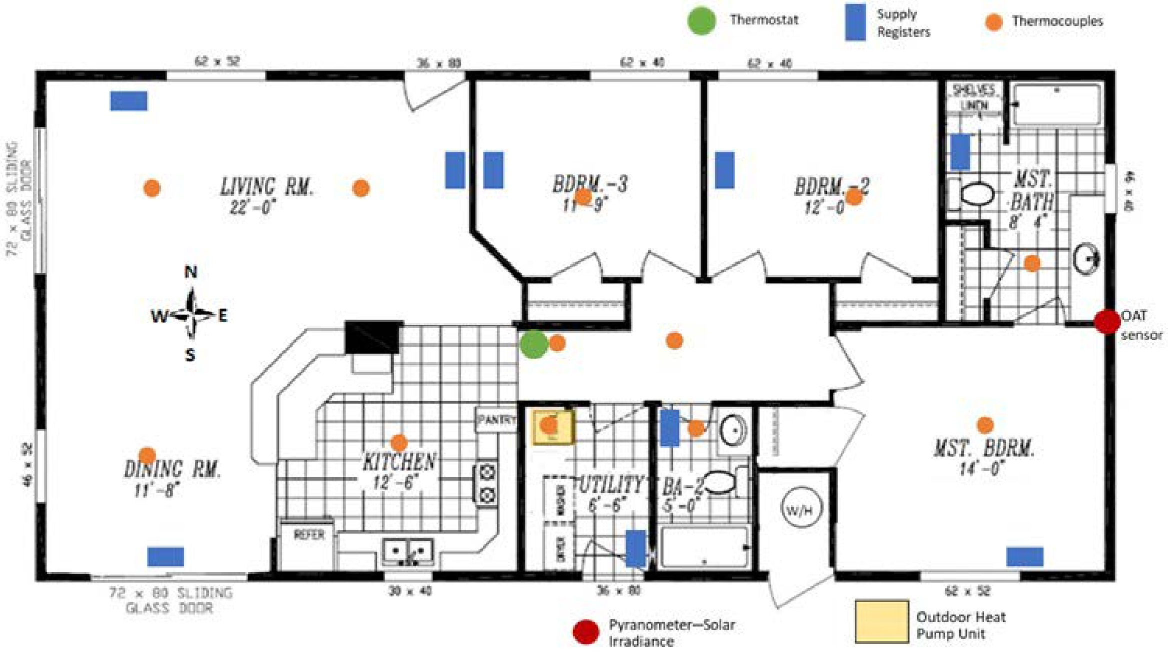


Figure 2.6. Lab homes floor plan with locations of heating/cooling registers, pyranometer, and thermocouples measuring air temperatures

(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 2.7. Washington lab homes, heating/cooling equipment, and data acquisition systems

(a) outdoor heat pump unit, (b) indoor heat pump unit showing the return air entrance on the top section, (c) indoor unit with front covers removed showing the coil in the top section and the blower and electric resistance heater unit in the lower section, (d) a close-up of the coil and filters using electrical and duct tape to block airflow to represent dirty or clogged filters, (e) return air temperature sensor located above the coil inside the upper section, (f) thermocouple and humidity sensor to measure room conditions, (g) rooftop weather station measuring wind speed and direction and shaded outdoor temperature sensor, (h) blue pyranometer (inside red circle) measuring solar irradiation and shaded outdoor temperature sensor, (j) flexible conduits through which wires run from current transducers in the electrical box behind the front electrical panel and the Campbell Scientific enclosure, (k) open Campbell enclosure showing wiring, logger, power supply, multiplexers, and Campbell Scientific CR1000 logger, and (l) another open enclosure showing sensor wiring, thermocouple amplifier and multiplexer, power supply, and Campbell CR1000X logger.

The homes have all electric equipment and appliances with heating and cooling provided by identical thermostatically controlled 13 seasonal energy efficiency ratio / 8.0 heating seasonal performance factor ducted central heat pumps with a rated cooling capacity of 29,400 Btu/hour (nominally 2.5 tons) and a nominal supply airflow rate of 850 scfm. Each heat pump has a rated heating capacity of 30,000 Btu/hour with 16 kW (approximately 54,590 Btu/hour) of electric resistance auxiliary heat. The thermostat is located in the hallway between the primary bedroom, bedrooms two and three, bathroom two, and the utility room. Heat is also available from Cadet fan-powered wall heaters throughout each home. The baths and kitchen have exhaust fans. The homes also include water heaters, refrigerators, washers and dryers, and dishwashers.

Instrumentation and a data acquisition system were installed in each house. Instrumentation includes energy metering with 42 individually monitored breakers, half of which are controllable or whole-house meters and a smart billing meter. Fifteen thermocouples measure indoor air temperatures, and two sensors measure indoor relative humidity. Twenty-two thermocouples measure the temperatures of the interior and exterior surfaces of window glazing. Two locatable mean radiant temperature sensors on tripods are available in the homes, and a “handheld” air velocity meter has been used for measuring flow rate and temperature of supply air flowing through the indoor air handler, which was in the project to determine the impact of blocked filters on flow rate and heat pump performance. Two Campbell Scientific data loggers are used for collecting and logging most of the data. Portable temperature sensors with onboard logging capability have also been used occasionally to supplement the main sensors (e.g., the temperatures of interior wall surfaces at various positions). Temperature and humidity data have been collected down to 1-minute time intervals, and power measurements as frequently as once every 10 seconds. Weather variables and solar irradiance are also measured on-site. Figures 2.7(a) through 2.7(x) provide photos of instrumentation and the heating and cooling equipment.

3 Home Thermal Model and Parameter Identification Development

3.1 Home Thermal Model

In this section, a home thermal model, which is built upon the energy conservation law, is formulated through the analysis of heat transfer processes between indoors and outdoors. The model represented by a second-order dynamic equation is used to capture the thermal dynamics of the indoor space and wall of a home.

3.1.1 Heat Transmissions Through Temperature Differences

For a 3R2C model application (Ogunsola, Song, and Wang 2014; Ogunsola and Song 2015), the exterior wall needs to be replaced by a wall for all exterior walls having different orientations, i.e., different orientations of the walls require them to be modeled individually. However, homes usually have one thermal zone (the entire house in most cases, and generally not more than two zones). Therefore, the home envelope (for a home with one zone) may be consolidated into one virtual envelope with the orientation-dependent wall temperature (T_{ie}) represented by the weighted average of the impacts on envelope elements having different orientations, as shown in Figure 3.1. The thermal properties of the virtual envelope are the weighted average of the thermal resistance and heat capacity, i.e., R_{ve} and $C_{ve,in}$ of all the envelope components. For internal space, the indoor air is represented by one uniform air temperature (T_{in}) in a thermal zone and its associated air thermal capacity (C_{air}) and thermal resistance (R_{air}). Therefore, heat transmissions through all the envelope components and internal space can be represented by two heat transfer relationships. The first relationship is driven by the temperature difference between the outdoor air temperature (T_o) and the wall temperature (T_{ie}). The second one is driven by the temperature difference between the interior wall surface temperature (T_{ie}) and indoor air temperature (T_{in}). Both relationships take into account the consolidated thermal properties of all the envelope components and internal spaces and contain parameters to be estimated using home operational data. The two relationships are shown in Equations (3.1) and (3.2), respectively.

$$C_{ve,in}\frac{dT_{ie}}{dt}=\frac{T_o-T_{ie}}{R_{ve}}+\frac{T_{in}-T_{ie}}{R_{air}} \tag{3.1}$$

$$C_{air}\frac{dT_{in}}{dt}=\frac{T_{ie}-T_{in}}{R_{air}} \tag{3.2}$$

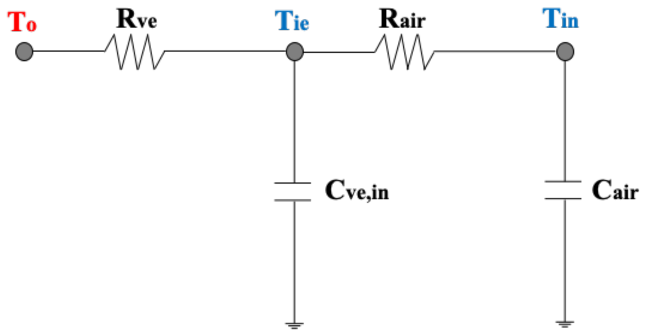


Figure 3.1. One virtual envelope

3.1.2 Solar Impacts

Solar radiation transfers heat inside a home through an envelope that consists of all the structural elements that separate the conditioned indoor spaces from unconditioned indoor spaces and the outdoors. The heat transfer takes place through two mechanisms. One is to heat the exterior opaque surfaces of the home. The heat received by the opaque envelope elements is first absorbed by the total heat capacity of the opaque envelopes and then released into the indoor air through conduction and convection. The other mechanism is to heat the indoor structural components and furnishings through which solar radiation is transmitted via glazing, such as windows and skylights. Some of the solar heat gain absorbed by the interior furnishings and structural components (e.g., walls) is immediately transferred to the indoor air by convection, and the rest is conducted into the structure or furnishings and gradually released later, thus heating the indoor air. In a traditional RC thermal model, the two mechanisms are described separately. The heat transfer of solar radiation on the exterior surfaces of opaque structural components is described by the sol-air temperature, which is orientation-dependent, while the solar gain through fenestration is separately described as radiative heat gains (McQuiston, Parker, and Spitler 2004). In this home model, however, a third-order polynomial, shown in Equation (3.3), is used to describe the overall attenuation from the normal direct irradiation that includes direct irradiation and diffuse irradiation to the solar heat received by all the envelope components, including the opaque and fenestration components, and eventually contributed to the internal space. The coefficients of the polynomial in Equation (3.3), representing the home thermal responses to solar inputs, are estimated using a parameter estimation scheme to be introduced in Section 3.2.

$$Q_{sol}(G) = a_1 G + a_2 G^2 + a_3 G^3 \tag{3.3}$$

where G is the global horizontal irradiation (W/m^2); $Q_{sol}(G)$ is the space air temperature increase that represents solar impacts on a home; and a_1, a_2, and a_3 are coefficients determined empirically from home operational data.

3.1.3 Wind Impacts

Wind impacts home thermal load through two mechanisms: changes in convection heat transfer coefficient and infiltration. To investigate the wind impacts, a preliminary study was carried out in March 2016 in an unoccupied home, and distinctive heat losses were observed for different wind speeds for the same outdoor air temperature. As shown in Figure 3.2, when the wind speed was at 1.34 m/s, the heat loss rate was approximately 10% higher than the loss rates at 0.67 m/s wind speed and more than 50% higher than the loss rates at close to 0 m/s. This showed that wind impacts cannot be ignored for home thermal load studies and suggested that the amount of heat gains or losses can be related to wind speeds. The heat loss rates were calculated using operational data of a gas heater logged over 1-minute intervals.

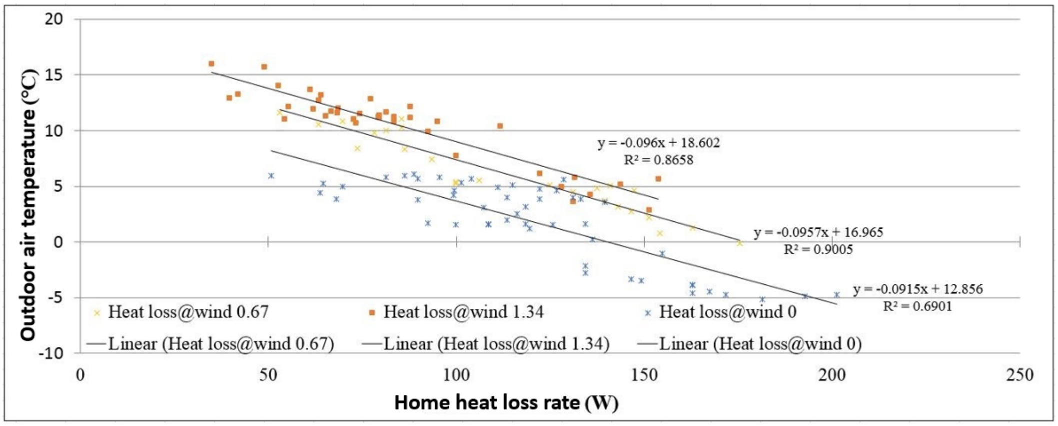


Figure 3.2. Home heat loss rate vs. outdoor air temperature for three different wind speeds

Due to the difficulties in directly calculating the infiltrated airflow rate (Gowri, Winiarski, and Jarnagin 2009; Waite and O'Brien 2010) and quantifying changes in the convection heat transfer coefficient, in this study the quadratic equation $b_1' \; W + b_2' \; W^2$ is used to capture the wind impacts for each specific house with the values of its parameters estimated through data training. Therefore, the rate of heat transfer by wind effects can be expressed by

$$q'_{vw} = \frac{T_o - T_{in}}{1/(b'_1 W + b'_2 W^2)} = \frac{T_o - T_{in}}{R_{vw}} \tag{3.4}$$

where q'_{vw} is the heat transfer rate due to wind effects; W is the wind speed; b'_1 and b'_2 are the empirically determined coefficients; and R_{vw} is a variable resistance dependent on the wind speed and airtightness of a specific home.

3.1.4 Internal Heat Gain Impacts

For homes, the dominant thermal mass (i.e., product of the mass and the specific heat capacity) comes from the envelopes (Kosny et al. 2001; Johra and Heiselberg 2017). This is because the heat capacity per specific volume of concrete, glass, wood/plastic, and materials for envelope elements is 1,000 times higher than the heat capacity per volume of air. It is also because the interior structure of a house is much smaller than that of a commercial building given the need for corridors, stairs, and elevators, in addition to interior zones, in a building. Meanwhile, for residential buildings, the volume of the envelope is significantly larger than the volume of partition walls and furnishings. On the other hand, internal heat gains have relatively small impacts compared with envelope heat gains/losses (Kim and Moon 2009). Herein, our hypothesis is to treat the internal heat gains as one input, Q_{int}, in this study. We have conducted two experiments to validate the hypothesis, one by introducing moderate heat using an electric heater (5,118 btu/hr) and the other one by introducing intensive heat using two electric heaters (total 10,236 btu/hr). Through the experiments, we found out that the errors caused by treating the internal gain as a constant throughout a day is within the model uncertainty (a success criterion).

3.1.5 Formulate Heat Transfer Processes

By integrating the contributions of Sections 3.1.1–3.1.4, the governing equation for the home thermal model can be expressed as

$$C_{ve,in}\frac{dT_{ie}}{dt} = \frac{T_o - T_{ie}}{R_{ve}} + \frac{T_{in} - T_{ie}}{R_{air}} \tag{3.5}$$

$$C_{air}\frac{dT_{in}}{dt} = \frac{T_{ie} - T_{in}}{R_{air}} + q'_{vw} + Q_{sol} + Q_{int} + Q_{sys} \tag{3.6}$$

where Q_{int} represents the sum of all the internal heat gains and Q_{sys} is the HVAC output. The circuit diagram for Equations (3.5) and (3.6) are shown in Figure 3.3.

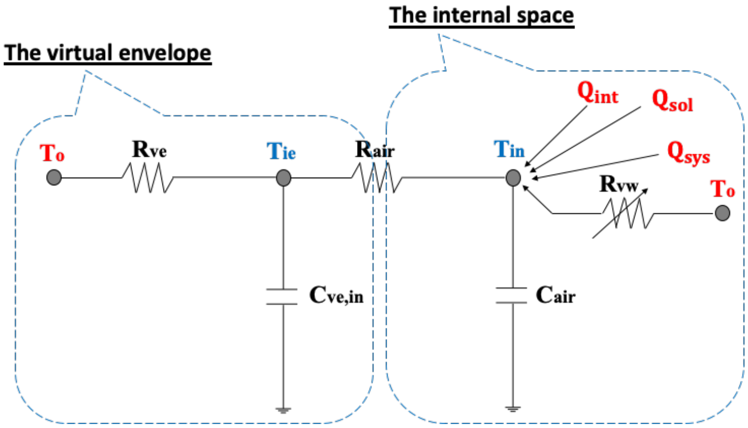


Figure 3.3. Schematic diagram of the 2R2C network

3.1.6 Summary

Substituting Equations (3.3) and (3.4) into Equation (3.6) and rearranging Equations (3.5) and (3.6), we obtain a continuous-time model

$$\frac{dT_{ie}(t)}{dt} = \frac{1}{\tau_1}[T_o(t) - T_{ie}(t)] + \frac{1}{\tau_2}\left[\left(T_{in}(t) - T_{ie}(t)\right)\right] \tag{3.7}$$

$$\begin{aligned}\frac{dT_{in}(t)}{dt} = &-\frac{1}{\tau_3}T_{in}(t)\\ &+\frac{1}{\tau_3}\Big[T_{ie}(t)+\big(T_o(t)-T_{in}(t)\big)\big(b_1W(t)+b_2W^2(t)\big)\\ &+\big(a_1G(t)+a_2G^2(t)+a_3G^3(t)\big)+\big(Q_iu_i(t)+Q_su_s(t)\big)\Big],\end{aligned} \tag{3.8}$$

where $\tau_1 = C_{ve,in}R_{ve}$ and $\tau_3 = C_{air}R_{air}$ are the time constants of the envelope and indoor air of a home, respectively; $\tau_2 = C_{ve,in}R_{air}$, $a_1 = a_1'R_{air}$, $a_2 = a_2'R_{air}$, $a_3 = a_3'R_{air}$, $b_1 = b_1'R_{air}$, and $b_2 = b_2'R_{air}$ are the corresponding coefficients associated with R_{air}; and $Q_i = Q_{int}R_{air}$ and $Q_s = Q_{sys}R_{air}$, where the internal heat gain Q_{int} and HVAC system output Q_{sys} are treated as constants associated with the internal activity schedules u_i and HVAC system on/off signal u_s, respectively.

All the coefficients in Equations (3.7) and (3.8), which represent the thermal responses of the envelope and indoor air of a home to the inputs of outdoor air temperature, interior wall surface temperature, wind, solar, internal heat gains, and HVAC system output, are estimated using a parameter estimation scheme introduced next in Section 3.2. Note: Equations (3.7) and (3.8) only consider sensible heat transfer for temperature prediction. The latent load impacts space humidity and therefore is not considered in the model.

3.2 Parameter Identification Methods

The formulated home thermal model in Equations (3.7) and (3.8) includes several unknown parameters that need to be estimated. In this section, a parameter estimation scheme is introduced using Euler's approximation and two different methods, namely, the least-squares method and the optimization method for different implementation scenarios. The two different methods are developed for the data collected from different seasons. In general, if the algorithm starts with the HVAC (including both heating and cooling) season, the optimization method needs to be applied for a better result. In contrast, in the transition season, the least-squares method needs to be adopted. We will explain in more detail in Section 4 the benefits of the two different methods. In this section, we focus only on introducing them.

3.2.1 Model Discretization

Because the home thermal model in Equations (3.7) and (3.8) is a continuous-time model, it needs to be discretized in order to use measured input and output data for parameter estimation. The continuous-time model is converted into a discrete-time model by applying Euler's method. With this method, the left-hand side of Equations (3.7) and (3.8) becomes

$$\frac{dT_{ie}(t)}{dt} = \frac{T_{ie}(t)-T_{ie}(t-1)}{\Delta t} \tag{3.9}$$

$$\frac{dT_{in}(t)}{dt} = \frac{T_{in}(t) - T_{in}(t-1)}{\Delta t} \tag{3.10}$$

where Δt is the sampling interval between two consecutive measurements.

By substituting Equations (3.9) and (3.10) into Equations (3.7) and (3.8), respectively, the continuous-time model becomes a discrete-time model of the form

$$T_{ie}(k) - T_{ie}(k-1) = \frac{\Delta t}{\tau_1}[T_o(k) - T_{ie}(k)] + \frac{\Delta t}{\tau_2}[(T_{in}(k) - T_{ie}k)] \tag{3.11}$$

$$\begin{aligned} T_{in}(k) - T_{in}(k-1) \\ &= -\frac{\Delta t}{\tau_3}T_{in}(k) \\ &+ \frac{\Delta t}{\tau_3}\big[T_{ie}(k) + \big(T_o(k) - T_{in}(k)\big)\big(b_1W(k) + b_2W^2(k)\big) \\ &+ \big(a_1G(k) + a_2G^2(k) + a_3G^3(k)\big) + \big(Q_iu_i(k) + Q_su_s(k)\big)\big] \end{aligned} \tag{3.12}$$

where k denotes discrete time. Although Equations (3.11) and (3.12) contain polynomial terms, the equations are *linear* in the unknown parameters. Therefore, in principle, the least-squares method may be used to optimally estimate these unknown parameters. It is, however, possible to improve the estimation performance by taking advantage of the fact that during certain periods, some of the terms in Equations (3.11) and (3.12) are naturally zero. For example, the solar term in Equation (3.12) should be zero at night, while the term containing the HVAC on/off signal should be zero whenever the AC is off. These observations suggest that instead of estimating all the unknown parameters *at once*, we could achieve a better estimation performance by applying the least-squares method in multiple steps, with each step devoted to estimating only a subset of the unknown parameters since some of the terms are naturally absent. This gives rise to what we called a stepwise parameter estimation scheme, to be described in Section 3.2.2. Herein, as an illustration, Equation (3.11) in the first step of the parameter estimation can be written in matrix form as

$$X\beta = Y \tag{3.13}$$

where *X* and *Y* are matrices containing the measured variables representing the inputs and output of the home thermal model, and *β* is the vector of unknown parameters to be estimated.

Assuming that X has full column rank, the least-squares solution to Equation (3.13) is:

$$\hat{\beta} = (X^TX)^{-1}X^TY \tag{3.14}$$

where $\hat{\beta}$ is the optimal estimate of the unknown parameters, and

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}; \qquad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}; \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}. \tag{3.15}$$

where n represents the number of time steps used for estimation and m represents the number of unknown parameters to be estimated. The detailed definitions of X, Y, and β can be found in Equations (B.1)–(B.8) in Appendix B.

3.2.2 Parameter Estimation Scheme Using the Least-Squares Method

The home thermal model in Equations (3.7) and (3.8) requires estimation of 10 parameters using a dataset of 7 known inputs: indoor air temperature (T_{in}), outdoor air temperature (T_o), interior wall surface temperature (T_{ie}), wind speed (W), normal direct irradiation (G), internal activity schedules (u_i), and HVAC system on/off signal (u_s). Of all the parameters to be estimated, the time constants (τ_1 and τ_3) in Equations (3.7) and (3.8) are most important in ensuring an accurate representation of the thermal properties of a home, which include the home envelope and internal space. The values of these two parameters do not change just because the HVAC system is turned on or off. Therefore, to minimize errors introduced by the indoor air and wall surface temperatures during HVAC on/off, we proposed a stepwise parameter estimation scheme based on the least-squares method described earlier in Equations (3.13)–(3.15). The parameter estimation process consists of two steps:

(1) Identify τ_1 and τ_2 by solving a least-squares problem formed by Equation (3.11) and measurements of the indoor air temperature $T_{in}(k)$, outdoor air temperature $T_0(k)$, and interior wall surface temperature $T_{ie}(k)$, focusing only on time periods when the HVAC system is off;

(2) Identify τ_3, b_1, b_2, a_1, a_2, a_3, Q_i, and Q_s by solving another least-squares problem formed by Equations (3.12) and all the measurements, thus completing the parameter estimation process.

Figure 3.4 shows a schematic diagram describing the parameter estimation process. Details of the parameter estimation are presented in Appendix B.

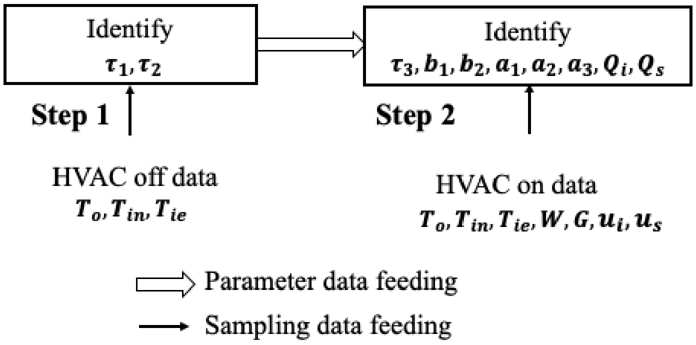


Figure 3.4. A schematic diagram describing the parameter estimation

3.2.3 Parameter Estimation Scheme Using the Optimization Method

An alternative way to estimate the unknown parameters of the model is what we called the optimization method. With this method, the input and output are the same as before. However, the parameter estimation process consists of the following two steps:

(1) Identify τ_1 and τ_2 by solving an optimization problem formed by Equation (3.1) and measurements of the indoor air temperature $T_{in}(k)$, outdoor air temperature $T_0(k)$, and interior wall surface temperature $T_{ie}(k)$, covering all time steps regardless of whether the HVAC system is on or off. The objective function of the optimization problem is the sum of the squares of the differences between the measured and modeled interior wall surface temperatures when the unknown parameters take on a specific set of values:

$$\underset{P}{minimize} J = \sum_{k=1}^{N} \left(T_{ie_{model}}^{k} - T_{ie}^{k}\right)^2 \text{ ,where } \mathrm{P} = \left[\tau_1, \tau_2, T_{ie}^{0}, T_{in_{measure}}\right] \tag{3.16}$$

This objective function can be minimized using the SLSQP (Kraft 1988) or Nelder-Mead's algorithm (Nelder and Mead 1965). The SLSQP algorithm is used in the simulation results presented in Section 4. The benefit of using SLSQP is that one can set up the constraints and bounds for the unknown parameters, in this case τ_1 and τ_2. Although the values of these two parameters may be different for different houses and different surrounding environments, it is reasonable to assume that they fall in certain ranges. Based on our experiments, we found that $\tau_1 \in$ (2000,8000) and $\tau_2 \in$ (20,5000) are realistic bounds. We also impose an additional constraint on their ratio since it is known that the values of τ_1 and τ_2 are typically proportionally related. Specifically, in the simulation result, the constraint is taken to be $5 \le \frac{\tau_1}{\tau_2} \le 20$. Finally, as required by the SLSQP algorithm, we let the initial guess of the values of τ_1, τ_2 be given by the result from Step 1 of the least-squares method described in the previous section.

(2) Identify τ_3, b_1, b_2, a_1, a_2, a_3, Q_i, and Q_s by solving a least-squares problem formed by Equation (3.14) and all the measurements, regardless of whether the HVAC system

is on or off, thus completing the parameter estimation process. A detailed description of these two steps can be found in Appendix B.

In general, to estimate the parameters of a second-order differential/difference equation, one may use the optimization method to train all parameters in the two equations simultaneously. However, due to the complexity of the home thermal model and the sensitivity of the resulting parameter estimates, we instead proposed the two-stage parameter estimation scheme.

4 Home Thermal Model Verification

The data collected from the four laboratory testing houses were used to verify the model and parameter identification methods. For model verification purposes, we collected data at 30-second time intervals using thermocouples we installed.

4.1 Model Verification Using Data From the Norman Test House

To run the algorithm, the first step is to identify the parameters of the thermal model using the first set of available data. However, we found that the parameter training method is sensitive to the data collected in different seasons, namely, a transition season and an HVAC season. Because the algorithm can start running at any season of the year, we have developed different methods—least-squares and optimization—to be applied to different seasons for best results.

In this section, we show the differences between the two methods when the parameter training algorithm starts with data collected in different seasons. A total of four tests are considered: the least-squares method with transition season data, the least-squares method with HVAC season data, the optimization method with transition season data, and the optimization method with HVAC season data. At the end of this section, we include a study on how the length of the data impacts the model training results using 6-day and 14-day data.

4.1.1 Tests Using the Least-Squares Method

4.1.1.1 Tests Using Model Trained by Transition Season Data

Table 4.1 shows the results of estimating the home thermal model parameters using the first 14 consecutive days of training data collected in May, which is considered a transition season. The trained model is then applied to the operational data collected in June, July, and August to predict 24-hours-ahead space air temperature. The predicted space air temperature is compared with measured space air temperature to verify the effectiveness of the model. Our success criterion for this verification, as indicated in the first section, is that the absolute errors between the measured and predicted space air temperatures are less than 2°F at a 90% confidence level.

Table 4.1. Estimated Parameters

Length of training data	τ_1	τ_2	τ_3	b_1	b_2
First 14 consecutive days	2383	262.6	7.897	0.005447	-0.0002832
Length of training data	a_1	a_2	a_3	Q_s	
First 14 consecutive days	5.422	-9.487	5.568	-2.086	

Figure 4.1 to Figure 4.3 present the verification results for the model trained using the 14-day transition data, and Table 4.2 presents the statistics. As can be seen, the absolute errors are less than 2°F at a 95% confidence level, exceeding the success criterion except for the result in August, which shows an absolute error of 2.47°F at a 95% confidence level, slightly above the 2°F error at a 90% confidence level. This finding has been reported in the progress report at the end of the first budge period. Therefore, the least-squares method has passed the performance verification when the transition season data are used to train the thermal model.

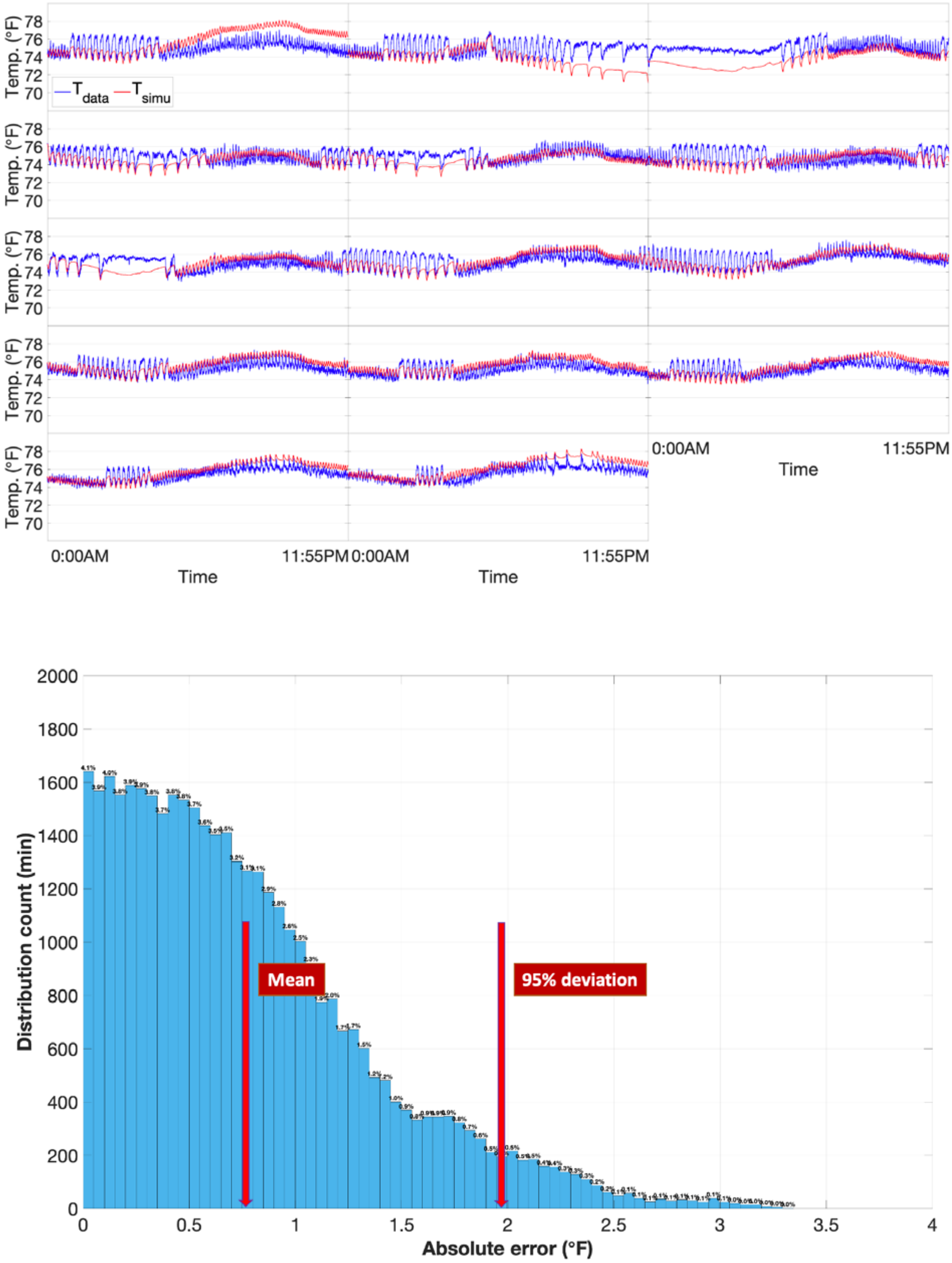


Figure 4.1. Validation using data from June 18–July 1, 2020 (total of 14 days when AC is on)

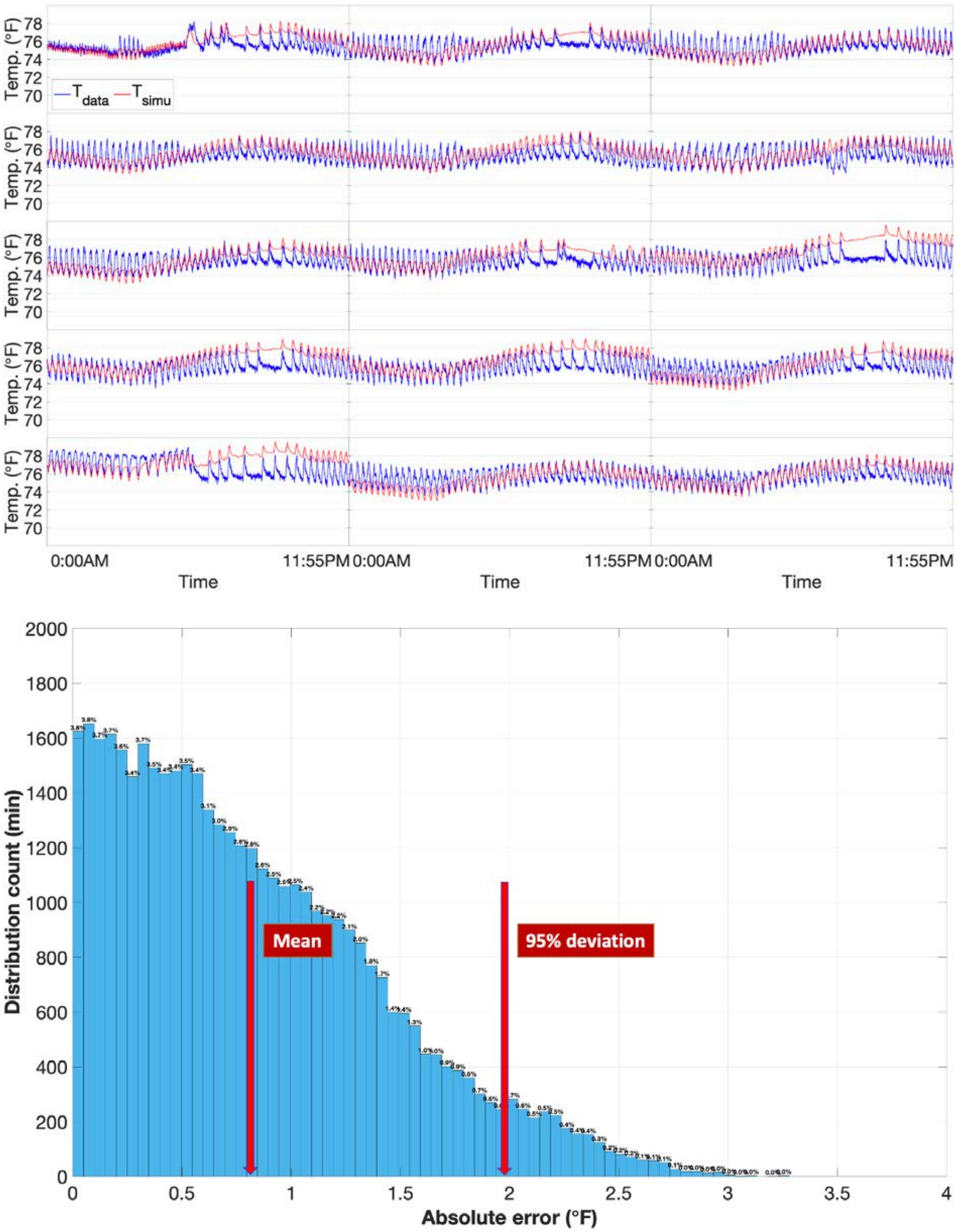


Figure 4.2. Validation using data from July 2–9, 17–20, and 24–26, 2020 (total of 15 days when AC is on)

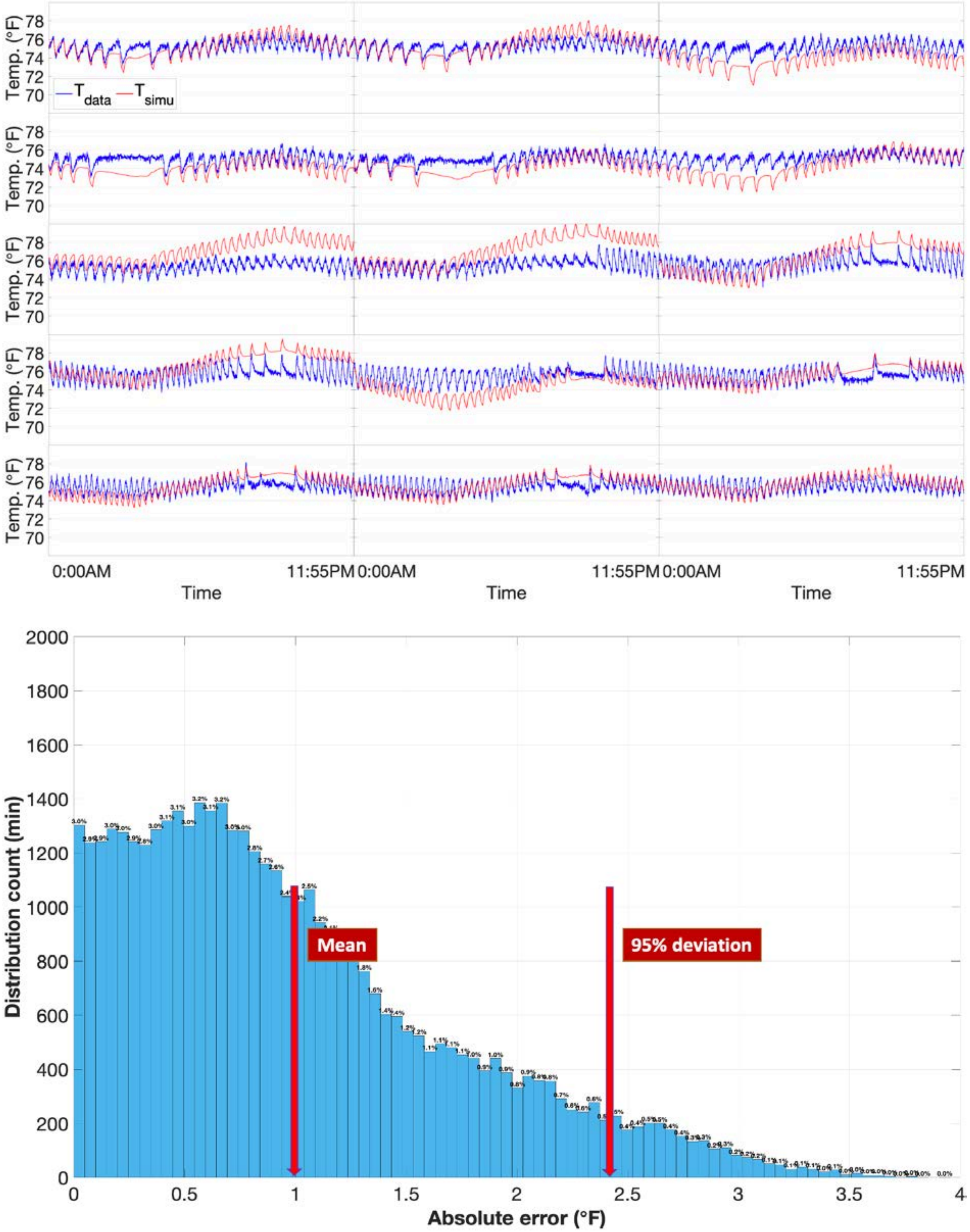


Figure 4.3. Simulation results on August 1–10 and 12–16, 2020 (total of 15 days)

Table 4.2. Absolute Error Comparison Using Thermocouple Temperature for Different Months

Month	Mean absolute error, °F	Maximum absolute error, °F	Absolute error at 95% confidence interval, °F
June	0.77	3.34	1.96
July	0.83	3.28	1.85
August	0.99	3.95	2.47

4.1.1.2 Tests Using Model Trained by HVAC Season Data

Figure 4.4 shows the predicted and measured space air temperature when the model is trained using data from August, which is considered a heavy HVAC season. The results are obviously not acceptable. The heavy HVAC season data predominantly contain information about AC impact and therefore are not effective in training building parameters in the thermal model that require heat transfer impacts through envelope. This is why the optimization method is developed to improve model performance for situations such as when the first set of data comes from the HVAC season.

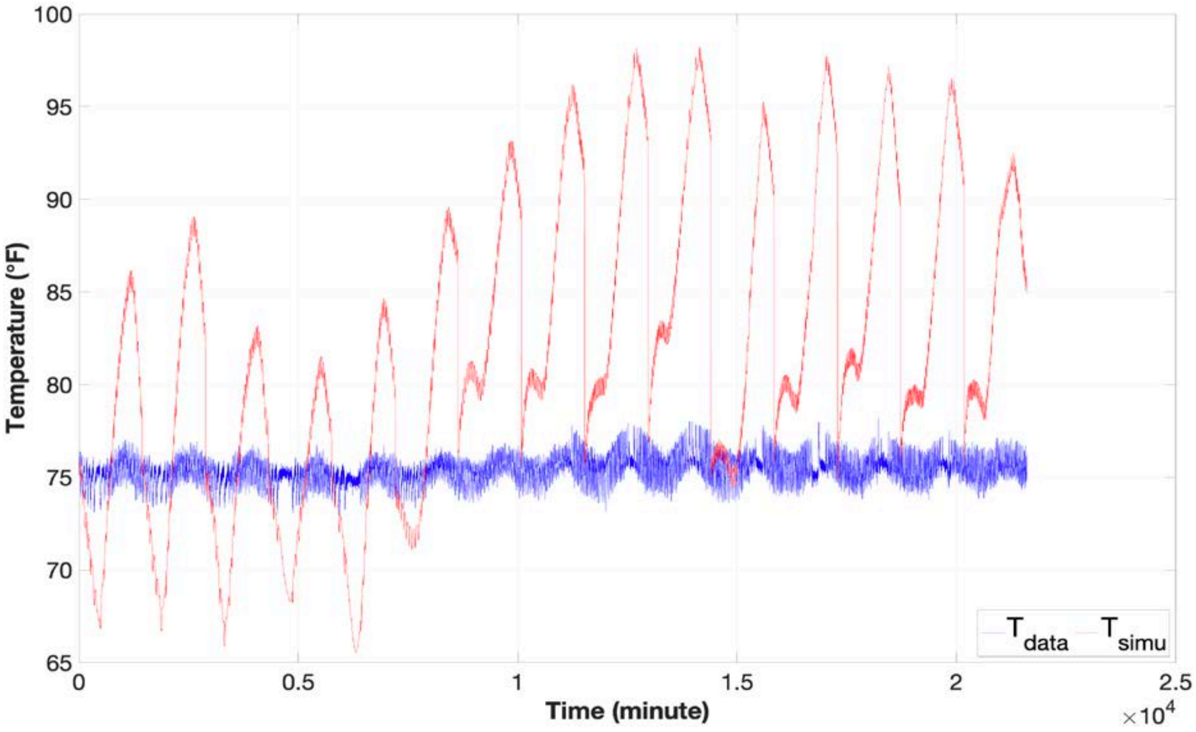


Figure 4.4. Data from August 1–10 and 12–16, 2020 (total of 15 days) for training and validation

4.1.2 Tests Using the Optimization Method

4.1.2.1 Tests Using Model Trained by Transition Season Data

As shown in Figure 4.4, the least-squares method may not achieve satisfactory temperature prediction if the training data are from the HVAC season, motivating us to propose the optimization method. In this section, we present the validation of this new training method. Following the same procedure as the least-squares method, we start from the transition season, i.e., using May data for training, and validate the model using data from other months. Then we change the training data to the HVAC season and, again, validate the performance using training data from other months.

Table 4.3 shows the results of estimating the home thermal model parameters using 13 consecutive days of training data.

Table 4.3. Training Results of Optimization Method Using May Data

Length of training data	τ_1	τ_2	τ_3	b_1	b_2
May 22–June 3 (total 13 days)	5078.520	397.118	7.897	0.005447	-0.0002832
Length of training data	a_1	a_2	a_3	Q_s	
May 22–June 3	5.422	-9.487	5.568	-2.086	

The top subplot of Figure 4.5 shows the measured and 24-hours-ahead predicted indoor air temperatures using the home thermal model with trained parameters from Table 4.3. As can be seen, the predicted temperatures match the measured temperatures quite well, with a mean absolute error of 0.65°F and a maximum absolute error of 2.54°F. Compared to the least-squares method, the optimization method performs better. The bottom subplot of Figure 4.5 displays a histogram of the absolute errors. A red marker is added to emphasize that 95% of the absolute errors are within 1.57°F. This suggests that the model is effective at capturing the home thermal dynamics by learning the properties of the test home.

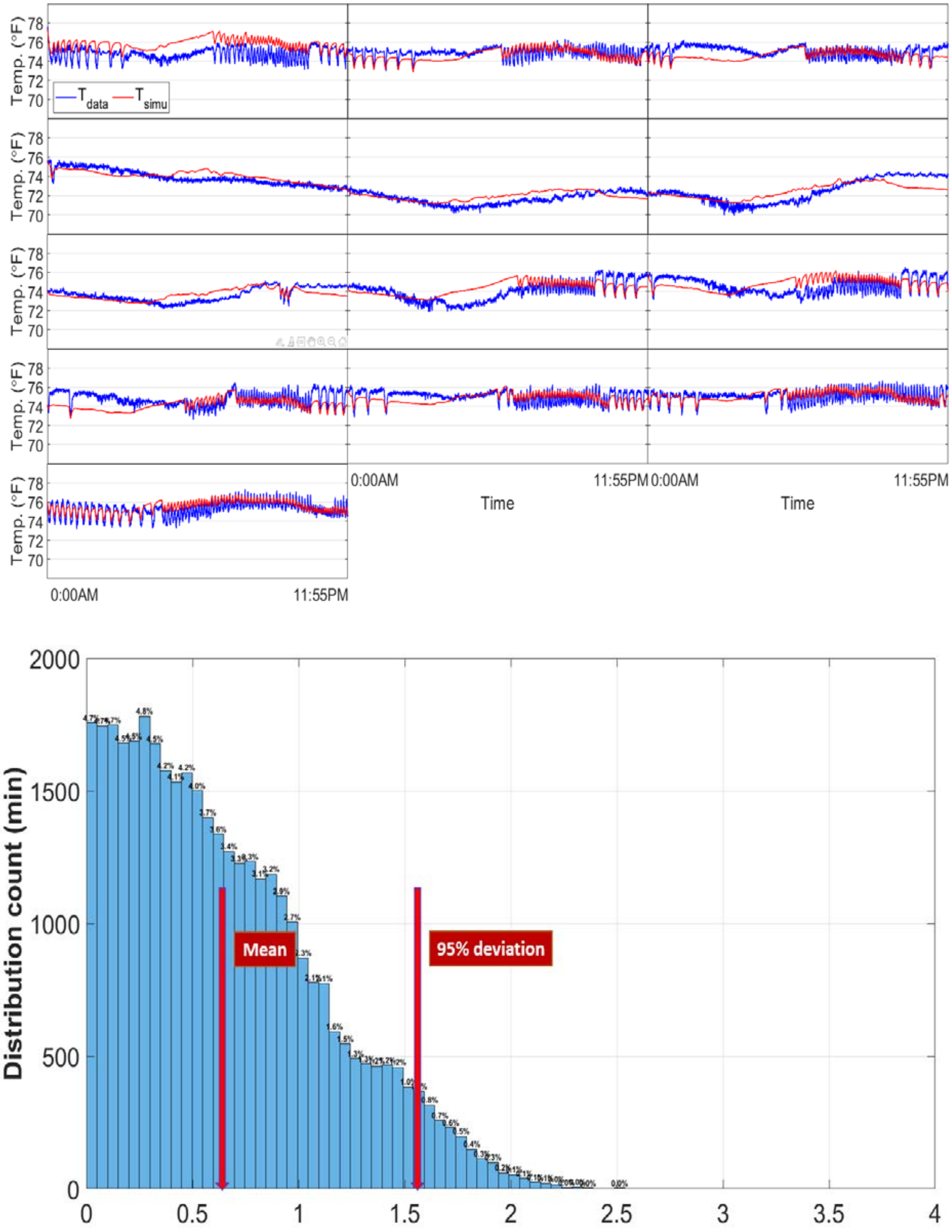


Figure 4.5. Simulation result using data from May 22–June 3, 2020 (total of 13 days). The red markers indicate the mean error and the error at 95% deviation.

Using the parameters trained in Table 4.3, we validate our model using data from June, July, and August. The simulation results and a histogram of the absolute errors are shown in Figure 4.6 to Figure 4.9. Table 4.4 states the mean absolute error, maximum absolute error, and absolute error at a 95% confidence interval. Compared with the least-squares method, the model trained by the optimization method shows a relatively better result for validation using data from those three months.

Table 4.4. Absolute Error Comparison Using Thermocouple Temperature for Different Months for the Optimization Method

Date	Mean absolute error, °F	Maximum absolute error, °F	Absolute error at 95% confidence interval, °F
May 28–June 11	0.62	3.09	1.52
June 18–July 1	0.61	2.42	1.55
July 2–9, 17–20, 24–26	0.58	2.77	1.41
August 1–August 16	0.70	2.72	1.61

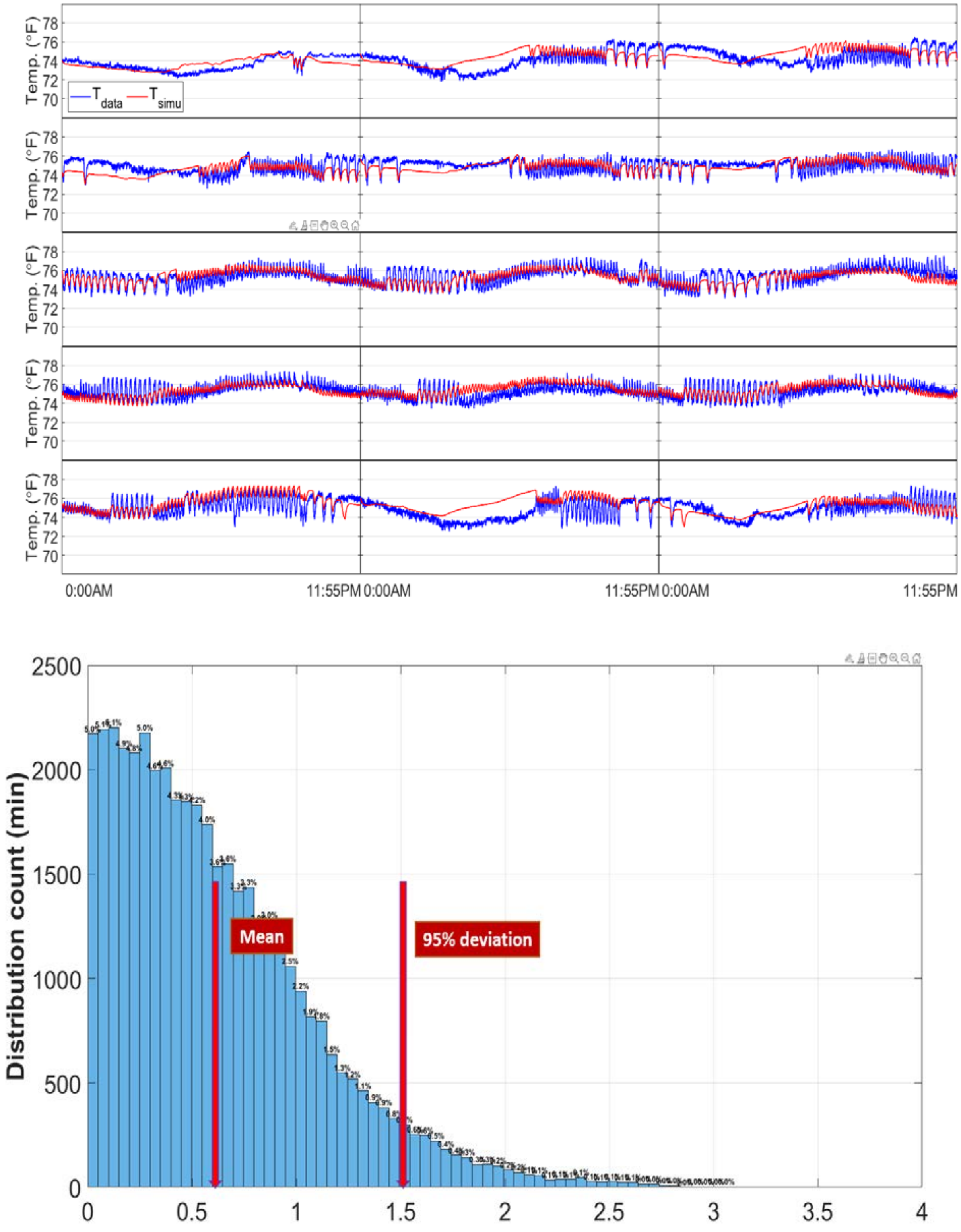


Figure 4.6. Validation results on May 28–June 11, 2020 (total of 15 days)

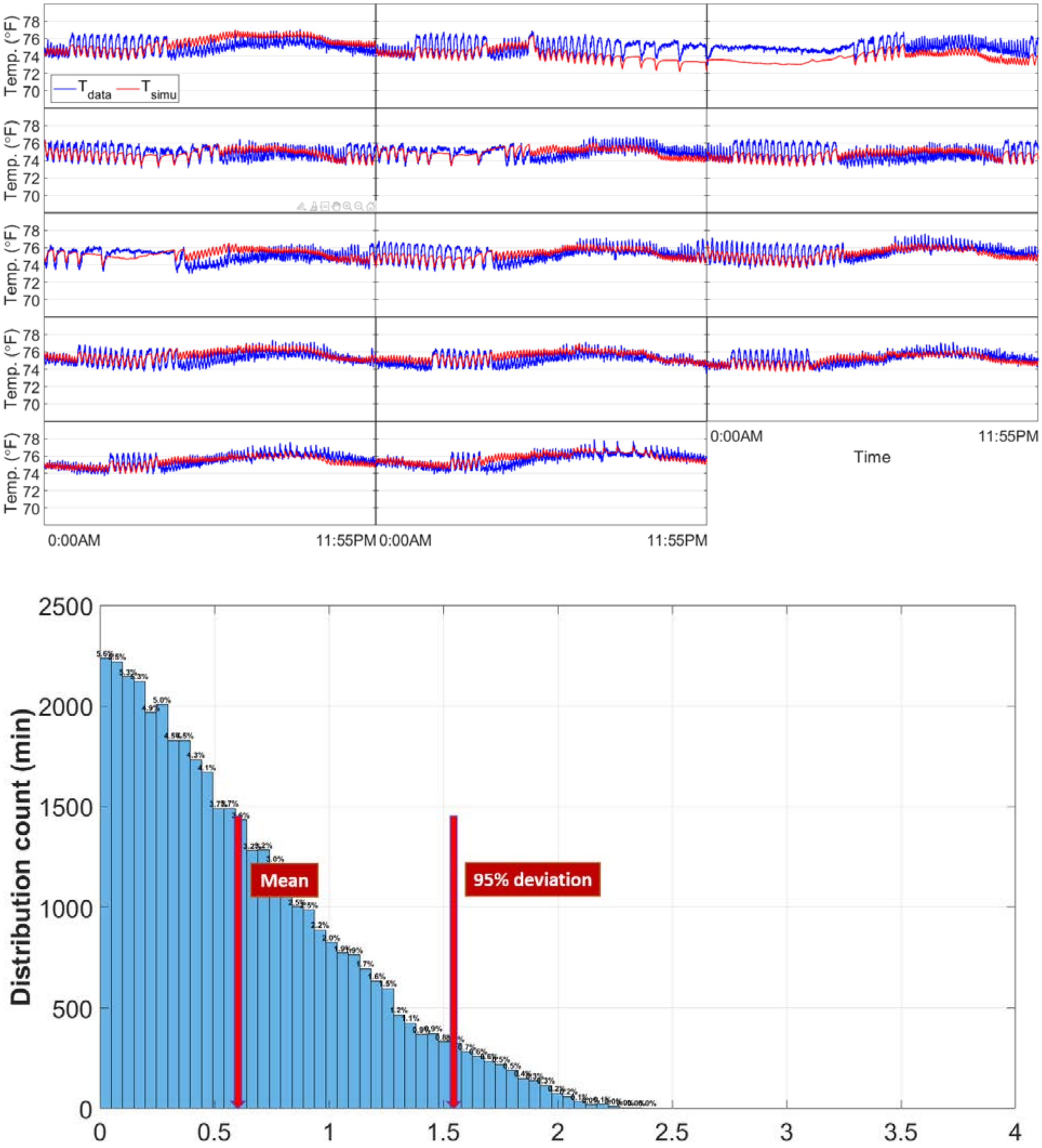


Figure 4.7. Validation results on June 18–July 1, 2020 (total of 14 days)

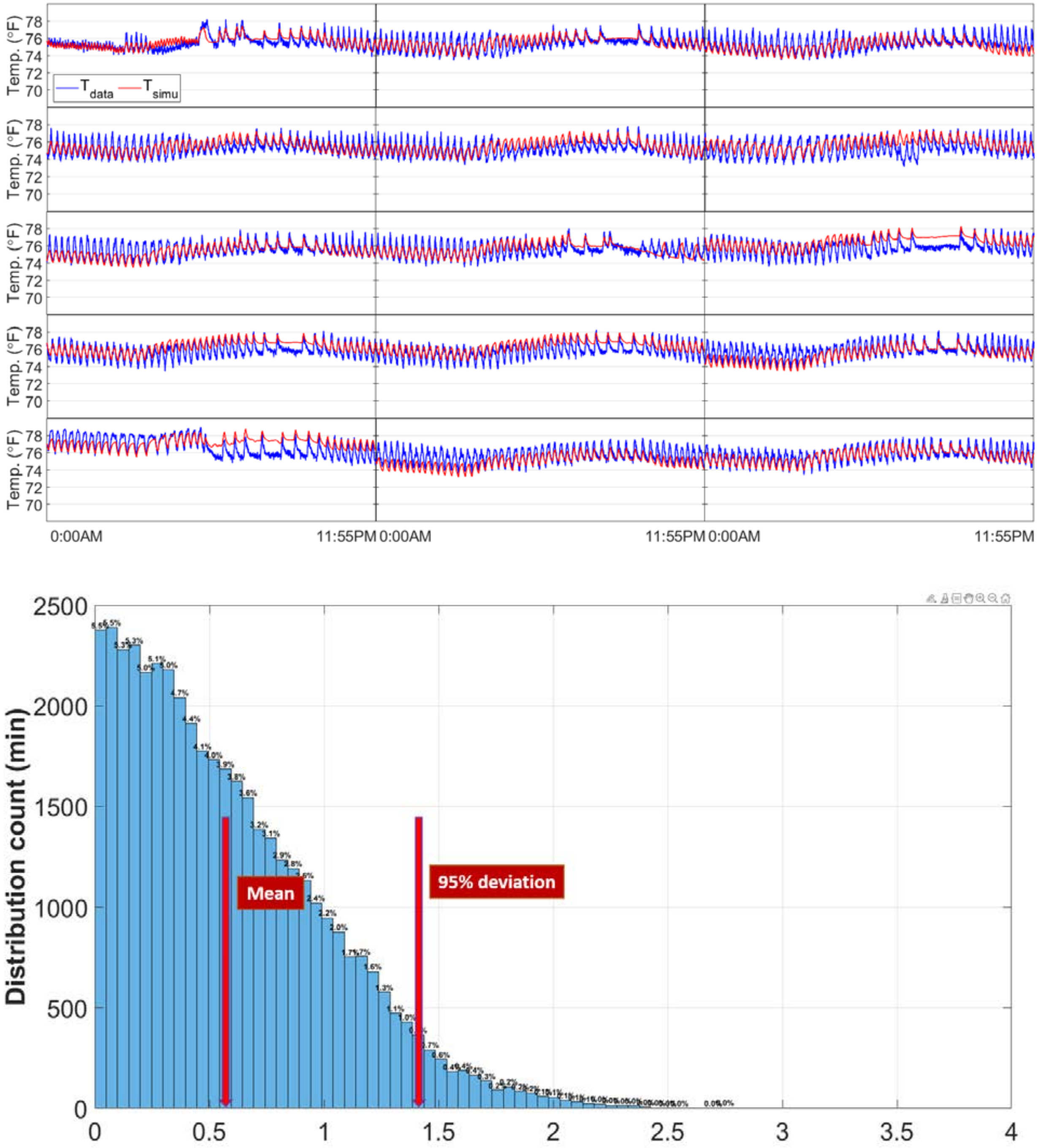


Figure 4.8. Validation results on July 2–9, 17–20, and 24–26, 2020 (total of 15 days)

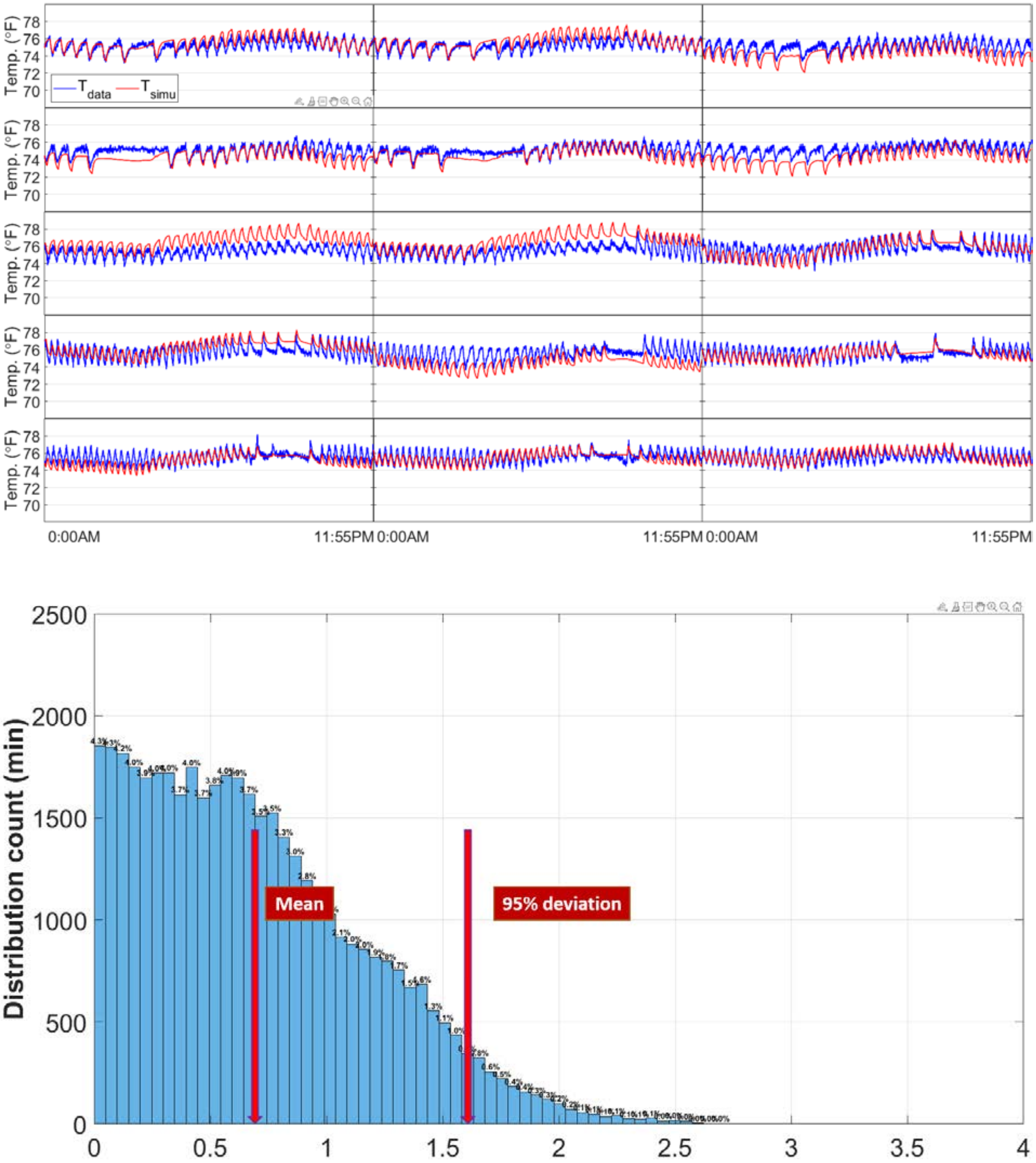


Figure 4.9. Validation results on August 1–16, 2020 (total of 15 days)

4.1.2.2 Tests Using Model Trained by HVAC Season Data

To better understand the efficacy of the optimization method, we also used it to train our model based on data from the HVAC season. Table 4.5 shows the results of estimating the home thermal model parameters using 15 consecutive days of training data.

Table 4.5. Training Results of Optimization Method Using August Data

Length of training data	τ_1	τ_2	τ_3	b_1	b_2
August 1–10, 12–16 (total 15 days)	7820.103	438.877	5.644	0.0148	-0.0009237
Length of training data	a_1	a_2	a_3	Q_s	
August 1–10, 12–16	-0.7219	2.579	-1.643	-1.102	

Figure 4.10 shows the measured and 24-hours-ahead predicted indoor air temperatures using the home thermal model with trained parameters from Table 4.5. Again, the predicted temperatures match the measured temperatures very well, with a mean absolute error of 0.83°F and a maximum absolute error of 3.2°F. Compared to the least-squares method above, the optimization method achieves a better performance. A histogram of the absolute errors is also shown in the figure. A red marker is added to indicate that 95% of the absolute errors are within 1.76°F. This implies that the model is once again effective in capturing the home thermal dynamics.

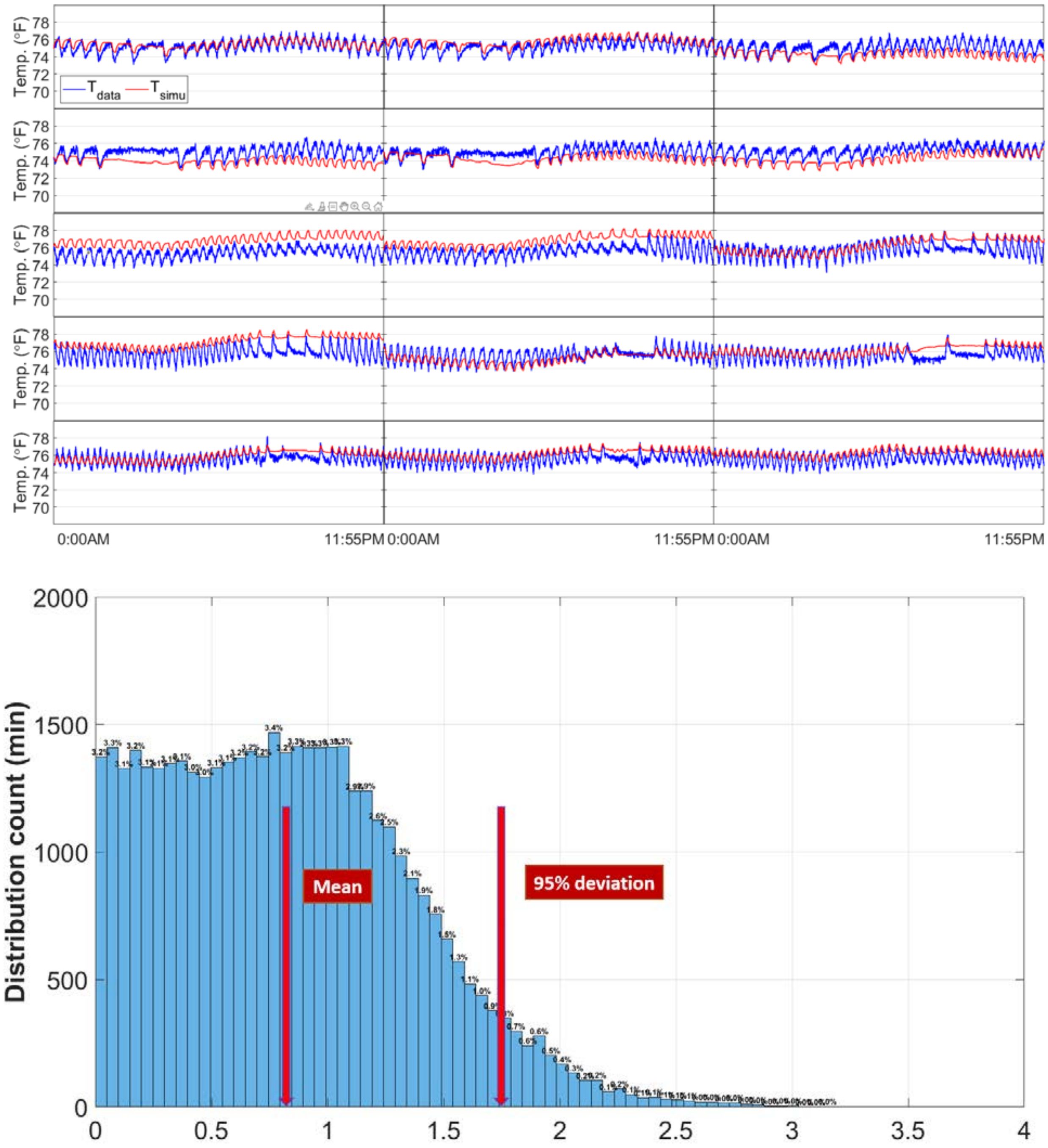


Figure 4.10. Simulation results using data from August 1–16, 2020 (total of 15 days). The red markers indicate the mean error and the error at 95% deviation.

Using the parameters identified in Table 4.5, we validate our model based on data from June, July, and August. The simulation results, along with a histogram of the absolute errors, are shown in Figure 4.11 to Figure 4.14. Table 4.6 summarizes the mean absolute error, maximum absolute error, and the absolute error at a 95% confidence interval. Compared with the least-squares method, the optimization method performs notably better. Indeed, all the absolute errors are lower than 1.96°F, even at a 95% confidence level, thus exceeding the success criterion. Therefore, we can conclude that the optimization method works well for both scenarios, i.e., either with the transition season data or HVAC season data. However, since the least-squares method is computationally much less demanding while providing satisfactory results for the transition season data, we recommend using the least-squares method when transition season data first become available, and using the optimization method when HVAC season data first become available.

Table 4.6. Absolute Error Comparison Using Thermocouple Temperature for Different Months for the Optimization Method

Date	Mean absolute error, °F	Maximum absolute error, °F	Absolute error at 95% confidence interval, °F
May 22–June 3	0.77	2.83	1.96
May 28–June 11	0.66	2.64	1.57
June 18–July 1	0.71	3.00	1.67
July 2–9, 17–20, 24–26	0.80	3.13	1.86

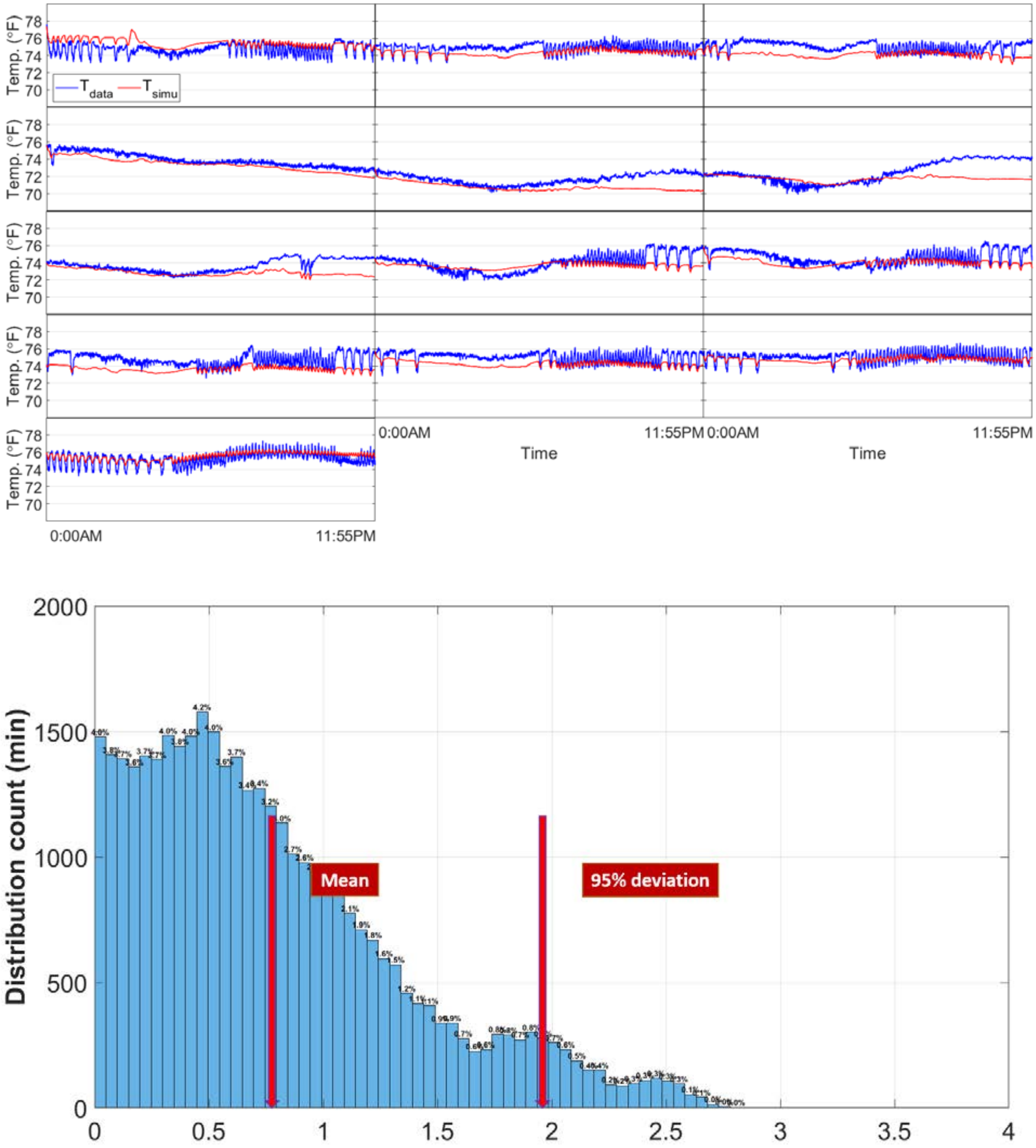


Figure 4.11. Validation results on May 22–June 3, 2020 (total of 13 days)

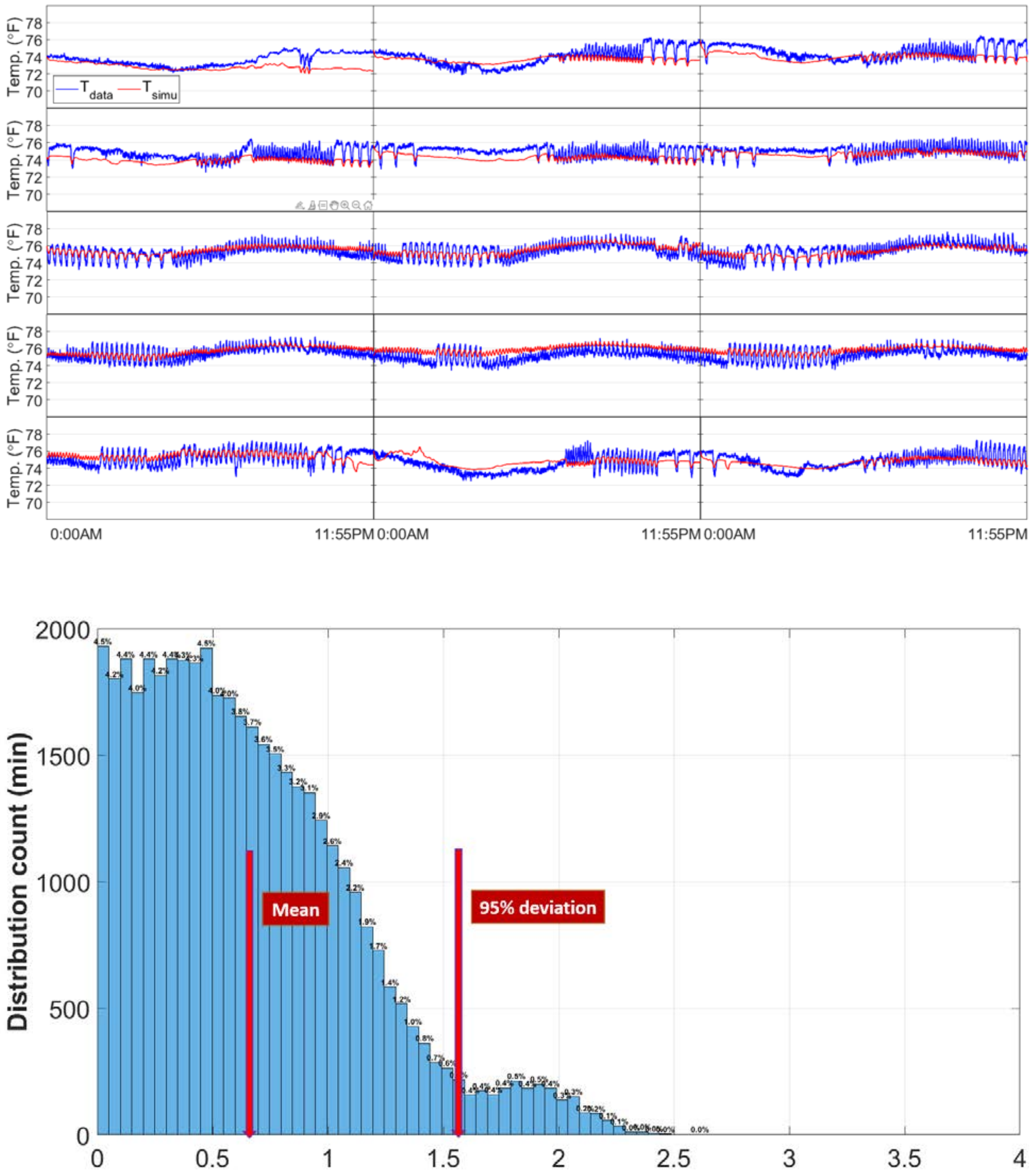


Figure 4.12. Validation using data from May 28–June 11, 2020 (total of 15 days)

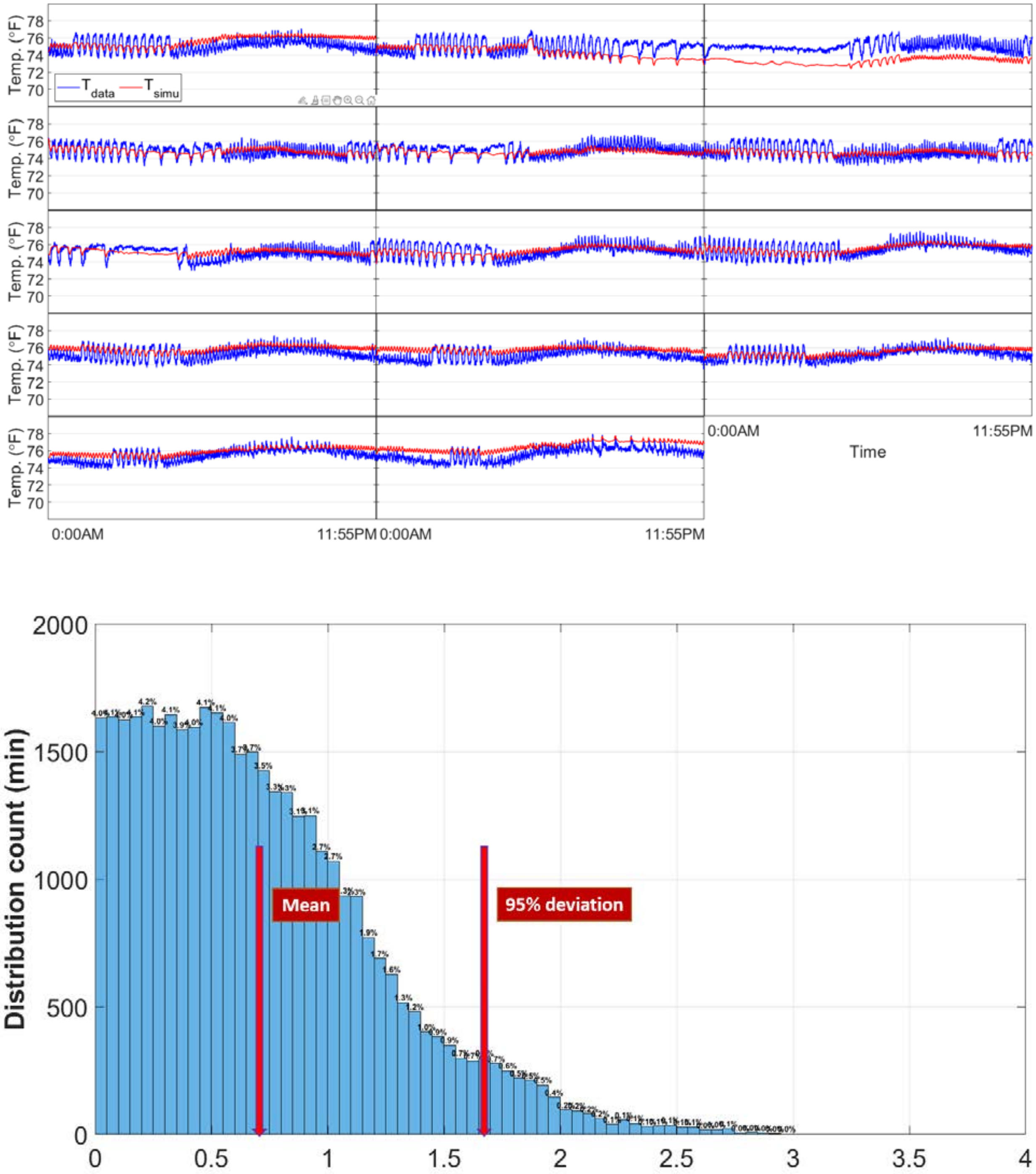


Figure 4.13. Validation using data from June 18–July 1, 2020 (total of 14 days)

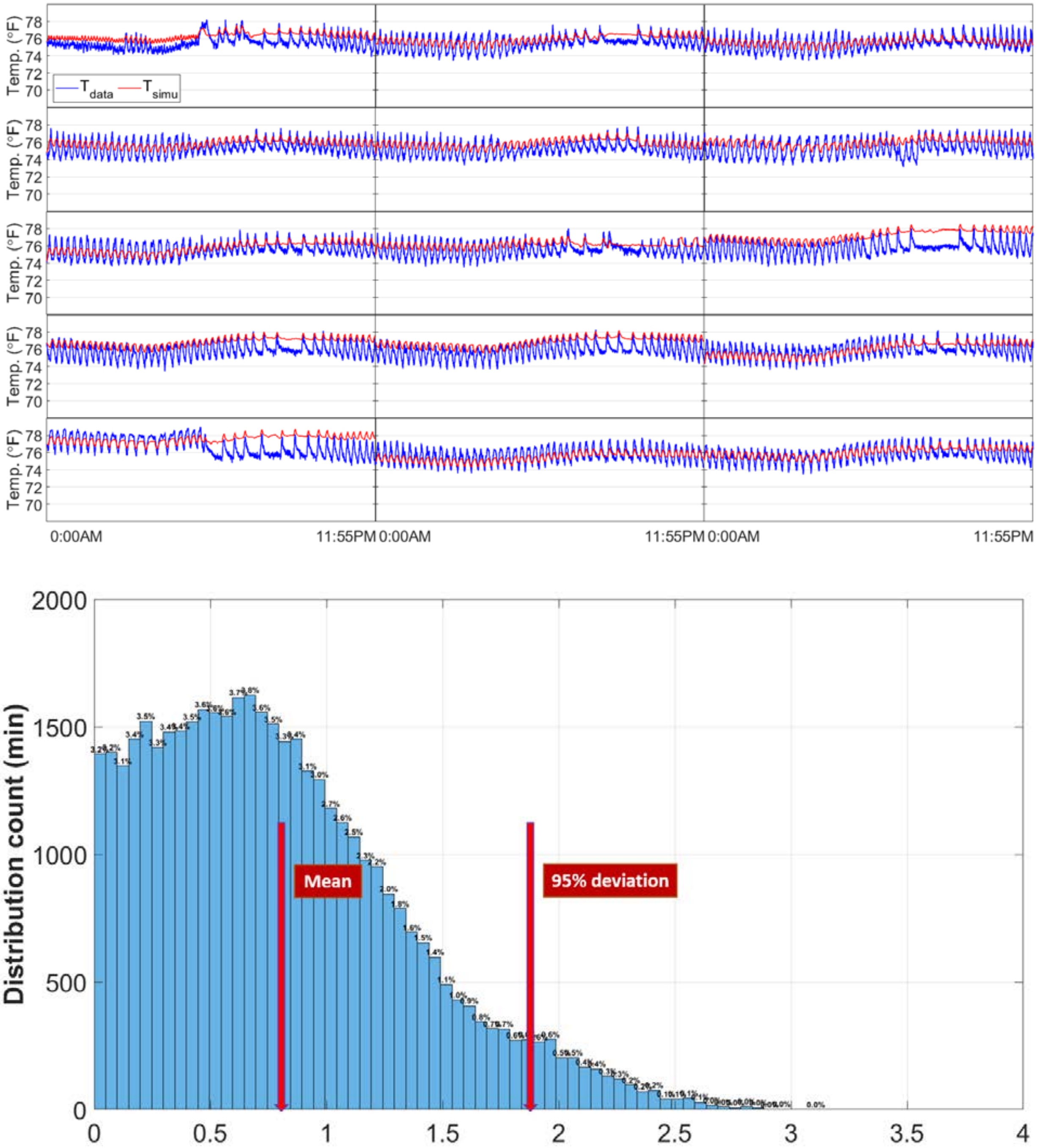


Figure 4.14. Validation using data from July 2–9, 17–20, and 24–26, 2020 (total of 15 days)

4.1.3 Tests Using Different Lengths of Training Data

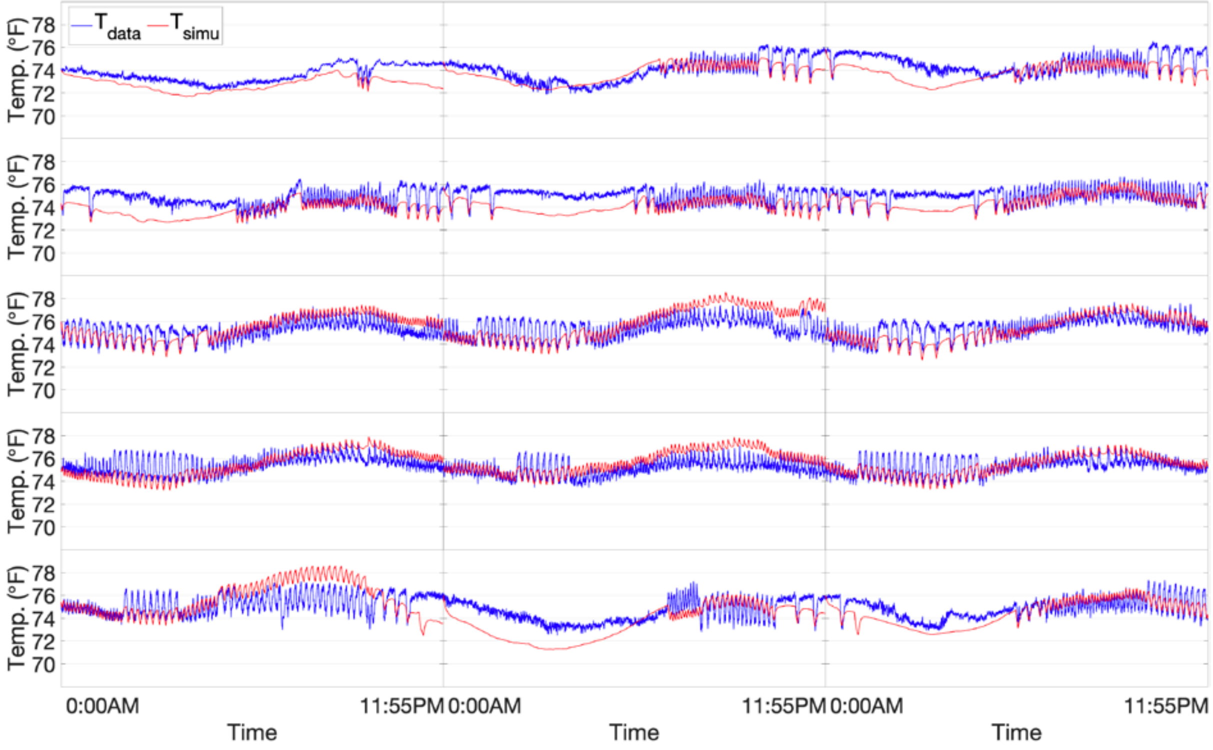
All the tests we have described so far are based on training the model using two weeks' worth of data (14 days). To understand how the length of data affects the results, we conducted another set of tests by varying the training data length, one with 6 days and the other with 14 days, as in previous sections. Here, for simplicity we consider only the least-squares method.

Table 4.7 shows the results of estimating the home thermal model parameters using the first 6 and first 14 consecutive days of training data collected in May, which is considered a transition season. The first 6 and first 14 consecutive days of training data are chosen because they are the earliest available data. Because two distinct sets of training data are used, we obtain two distinct home thermal models with different parameters, which are then used to perform simulations.

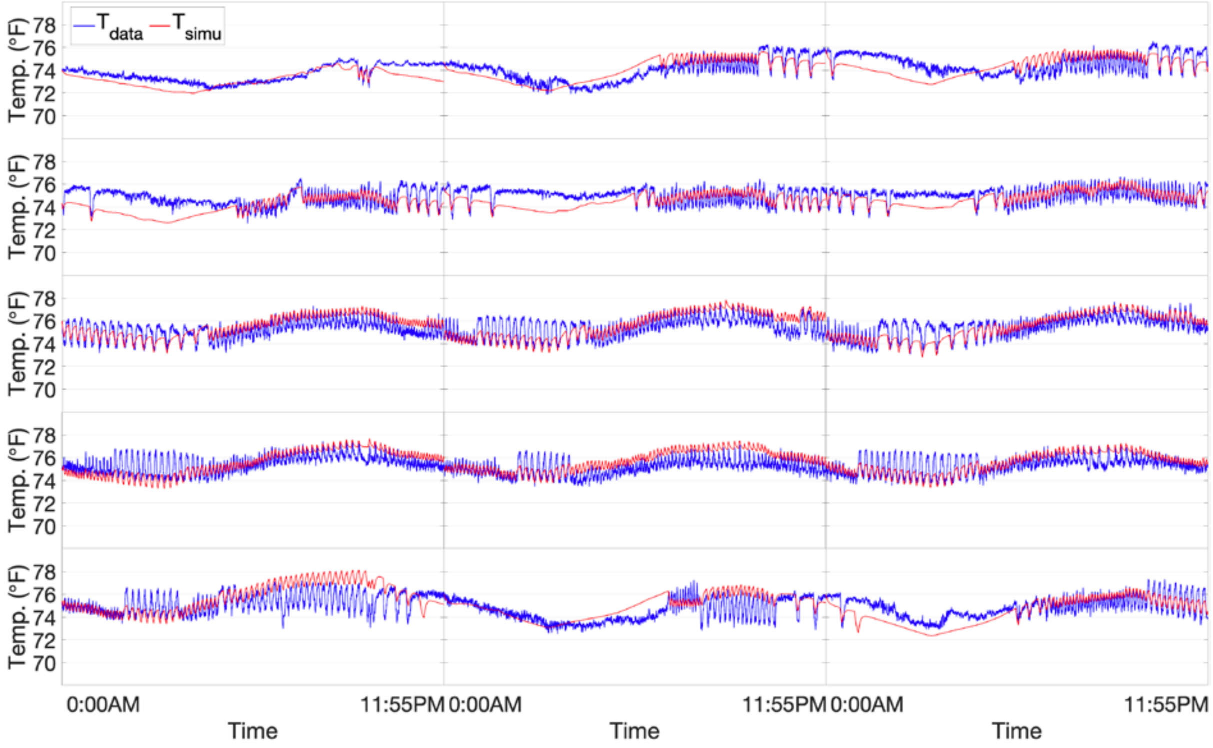
Table 4.7. Estimated Parameters for the Two Identified Models

Length of training data	τ_1	τ_2	τ_3	b_1	b_2
First 6 consecutive days	2636	278.6	7.584	0.01272	-0.0002349
First 14 consecutive days	2383	262.6	7.897	0.005447	-0.0002832
Length of training data	a_1	a_2	a_3	Q_s	
First 6 consecutive days	4.331	-8.457	5.218	-2.200	
First 14 consecutive days	5.422	-9.487	5.568	-2.086	

Figure 4.15 shows the measured and 24-hours-ahead predicted indoor air temperatures using the two models. From the figure, we see that the predicted temperatures follow the measured temperatures reasonably well, achieving mean absolute errors of 0.82°F and 0.80°F and absolute errors of 1.85°F and 1.90°F at a 95% confidence interval, respectively. Figure 4.16 compares the histograms of the absolute errors for the two models, from which we can see that the two models are equally effective in capturing the home thermal dynamics.

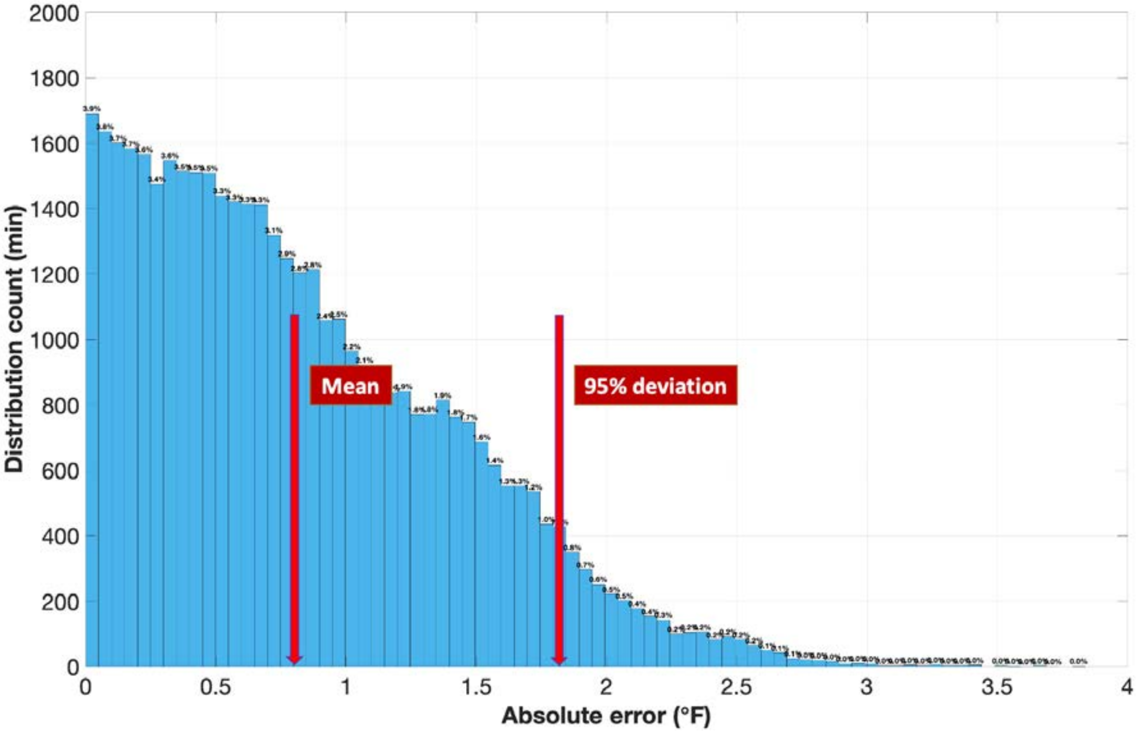


(a) for the home thermal model trained using the first 6 consecutive days of data

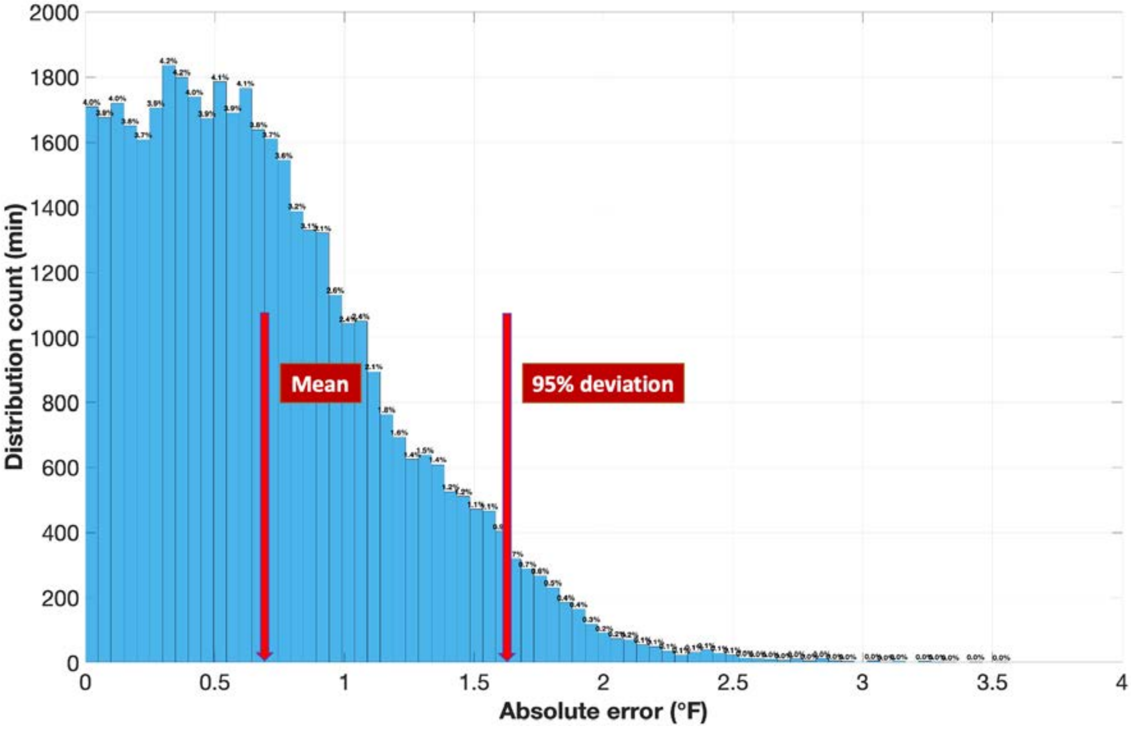


(b) for the home thermal model trained using the first 14 consecutive days of data

Figure 4.15. Comparison of measured and predicted indoor air temperatures



(a) for the home thermal model trained using the first 6 consecutive days of data



(b) for the home thermal model trained using first 14 consecutive days of data

Figure 4.16. Comparison of the distributions of the absolute errors. The red markers indicate the mean error and the error at a 95% confidence interval.

Analyzing Figure 4.15 and Figure 4.16, we observe that the simulation results for the two models, trained using the first 6 and first 14 consecutive days of data, are very close to each other. In addition, as shown in Table 4.8, the mean, maximum, and absolute errors at a 95% confidence interval for the model trained using the latter are slightly better than that using the former, but not by much. Therefore, it is fair to say that the two models have similar accuracy in predicting indoor air temperature 24 hours into the future. In other words, having longer training data does not actually bring significant advantages.

Table 4.8. Absolute Error Comparison for the Two Identified Models

Length of training data	Mean absolute error, °F	Maximum absolute error, °F	Absolute error at 95% confidence interval, °F
First 6 consecutive days	0.82	3.84	1.85
First 14 consecutive days	0.70	3.56	1.62

4.2 Model Verification Using Data From the Miami Test House

To investigate how the proposed model and parameter estimation methods perform in a region with a climate different from Oklahoma, we collected thermocouple data from the Miami test house for 15 days during the 2021 cooling season, from November 24 to December 9, at 1-minute intervals. In this section, we describe how we used the first 6 consecutive days of data to train the model and the last 6 days of data to validate its performance. The least-squares method is adopted throughout.

4.2.1 Model Verification Using the Least-Squares Method

4.2.1.1 Tests Using HVAC Season Data

Table 4.9 shows the results of estimating the home thermal model parameters using the first 6 days of data collected in November, which is considered an HVAC season.

Table 4.9. Training Results of Least-Squares Method Using November Data

Training dataset	τ_1	τ_2	τ_3	τ_4
November 25–November 30	1416.4677	1416.8466	265.4775	1320.1172
Length of training data	a_1	a_2	a_3	Q_s
Total 6 days	1.01990	-0.1289	0.006195	-10.8866

In Table 4.9, the two coefficients b_1 and b_2, which represent the wind impact on a house, have been replaced by a new coefficient called τ_4 due to the lack of wind speed measurements at the Miami house. To carry out the replacement, recall that the second part of the governing equation of the home thermal model is:

$$C_{air}\frac{dT_{in}}{dt}=\frac{T_{ie}-T_{in}}{R_{air}}+q'_{vw}+Q_{sol}+Q_{int}+Q_{sys}, \tag{4.17}$$

where q'_{vw} is the rate of heat transfer due to wind, which can be expressed by

$$q'_{vw}=\frac{T_o-T_{in}}{1/\left(b'_1W+b'_2W^2\right)}=\frac{T_o-T_{in}}{R_{vw}}.$$

Although wind speed measurements are not available, we still would like to account for the impact of q'_{vw} using the variable resistance R_{vw} instead of $\frac{1}{b'_1W+b'_2W^2}$. Dividing C_{air} on both sides of Equation (4.1), a new equation representing the second part of the home thermal model can be obtained as follows:

$$\begin{aligned}\frac{dT_{in}(t)}{dt}=&-\frac{1}{\tau_3}T_{in}(t)+\frac{1}{\tau_3}\left[T_{ie}(t)+\left(a_1G(t)+a_2G^2(t)+a_3G^3(t)\right)+\left(Q_iu_i(t)+Q_su_s(t)\right)\right]\\&+\frac{1}{\tau_4}\left[\left(T_o(t)-T_{in}(t)\right)\right],\end{aligned} \tag{4.18}$$

where $\tau_4=C_{air}R_{vw}$.

Figure 4.17 shows the measured and 24-hours-ahead predicted indoor air temperatures using the home thermal model with trained parameters from Table 4.9. Note that the predicted temperatures are able to closely follow the measured temperatures, achieving a mean absolute error of 0.30°F and a maximum absolute error of 1.23°F. The second part of Figure 4.17 shows a histogram of the absolute errors, where it can be seen that 90% of the absolute error are within 0.56°F. This encouraging result suggests that despite being applied to a different house in a different geographical region, the proposed model is capable of capturing the home thermal dynamics.

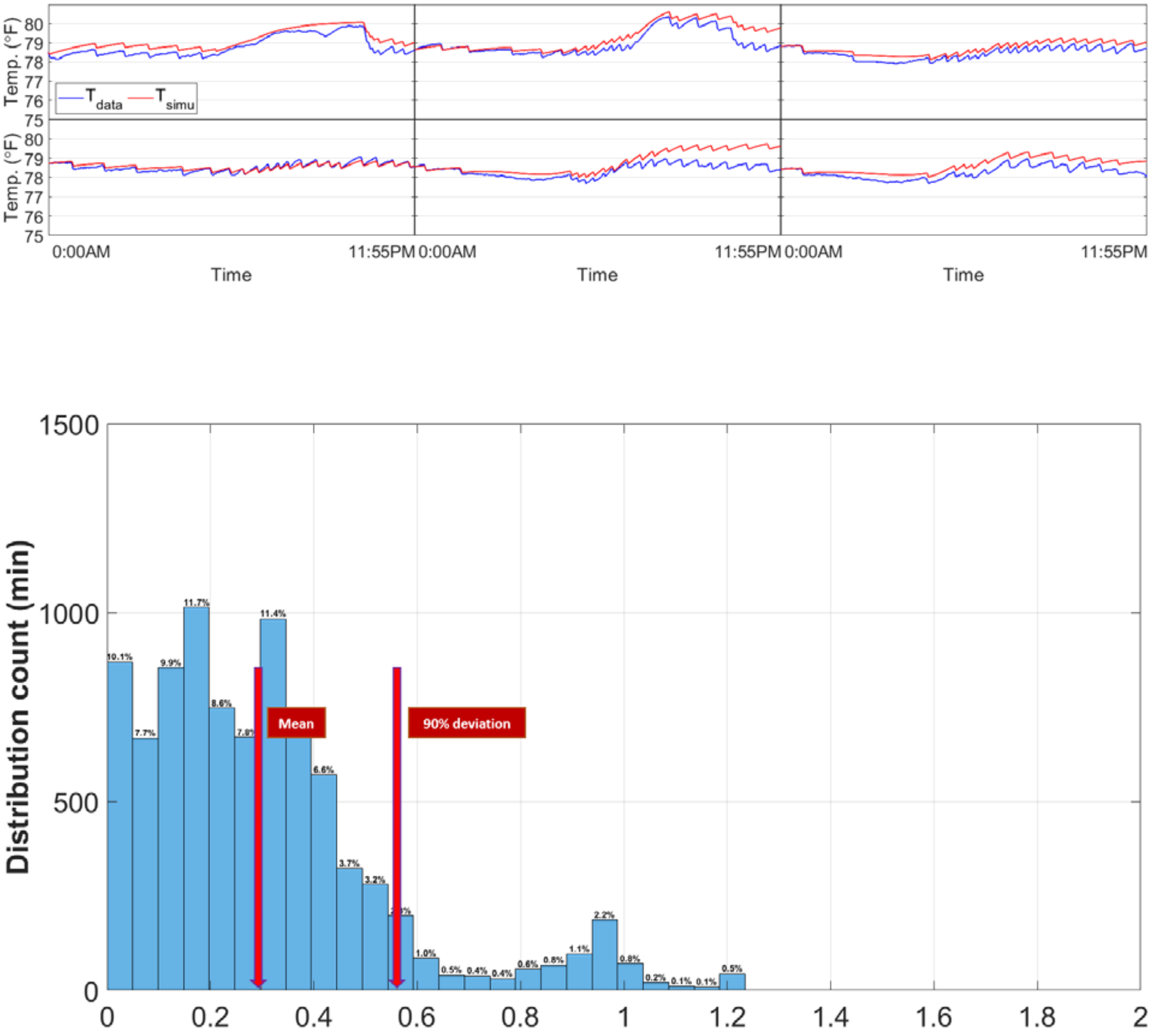


Figure 4.17. Simulation results using data from November 25–30, 2020 (total of 6 days)

Using the parameters identified in Table 4.9, we validate the model using data from December, which is considered a transition season. The simulation result and a histogram of the absolute errors are shown in Figure 4.18. Table 4.10 lists the mean absolute error, maximum absolute error, and the absolute error at a 90% confidence interval. **Observe that 90% of the absolute errors are lower than 1.42°F, thus satisfying the success criterion.**

Table 4.10. Absolute Error Comparison Using Thermocouple Temperature

Date	Mean absolute error, °F	Maximum absolute error, °F	Absolute error at 90% confidence interval, °F
December 1–December 8	0.64	1.94	1.42

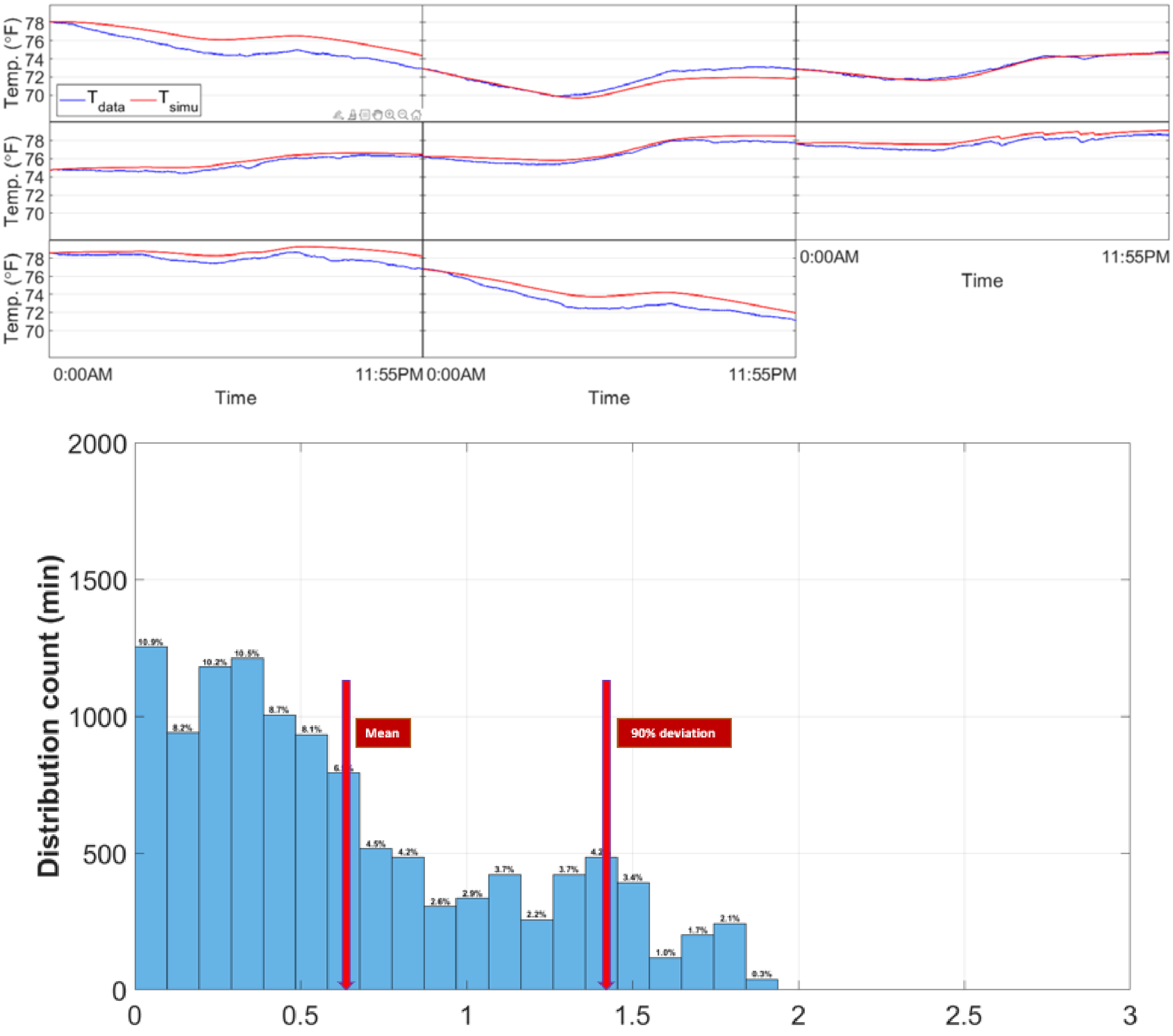


Figure 4.18. Validation results on December 1–8, 2020 (total of 8 days)

4.2.2 Tests Using Transition Season Data

Table 4.11 shows the results of estimating the home thermal model parameters using 8 days of data collected in December, which is considered a transition season in Miami.

Table 4.11. Training Results of Least-Squares Method Using December Data

Training dataset	τ_1	τ_2	τ_3	τ_4
December 1 – December 8	979.6927	529.7334	656.9250	1513.3107
Length of training data	a_1	a_2	a_3	Q_s
Total 8 days	0.3426	0.09117	-0.007995	-24.2813

Figure 4.19 shows the measured and 24-hours-ahead predicted indoor air temperatures obtained using the home thermal model with trained parameters from Table 4.11. Observe that the predicted temperatures track the measured temperatures well, achieving a mean absolute error of 0.48°F and a maximum absolute error of 1.46°F. The second part of Figure 4.19 shows a histogram of the absolute errors, where it is seen that 90% of the absolute errors are within 1.20°F. This adds to the evidence that the model is able to accurately capture the home thermal dynamics.

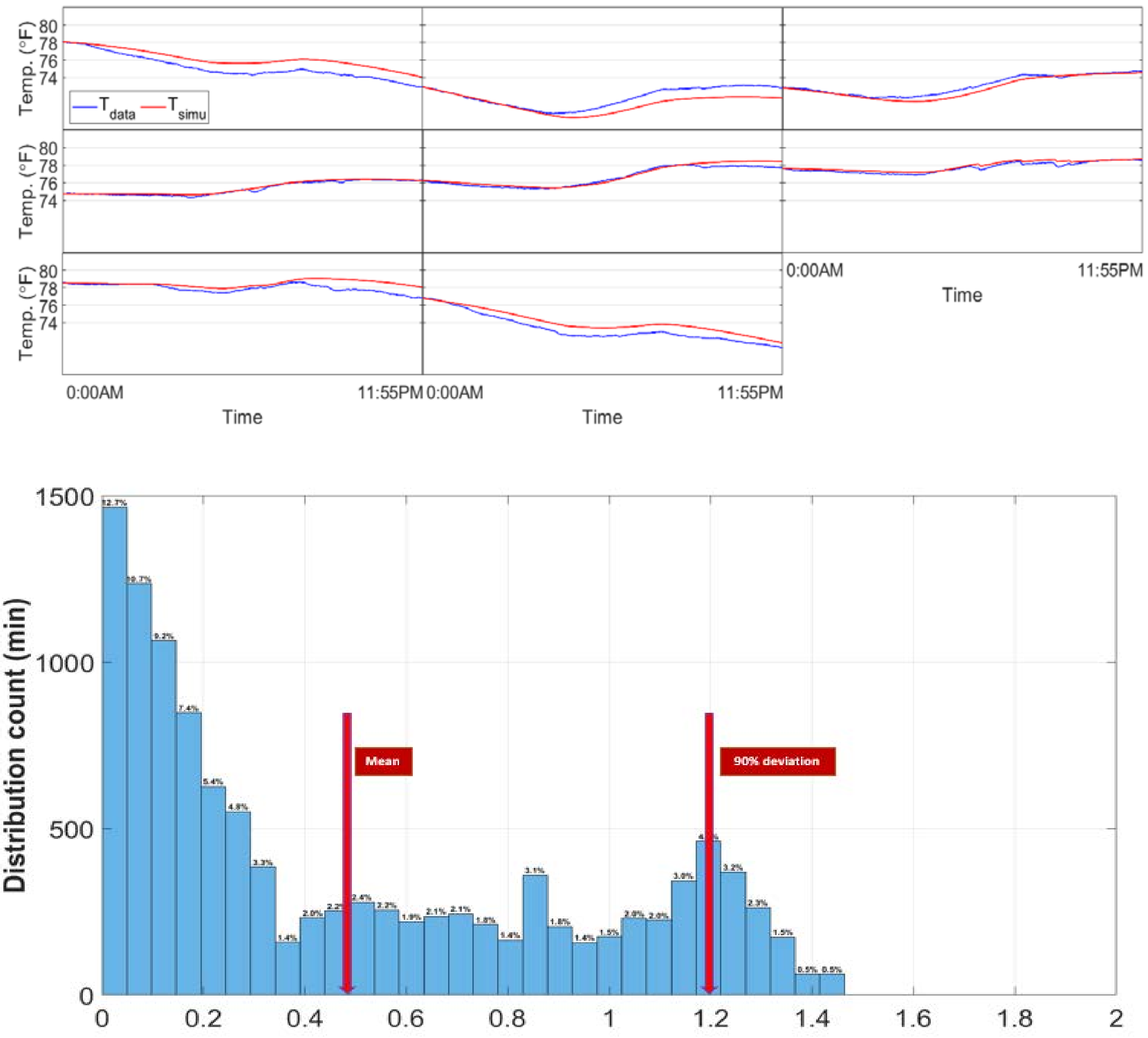


Figure 4.19. Simulation results using data from December 1–8, 2020 (total of 8 days)

Finally, using the parameters identified in Table 4.11, we validate the model based on data from November, which is considered an HVAC season in Miami. The simulation results and a histogram of the absolute errors are shown in Figure 4.20. Table 4.12 displays the mean absolute error, maximum absolute error, and the absolute error at a 90% confidence interval. **Because 90% of the absolute errors are lower than 0.68°F, the performance surpasses the success criterion.**

Table 4.12. Absolute Error Comparison Using Thermocouple Temperature for the Least-Squares Method

Date	Mean absolute error, °F	Maximum absolute error, °F	Absolute error at 90% confidence interval, °F
November 25–November 30	0.32	1.28	0.68

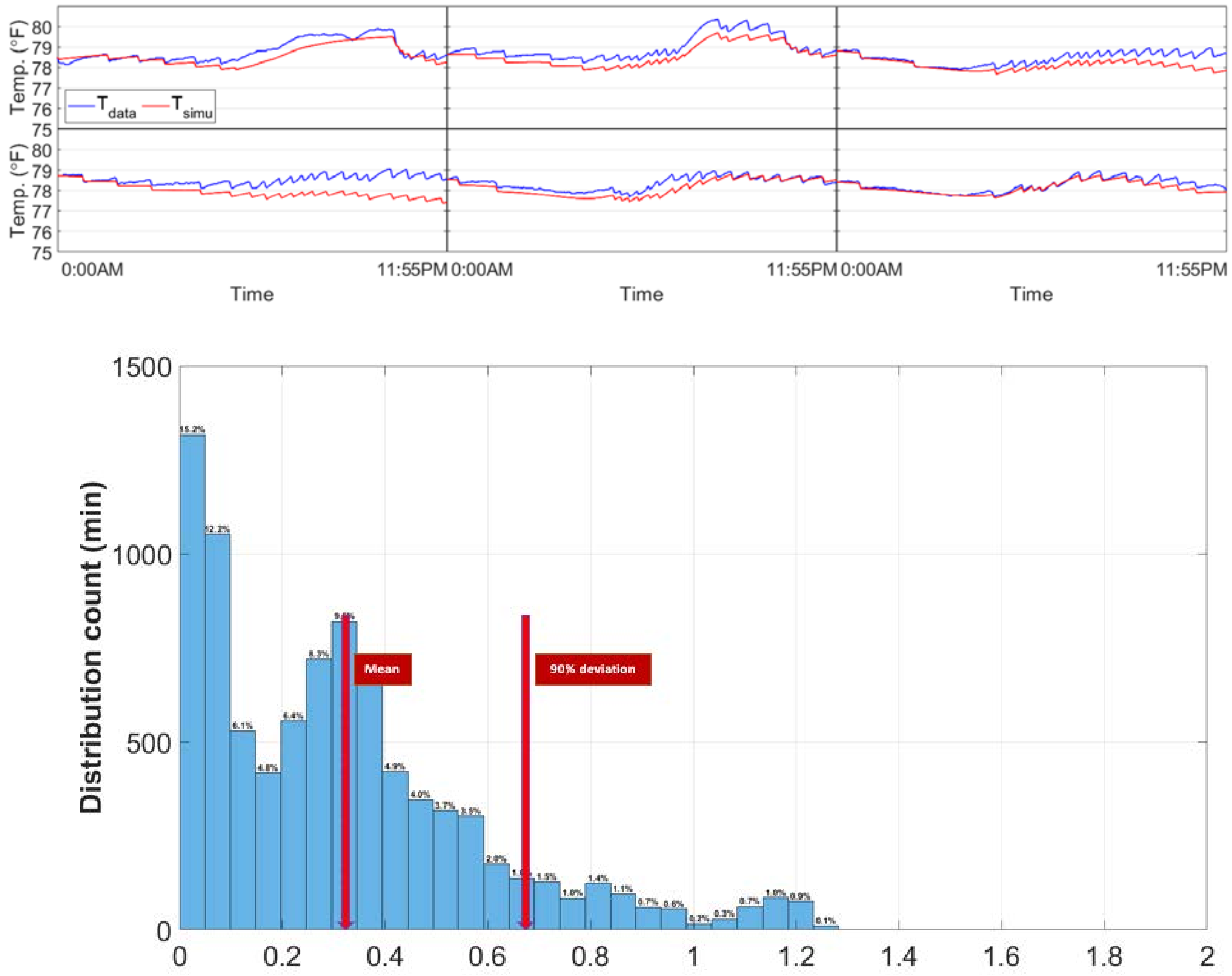


Figure 4.20. Validation results using data from November 25–30, 2020 (total of 6 days)

4.3 Model Verification Using Data From the PNNL Test Houses

Data from the two PNNL test houses were collected in a cooling season and used in model validation. The first set of data recorded the AC operation and building temperature at PNNL's Home A for a total of 14 days from July 3 to July 17, 2020, at 30-second intervals (July 15 has missing data). The second set of data contains the same information but is for PNNL's Home B. In Section 4.3.1, we use the first 6 consecutive days of data to estimate the model parameters for Home A and the last 6 consecutive days of data to validate the model prediction performance. In Section 4.3.2, we do the same for Home B. At the end of each section, we discuss data issues observed at both homes. The optimization method is applied throughout.

4.3.1 Test Results Using Data Collected From PNNL Home A

Table 4.13 shows the results of estimating the home thermal model parameters using the first 6 days of data collected in July from PNNL Home A.

Table 4.13. Training Results of Optimization Method Using PNNL Home A Data

Training dataset	$\boldsymbol{\tau_1}$	$\boldsymbol{\tau_2}$	$\boldsymbol{\tau_3}$	b_1	b_2
July 3–July 8	4694.8809	804.1474	4679.1909	168.7959	59.3138
Length of training data	$\boldsymbol{a_1}$	$\boldsymbol{a_2}$	$\boldsymbol{a_3}$	$\boldsymbol{Q_s}$	
Total 6 days	-190.9453	3.7499	-0.2754	-247.62	

Figure 4.21 shows the measured and 24-hours-ahead predicted indoor air temperatures using the home thermal model with trained parameters from Table 4.13. As can be seen, the predicted and measured temperatures are close to each other for the most part, and the mean absolute error is merely 0.94°F. However, the maximum absolute error is 4.99°F, which is quite large. The second part of Figure 4.21 displays a histogram of the absolute errors, from which one can see that the large maximum absolute error is a rare outlier. Indeed, 90% of the absolute errors are within 2.09°F.

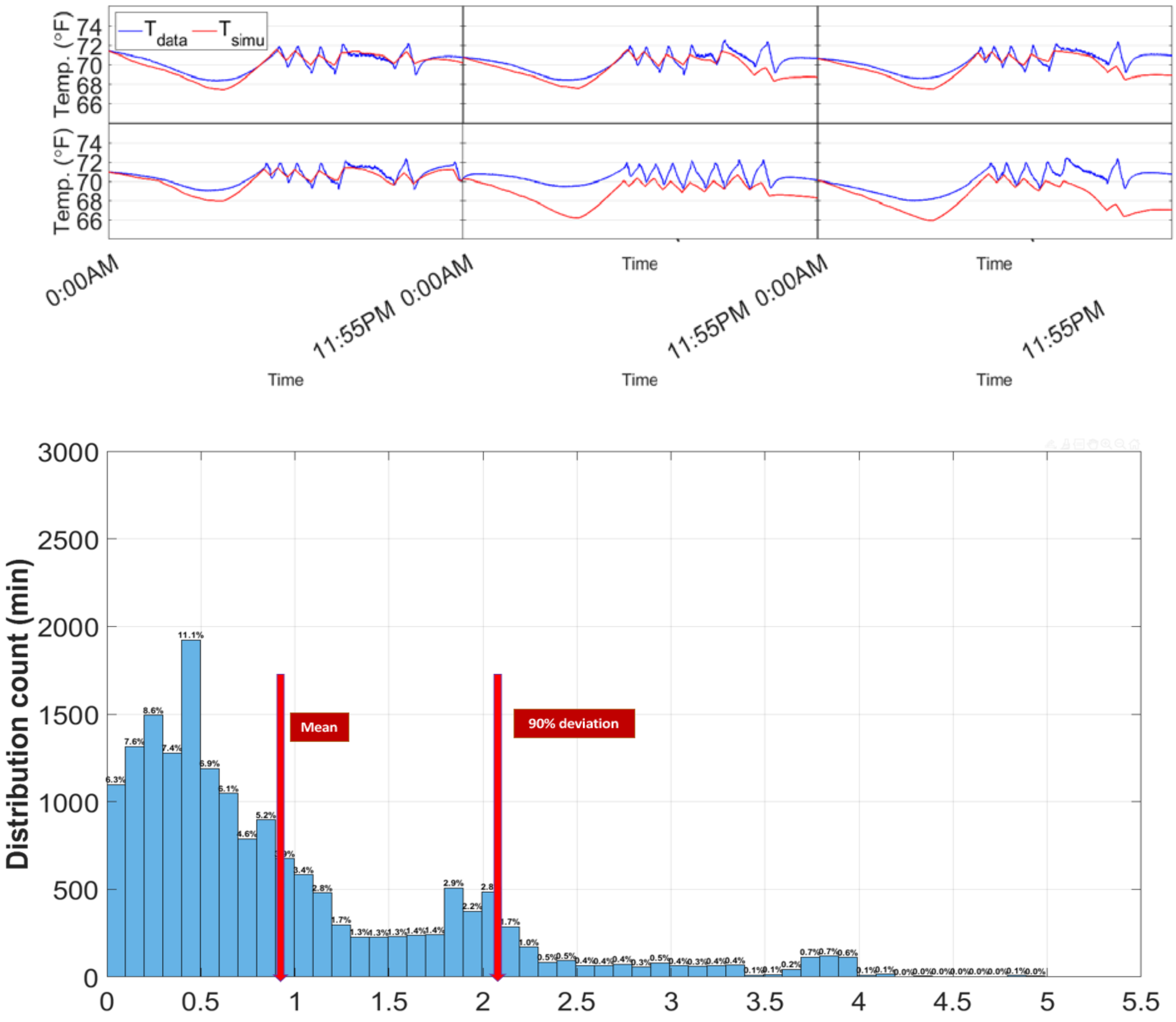


Figure 4.21. Simulation results using data from July 3–8, 2020 (total of 6 days)

Using the parameters identified in Table 4.13, we validate the model using the last 6 days of data, from July 9 to July 14, 2020. The simulation results and a histogram of the absolute errors are shown in Figure 4.22. Table 4.14 indicates the mean absolute error, the maximum absolute error, and the absolute error at a 90% confidence interval. Notice that the maximum absolute error of 6.12°F is very large.

Table 4.14. Absolute Error Comparison Using Thermocouple Temperature at PNNL Home A

Date	Mean absolute error, °F	Maximum absolute error, °F	Absolute error at 90% confidence interval, °F
July 9–July 14	1.18	6.12	2.60

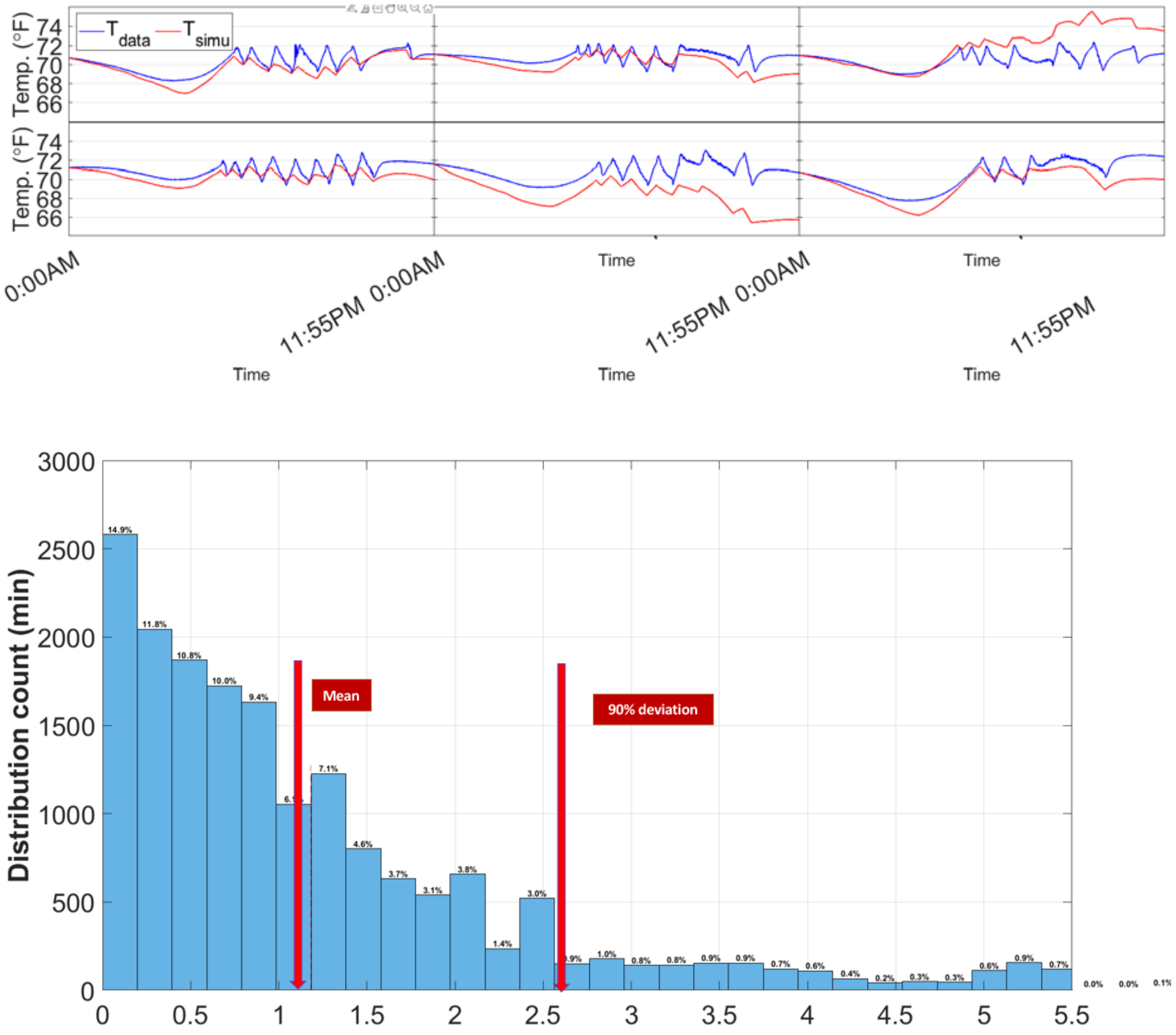


Figure 4.22. Validation results using data from July 9–14, 2020 (total of 6 days)

From the above simulation and validation results, we observe that the home thermal model is capable of capturing a building's thermal dynamics. However, its performance at PNNL's Home A is worse than that at both the Norman Test House and Miami Test House. In the following text, we provide an explanation of the discrepancy in performance. Physics-based gray box models such as our home thermal model could generate reasonably accurate predictions based on a relatively small amount of data. Therefore, data quality is important because it can significantly affect prediction accuracy. Figure 4.23 shows the two days of AC and building temperature data for PNNL Home A, from the training dataset of July 3 to July 8, 2020. The red curve is the indoor air temperature, the blue curve is the wall surface temperature, and the gray-shaded step plot is the AC on/off signal. As can be seen in Figure 4.23, the wall surface temperature went below the indoor air temperature when the AC was turned on, indicated by the green arrows. This is physically impossible. The wall surface measurement is between the outdoor air and indoor air and therefore impacted by the heat transfers with both. When the outdoor air temperature (the gray curve) was consistently high at above 80˚F (in fact, it was above 90˚F on July 3, 2020 and 85˚F on July 4, 2020), the only reason for the wall surface temperature (blue curve) to drop below 68˚F is by heat loss to the indoor air (red line). However, the indoor air temperature was either close to the wall surface temperature or above it. Therefore, there is a chance that the sensors for Home A measurements have errors. PNNL Home A did not originally have a wall surface temperature sensor installed. The wall surface temperature was measured using an add-on, portable temperature data logger. Because the home thermal model was trained using the PNNL Home A data, which had relatively low quality due to the errors, the prediction accuracy suffered as a result.

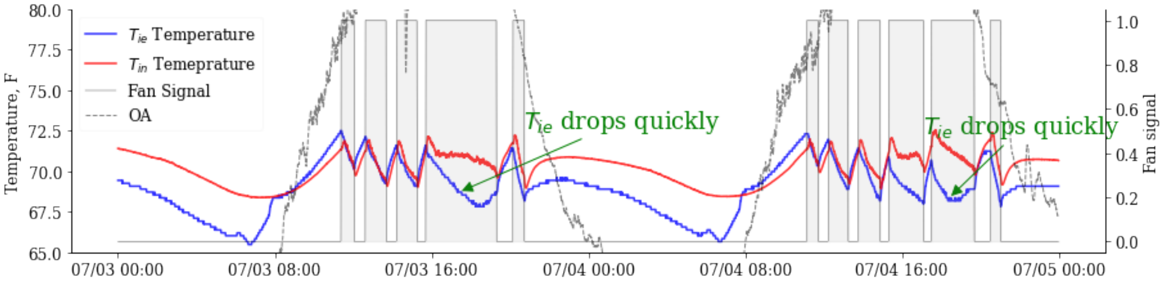


Figure 4.23. Indoor air temperature issue at PNNL Home A; data are from July 3–4.

4.3.2 Test Results Using Data Collected From PNNL Home B

Table 4.15 shows the results of estimating the home thermal model parameters using the optimization method. Figure 4.24 shows the measured and 24-hours-ahead predicted indoor air temperatures using the home thermal model with trained parameters from Table 4.15. It can be observed from the figure that the predicted temperatures match the measured temperatures very well. In fact, the histogram of absolute errors in Figure 4.24 shows that the mean absolute error is 0.7°F, the maximum absolute error is 2.9°F, and 90% of the absolute errors are within 1.6°F. However, although the simulation results look good, one of the parameters, $\tau 3$, turns out to be negative when it is expected to be positive. This is likely a consequence of problematic data, which will be explained in the following text.

Table 4.15. Training Results of Optimization Method Using PNNL Home B Data

Training dataset	τ_1	τ_2	τ_3	b_1	b_2
July 3–July 8	5332.76626	624.23289	-1589.3526	-1.1128	0.14402
Length of training data	a_1	a_2	a_3	Q_s	
Total 6 days	-53.9853	70.5305	-29.2909	80.3441	

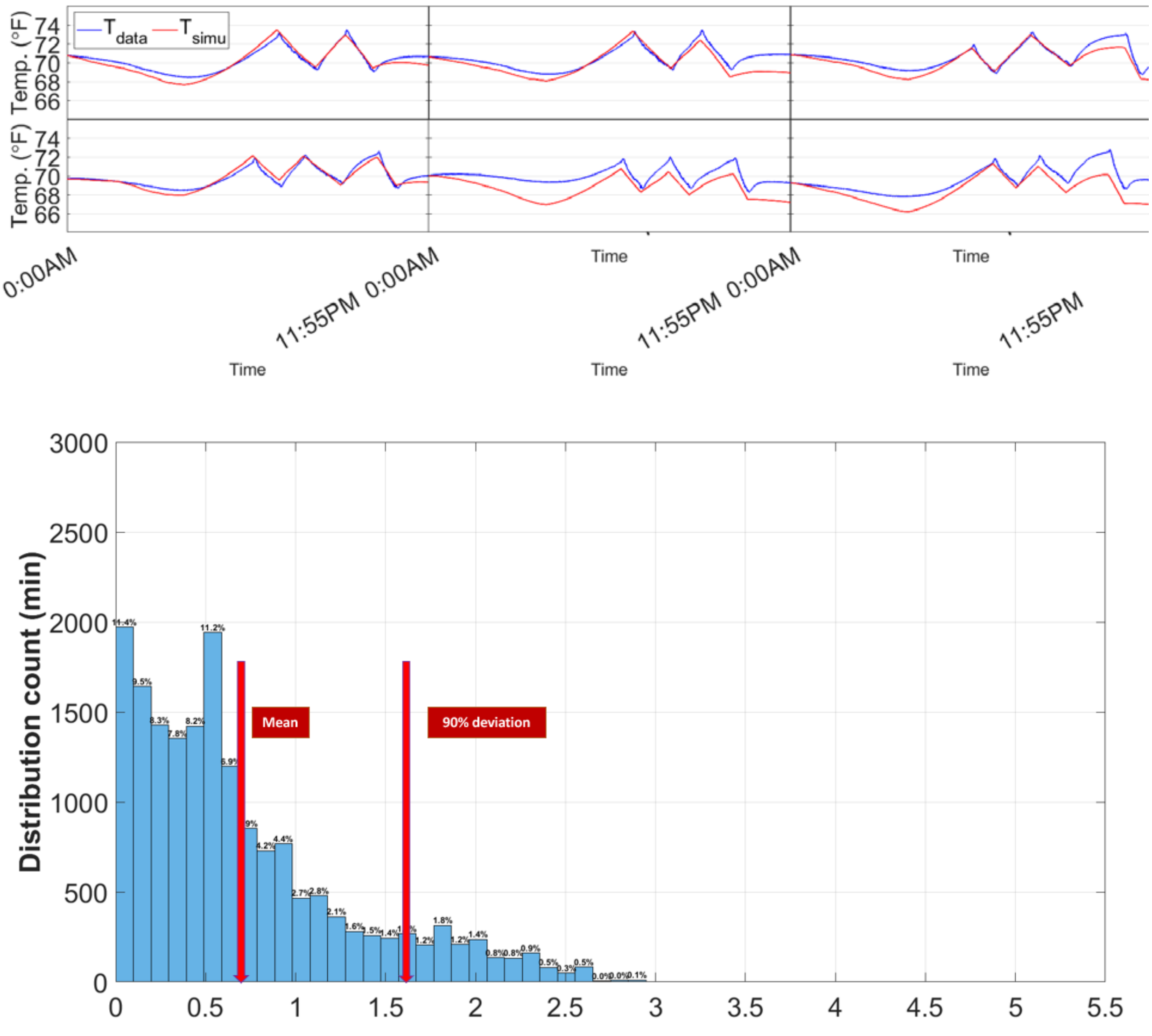


Figure 4.24. Simulation results using data from July 3–8, 2020 (total of 6 days)

Figure 4.25 shows the two days of AC operation and building temperature data for PNNL Home B in the training set from July 3 to July 8, 2020. The red curve is the indoor air temperature, the blue curve is the wall surface temperature, the gray dashed curve is the outdoor air temperature, and the AC signal is shown using a gray shaded step plot. From this figure, we see similar but more severe errors compared to PNNL Home A. The wall surface temperature went much lower than the indoor air temperature when the AC was turned on, indicated by the green arrow. This severe sensor error, however, led to τ_3 being negative. These factors suggest that data quality control is necessary before performing model identification.

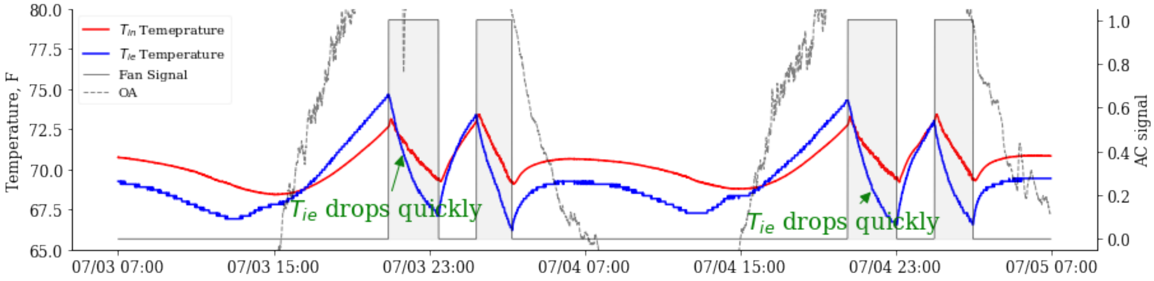


Figure 4.25. Temperature measurement issue at PNNL Home B; data are from July 3–July 4, 2020.

5 Development and Verification of Optimal Operation Algorithm

An MPC-based optimization agent has been developed to achieve optimal AC operation for cost savings. The MPC-based optimization agent adopts the second-order home thermal model developed in Section 3. A schematic that represents the MPC-based optimization agent is shown in Figure 5.1. The agent makes use of a set of thermal comfort criteria, the utility rate, and the HVAC system power use.

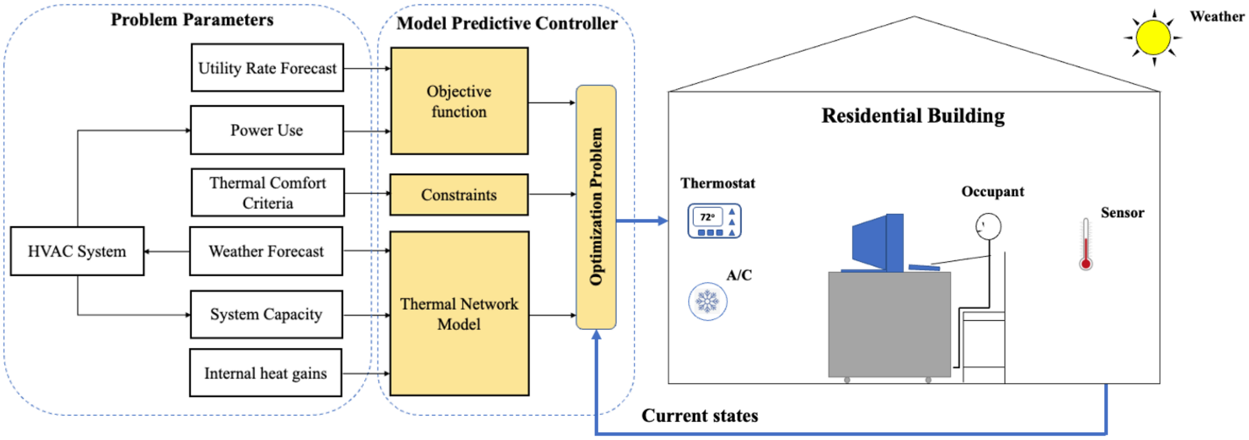


Figure 5.1. Schematic of MPC-based optimization agent

5.1 MPC-Based Optimization Methodology

Consider a home equipped with an HVAC system. Let Δt, i, N_p, and $\mathcal{H}_k = \{k,\ k+1, \ldots, k+N_p-1\}$ denote the sampling period, discrete time, total number of discrete-time slots over a prediction horizon, and the set of time indices in a prediction horizon at the current time k, respectively. In addition, let $P[i]$, $E[i]$, $T_{ie}[i]$, and $T_{in}[i]$ denote the utility rate, HVAC system power use, interior wall surface temperature, and indoor air temperature of the home, respectively. The HVAC on/off control signal $u_s[i]$ (i.e., $u_s[i] = 1$ if on and $u_s[i] = 0$ if off) is treated as an optimization variable. Specifically, at each time k, $u_s[i]$ for $i \in \mathcal{H}_k$ is chosen to minimize the objective function

$$J(u_s) = \sum_{i \in \mathcal{H}_k} P[i]E[i]u_s[i] \tag{5.19}$$

subject to the HVAC control constraint

$$u_s[i] \in \{0, 1\} \tag{5.20}$$

the temperature-based thermal comfort criteria

$$T_{lb}[i] \leq T_{in}[i] \leq T_{ub}[i] \tag{5.21}$$

and the discretized second-order thermal network model (Wang et al. 2022)

$$T_{ie}[i+1] = (1-\alpha_1-\alpha_2)T_{ie}[i] + \alpha_1 T_o[i] + \alpha_2 T_{in}[i] \tag{5.22}$$

$$\begin{aligned} T_{in}[i+1] = (1-\alpha_3)T_{in}[i] \\ &+ \alpha_3(T_{ie}[i] + (T_o[i]-T_{in}[i])(b_1 W[i] + b_2 W^2[i]) \\ &+ (a_1 G[i] + a_2 G^2[i] + a_3 G^3[i]) + Q_{int}[i]u_{int}[i] + Q_s[i]u_s[i]) \end{aligned} \tag{5.23}$$

where $\alpha_1 = \frac{\Delta t}{\tau_1}$; $\alpha_2 = \frac{\Delta t}{\tau_2}$; and $\alpha_3 = \frac{\Delta t}{\tau_3}$; τ_1 and τ_3 are time constants of the virtual envelope and indoor air of a home, respectively; τ_2 is a parameter representing the combined effect of the virtual envelope and indoor air; $T_o[i]$ is the outdoor air temperature; $T_{lb}[i]$ and $T_{ub}[i]$ are lower and upper bounds on $T_{in}[i]$ that ensure thermal comfort; a_1, a_2, and a_3 are parameters representing the effect of solar radiation; b_1 and b_2 are parameters representing the effect of wind; $G[i]$ is the global horizontal solar irradiation; $W[i]$ is the wind speed; $Q_{int}[i]$ is the internal heat gain; $Q_s[i]$ is the HVAC system output (also known as scaled cooling capacity); and $u_{int}[i]$ is the home occupancy signal (i.e., $u_{int}[i] = 1$ if occupied and $u_{int}[i] = 0$ otherwise). The home thermal model in Equations (5.4) and (5.5) can also be expressed in the form

$$\begin{bmatrix} T_{ie}[i] \\ T_{in}[i] \end{bmatrix} = \Phi[i,1]\begin{bmatrix} T_{ie}[1] \\ T_{in}[1] \end{bmatrix} + \sum_{j=2}^{i} \Phi[i,j]\,(B[j-1]u_s[j-1] + d[j-1]) \tag{5.24}$$

where $T_{ie}[1]$ and $T_{in}[1]$ are initial values, which are given; $\Phi[i,j] = \prod_{p=j}^{i-1} A\,[p]$ is the state transition matrix (Brogan 1991); and $A[i], B[i], \text{and } d[i]$ are matrices given by

$$A[i] = \begin{bmatrix} 1-\alpha_1-\alpha_2 & \alpha_2 \\ \alpha_3 & (1-\alpha_3) - \alpha_3(b_1 W[i] + b_2 W^2[i]) \end{bmatrix}$$

$$B[i] = \begin{bmatrix} 0 \\ \alpha_3 Q_s[i] \end{bmatrix}$$

$$d[i] = \begin{bmatrix} \alpha_1 T_o[i] \\ \alpha_3(T_o[i](b_1 W[i] + b_2 W^2[i]) + (a_1 G[i] + a_2 G^2[i] + a_3 G^3[i]) + Q_{int}[i]u_{int}[i]) \end{bmatrix}.$$

The optimization problem is an integer linear program (ILP) (Schrijver 1998) that may be solved using, for example, the CVX framework (i.e., a MATLAB-based modeling system for convex optimization (Grant and Boyd 2014; MathWorks 2020) along with a GUROBI solver (Gurobi Optimization 2021)). A solution to the optimization problem in Equations (5.1)−(5.5) is a sequence of optimal HVAC control signals $u_s^*[i]$ for $i \in \mathcal{H}_k = \{k,\ k+1, \dots, k+N_p-1\}$ that minimizes energy costs while maintaining an acceptable level of temperature-based thermal comfort over a prediction horizon. As the prediction horizon keeps being shifted forward, a new sequence of control signals, namely AC on/off command, is obtained at each time k. However, only the first element $u_s^*[k]$ of this sequence is applied, and the rest of the elements $u_s^*[k+1]$, $u_s^*[k+2]$, …, $u_s^*[k+N_p-1]$ are disregarded.

5.2 MPC-Based Optimization Agent Validation in Norman Test House

Voltron-based application code has been developed to test the MPC-based optimization agent in the Norman Test House. This lab test was conducted between October and November 2021, when the cooling season was almost finished. Therefore, the purpose of the lab test was to verify that the agent was implementable before we deployed the agent in 10 participating homes for field testing in Summer 2022. It was not intended for energy performance evaluation because it was not a typical HVAC season.

For testing purposes, we created different pricing signals throughout the day to see if the MPC-based optimization agent was able to run AC heavily in lower priced hours and reduce/eliminate AC operations during higher priced hours for cost savings. Figure 5.2 presents the initial test results. The test started the evening of October 10, 2021. As can be seen from the blue shaded areas, we intentionally made the price higher, at 42 cents per kWh from 11:30 p.m. to 4:30 a.m., and set it to 24 cents during other hours. Meanwhile, the indoor air temperature band was set between 62˚F and 72˚F to allow temperature to float within 10˚F for testing purposes. The outdoor unit power measurement (the brown line in Figure 5.2) represents the AC on/off operation. When the outdoor power was close to 2 kW, it indicated that the AC was on. When the power measurement was close to zero, it indicated that the AC was off. Although the AC indeed did not turn on for most of time during higher priced hours, it was not obvious that the reduction/elimination of the AC operation in the higher priced hours was due to the MPC agent. Because the outdoor air temperature experienced a sudden decrease the night of October 10, 2021, the AC did not need to turn on in the lower priced hours or the following day, October 11, 2021. Therefore, the test was allowed to continue.

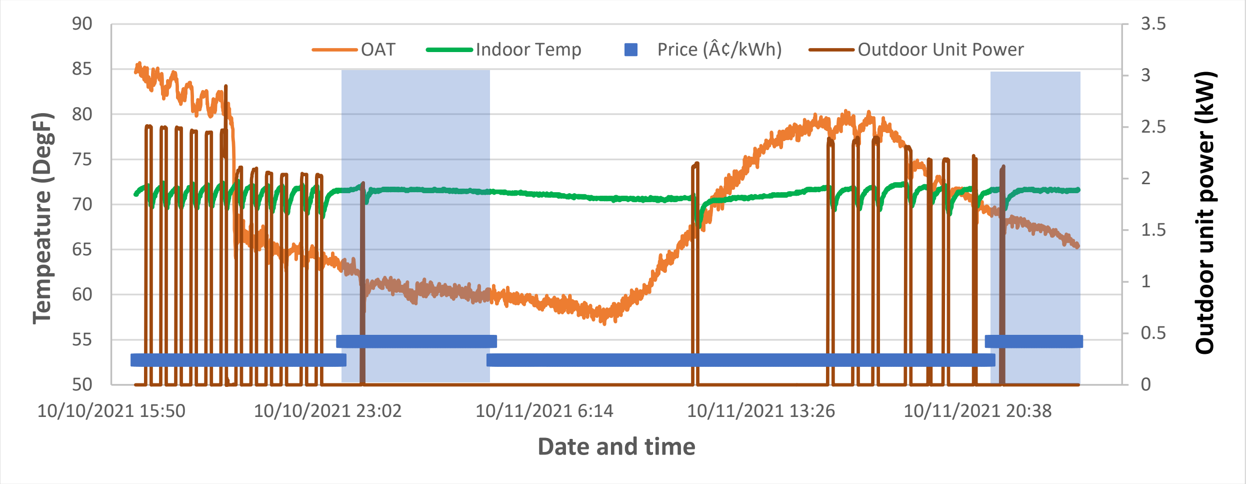


Figure 5.2. Initial MPC agent test results on October 11, 2021

Figure 5.3 shows the MPC agent test results on November 11, 2021. In Figure 5.3, the weather continued to get cold, with temperature ranging between 45˚F and 65˚F. To force a cooling operation so that the MPC agent would operate, the indoor temperature band was reduced to between 58˚F and 66˚F. The higher priced hours were also adjusted from 4 p.m. to 10 p.m. With these changes, AC operation in the higher priced hours was completely eliminated. In addition, the AC ran intensively prior to the higher priced hours to cool the indoor air down to approximately 58˚F, which was the lower bound of the temperature band, so that it could take advantage of the lower electricity price. The MPC agent operation is therefore proven to be effective by integrating the electricity price signal into its optimal decision-making.

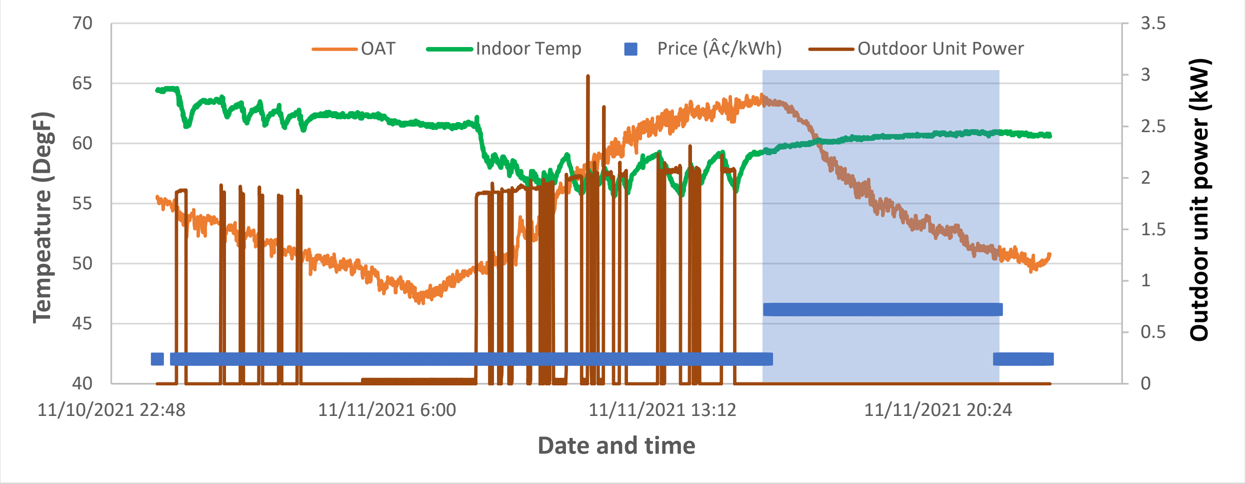


Figure 5.3. Additional MPC agent test results on November 11, 2021

6 Development and Verification of FDD Algorithm

6.1 Faults Considered

Figure 6.1 shows the various typical faults in residential HVAC systems and the extent to which they have been studied as expressed in the form of percentages. The information is obtained by analyzing 35 relevant publications we reviewed (Ejenakevwe and Song 2021). Incorrect refrigerant charge levels (RCL) and incorrect airflow rate are the most researched, as noted in related work (Chintala, Winkler, and Jin 2021; Cetin and Kallus 2016). Refrigerant leakage, which is the third most researched topic, can be classified with low RCL. Duct leakage is included as a common fault (Rogers, Guo, and Rasmussen 2019; Chintala, Winkler, and Jin 2021), but it can also be considered under incorrect airflow since it leads to low airflow. Also, from the chart, sensor fault is the least studied for residential HVAC systems. This is because residential HVAC systems typically have far fewer sensors compared to commercial HVAC systems. Thus, incorrect RCL and AFR are the two most dominant residential AC faults, and both contribute to changes in evaporator performance, which can be monitored by enthalpy changes between return air (RA) and supply air (SA). However, when both faults occur at the same time, enthalpy changes between RA and SA would not be an effective indicator for diagnosing and differentiating the two faults. In this case, we propose to use an indoor power meter to diagnose the AFR fault from indoor fan power measurements. Therefore, we have developed two different FDD methods for situations with and without power meter measurements.

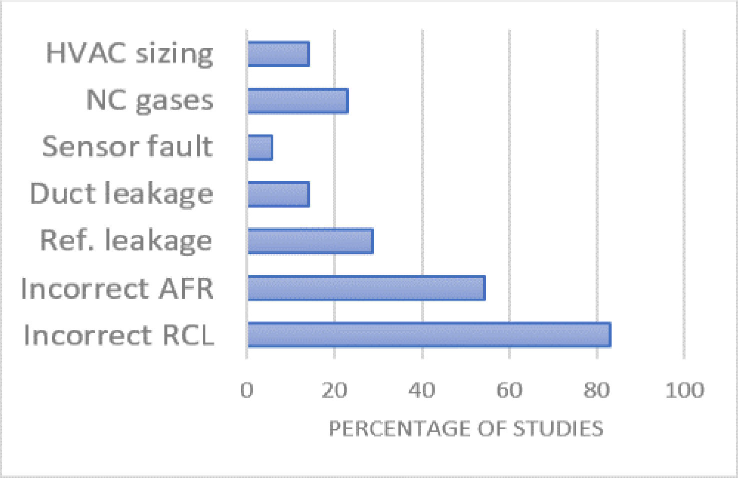


Figure 6.1. Level of studies on typical faults in residential HVAC systems

6.2 FDD Methodology

Figure 6.2 gives an illustration of the proposed Internet of Things (IoT)-based AFDD approach. Based on the aim of the proposed AFDD, which is to check for both installation and degradation faults, the FDD agent is implemented and investigated in two phases. In Phase 1, testing is performed to identify possible installation issues/faults, while Phase 2 is used to check for faults resulting from system degradation. Because Phase 2 checks for degradation, it commences only after the completion of Phase 1. Also, during Phase 1, the data collected and used for installation fault check are then used as baseline data for Phase 2. Before testing, the AFDD algorithm was developed.

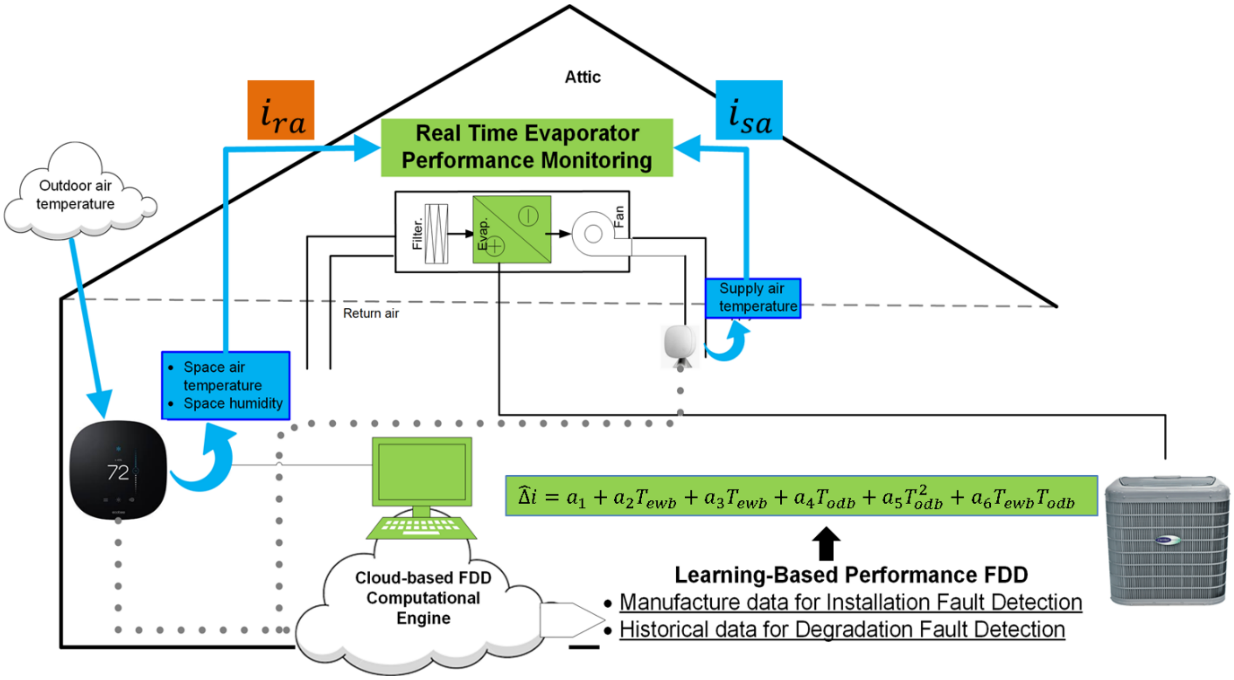


Figure 6.2. Schematic representation of proposed AFDD

6.3 Algorithm Development

6.3.1 Overview of Proposed FDD Methods

As mentioned, we have developed two different methods for FDD purposes with and without power meter measurements.

Method 1: For units without smart meter installation, enthalpy changes between return air (RA) and supply air (SA) can be used to diagnose the RCL and AFR faults. When an AC unit experiences the RCL fault, in general it is due to charge leaks. Therefore, dominant RCL faults are undercharged. With undercharged AC units, the fault can be detected by enthalpy changes between RA and SA, which are smaller than their normal values according to the theoretical vapor compression cycle. In this case, their normal values can be determined from manufacturer data (Phase 1) or collected baseline data at the beginning of the FDD installation (Phase 2). On the other hand, RCL faults can also be detected and differentiated by enthalpy changes between RA and SA because they show larger values than normal operation for the AC units with a constant-speed indoor fan. This study focuses on constant-speed, belt-driven fans as they are one of the most popular types of AC units used in U.S. homes. In this method, enthalpy changes between RA and SA can be measured/calculated using smart thermostat data and additional smart node sensors along with some reasonable engineering assumptions. A detailed discussion of this method is provided in Section 6.3.2.

Method 2: For units with smart meter installation along with smart thermostats, power measurements are particularly useful when both RCL and AFR faults occur at the same time. In this method, indoor fan power measurements are used to diagnose the AFR fault because for residential AC configuration (no moving dampers to cause resistance changes), fan power can directly reflect the flow rate variations caused by flow rate related faults for the most prevalent typical residential AC units driven by a conventional induction motor, which is the scope of this study. Then, enthalpy changes between RA and SA are used to diagnose the RCL fault. By evaluating the AFR fault first using fan power measurements, enthalpy changes between RA and SA can be used to diagnose overcharge, undercharge, and occurrent faults with AFR.

6.3.2 Proposed FDD Method 1

The approach illustrated in Figure 6.2 uses the enthalpy change across the evaporator coil as a fault detection index. The measured enthalpy change (Δi) is compared with the predicted enthalpy change ($\Delta \hat{\imath}$), as shown in Equation (6.1):

$$\delta_{\Delta i} = \Delta \hat{\imath} - \Delta i \tag{6.1}$$

With this approach, if $\delta_{\Delta i}$ exceeds a set deviation threshold, $\overline{\delta_{\Delta \imath}}$, a fault is suspected. The predicted $\Delta \hat{\imath}$ is calculated using a regression model with two measured inputs, namely the coil entering wet-bulb temperature and the outdoor air-dry bulb temperature, which have the most impact on AC cooling output. The regression model is given by

$$\Delta \hat{\imath} = \frac{a_1 + a_2 T_{ewb} + a_3 T_{ewb}^2 + a_4 T_{odb} + a_5 T_{odb}^2 + a_6 T_{ewb} T_{odb}}{60 \rho_{sa} Q_{sa}} \tag{6.2}$$

where $\Delta \hat{\imath}$ is the predicted enthalpy difference, in Btu/lbm (J/kg), across the evaporator, $a_1, a_2, a_3, a_4, a_5, a_6,$ are the regression coefficients, T_{ewb} is the entering wet-bulb temperature obtained from space air temperature/relative humidity measurements at smart thermostats, T_{odb} is the outdoor dry-bulb temperature obtained from an online weather station, and ρ_{sa} and Q_{sa} are the density and flow rate of the supply air, respectively.

Two phases are considered for the proposed AFDD algorithm, namely: Phase 1, which involves an installation fault check, and Phase 2, which involves a degradation fault check. For Phase 1, the model is trained using manufacturer data, such as rated data collected from the expanded performance table (EPT) of the HVAC system provided by the manufacturer. The trained model is then used to predict the enthalpy change across the evaporator in daily operation. In the prediction, the indoor wet-bulb temperature is obtained from space air temperature/relative humidity measurements at smart thermostats, while the outdoor dry-bulb temperature and relative humidity are obtained from an online weather station. For Phase 2, since the Q_{sa} data, which is required in Equation (6.2), is not available due to cost implications, the equation is slightly modified to give

$$\Delta\hat{\imath} = b_1 + b_2 T_{ewb} + b_3 T_{ewb}^2 + b_4 T_{odb} + b_5 T_{odb}^2 + b_6 T_{ewb} T_{odb} \tag{6.3}$$

and trained using data collected during Phase 1.

Next, the measured enthalpy change, Δi, across the evaporator, with which the performance of the evaporator performance is then monitored in real time, is obtained by taking the difference between the return air enthalpy and supply air enthalpy, i.e.,

$$\Delta i = i_{ra} - i_{sa} \tag{6.4}$$

Equations (6.2)–(6.4) are for wet cooling. For dry cooling, T_{ewb} in Equations (6.2) and (6.3) is replaced with T_{edb} (i.e., the entering dry-bulb temperature) while the measured enthalpy change is calculated from the following equation:

$$\Delta i = c_{p_{sa}} (T_{ra} - T_{sa}) \tag{6.5}$$

which is the sensible enthalpy change across the coil, where T_{sa} is the supply air temperature (SAT), and $c_{p_{sa}}$ is the specific capacity of the supply air, assumed to be constant.

The smart thermostat provides space air temperature and humidity measurements, which can be used to obtain the enthalpy of the air entering the evaporator, simply called the return air (RA) enthalpy, i_{ra}. In addition to the smart thermostat, a smart node sensor that is linked to the thermostat is strategically mounted on a diffuser to aid easy access for SAT measurements. The SAT is measured at the diffuser rather than at the exit of the evaporator due to the technical difficulty of installing a sensor right after the evaporator. However, the node sensor used to measure the SAT can only measure temperature. To obtain the SA enthalpy, i_{sa}, a second property is needed. This second property can be estimated from the consideration that for wet cooling, the SAT after the coil is almost saturated, thus suggesting a relative humidity, ϕ_{sa}, of about 90%–95%. Then, for dry cooling, the SA dew point (SADP) would be the same as the RA dew point, and RA dew point can be computed from T_{ra} and ϕ_{ra}. But, as the SAT is measured at the diffuser, ϕ_{sa} can significantly vary between the coil exit and the diffuser, thus making it more challenging to estimate ϕ_{sa} at the diffuser. So, instead of ϕ_{sa}, an estimated SADP after the coil is used because the SADP at the diffuser is the same as that at the evaporator exit. To estimate SADP after the coil, the following assumptions are required. The validation of these assumptions is presented in Section 6.5. Note that there is only a slight difference between the SAT at the diffuser and the SAT after the coil. This is ensured by using the diffuser closest to the indoor unit.

For residential units, the SADP is assumed to be around 55$\pm$2°F. This assumption is adopted based on the consideration that to maintain the indoor humidity in commercial HVAC systems, most HVAC manufacturers target an SADP of 55≥ 55°F. This can be extended to residential HVAC systems, especially for SAT ≥ 55°F. For SAT < 55°$\phi_{sa} \approx 100\%$. Thus, $SADP \approx SAT$. It follows that the SADP is estimated using:

$$SADP = \begin{cases} 55\,F, & if\ SAT \geq 55\,F \\ SAT, & if\ SAT < 55\,F \end{cases} \tag{6.6}$$

with a ± 2°F uncertainty. Other assumptions made in the proposed AFDD algorithm are:

First, the thermostat data accurately represents the RA condition. This relies on the assumption that the thermostat is well located in the space, close enough to the return air ducts. Second, the EPT data used for training the model in Phase 1 is accurate and representative of the expected behavior of the AC based on specifications. Third, infiltration is the main source of outdoor air intake in residential buildings, and this contributes significantly to the return air humidity. This assumption is based on the need to classify the evaporator coil condition as either wet or dry. The outdoor air dew point is thus used to make this classification.

Next, to forestall false fault alarms due to some disturbances in the system, the probability of a fault based on $\delta_{\Delta i}$ in Equation (6.1) exceeding $\overline{\delta_{\Delta \iota}}$ is subject to further scrutiny using the CUSUM (cumulative sum) control strategy used in related work (Grigg, Farewell, and Spiegelhalter 2003; Jain et al. 2019). The method involves taking cumulative sums, τ, of times when $\delta_{\Delta i} > \overline{\delta_{\Delta \iota}}$. If τ continues to increase and eventually reaches a set duration threshold, $\bar{\tau}$, then and only then will a fault alert be sent. This τ is a sum of consecutive times and is intended to avoid having a cumulative sum from disturbances over an extended period of time eventually lead to a false fault alarm. Thus, if for a given time step, $\delta_{\Delta i} > \overline{\delta_{\Delta \iota}}$, then τ increases by a time step. But if for the next time step, $\delta_{\Delta i} < \overline{\delta_{\Delta \iota}}$, then τ decreases by a time step. The flow chart in Figure 6.3 illustrates how the AFDD process works, in which the grayed flow path is inactive for this method. So, whenever $\delta_{\Delta i} < \overline{\delta_{\Delta \iota}}$, the system is considered to be fault-free. But, even if $\delta_{\Delta i} > \overline{\delta_{\Delta \iota}}$, but $\tau < \bar{\tau}$, the system is still considered to be fault-free, as it is assumed that such deviations could possibly be due to some exogenous uncertainties/disturbances, rather than a true fault.

6.3.3 Proposed FDD Method 2

Figure 6.3 illustrates the procedure for this method. The method is similar to Method 1, but with the incorporation of power measurements from a smart power meter. The power measurements are primarily used to diagnose RCL faults from AFR faults. Hence, two AFDD indices are employed for this method, i.e., enthalpy change and indoor fan power. The algorithm for enthalpy change is the same as that for Method 1. However, for indoor fan power, a baseline is first established using the manufacturer's specified indoor fan/blower power for the case of installation check (Phase I). An average of the indoor fan power consumption from data collected during Phase 1 is then obtained for Phase 2. The average indoor power is used because for HVACs with constant-speed fans (found in most residential units and within the proposed scope for this project), the fan power consumption remains constant and depends almost entirely on the indoor airflow rate through the system. However, the indoor fan power measurement experiences oscillations caused by signal noises and/or motor and fan dynamics. To address any possible uncertainty introduced into the AFDD algorithm due to such disturbances, a suitable threshold has been chosen and used in the algorithm. Details of this are provided in Section 6.4. Hence, the averaged indoor power from the normal operation dataset is believed to be appropriate for constant-speed HVAC systems. For the systems installed with ECM motors, the proposed method would be ineffective for AFDD; as a result, a machine learning technique can be employed to help predict the indoor fan power consumption. This study focuses on constant-speed, belt-driven fans, as they are one of the most popular types of AC units used in U.S. homes. Therefore, the measured indoor power (p) is compared with the predicted fan power ($\hat{p}$) and the deviation is calculated using a percentage as follows:

$$\delta_{IP} = \frac{\hat{p} - p}{\hat{p}} \times 100 \tag{6.7}$$

With a baseline of the expected indoor power consumption obtained from the product manual or Phase 1 data, subsequent power measurements are compared to this baseline and deviations above a set fault threshold are identified as faulty behavior, specifically related to incorrect indoor airflow. In the case where such indoor power deviations are below the set threshold, the enthalpy change across the evaporator is further used to check for the presence of incorrect charge fault following the process described for Method 1. Again, the CUSUM algorithm is used to forestall possible false alarms by monitoring for faulty behavior and triggering a fault alarm whenever the set duration threshold is exceeded.

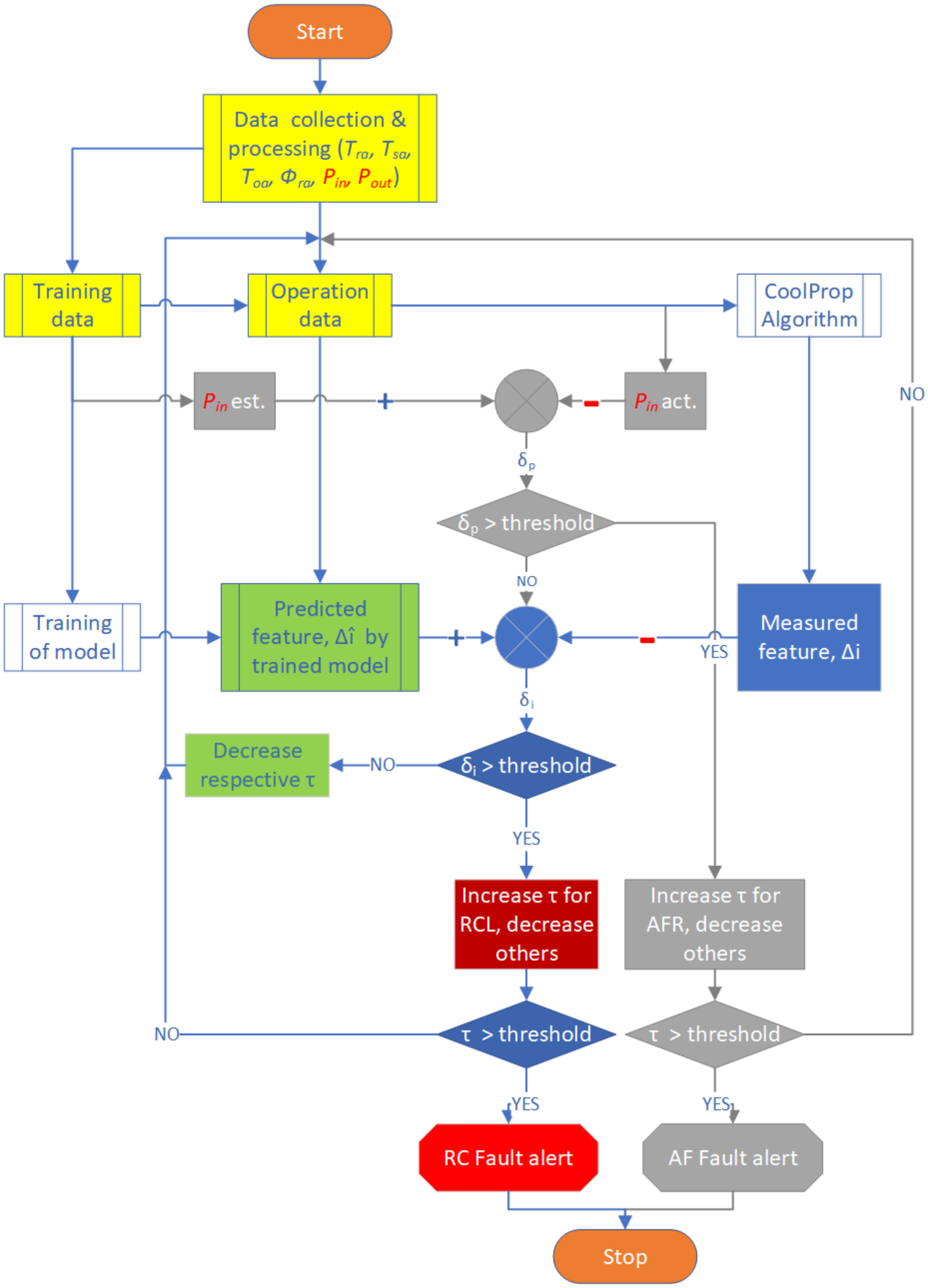


Figure 6.3. AFDD algorithm for IoT-based method

6.4 Determination of Fault Threshold Through Uncertainty Propagation Analysis

The selected feature, enthalpy change, cannot be measured directly and thus must be calculated from independent variables such as temperature and humidity or dew point, etc., using suitable packages such as CoolProp (Bell et al. 2013), an open-source database of fluid and humid air properties. However, because these independent variables are measured using sensors with built-in uncertainties, an uncertainty propagation analysis needs to be conducted to better understand the uncertainty in the enthalpy difference. This will help ensure that deviations caused by such uncertainty are not misjudged as a fault. The uncertainty propagation analysis is carried out based on the root sum square (RSS) method:

$$\delta Y = \sqrt{\left(\frac{\partial Y}{\partial X_1}\delta X_1\right)^2 + \left(\frac{\partial Y}{\partial X_2}\delta X_2\right)^2 + \cdots + \left(\frac{\partial Y}{\partial X_n}\delta X_n\right)^2} \qquad (6.8)$$

where δY is the uncertainty in the dependent variable $Y \text{ and } \delta X_n$ is the uncertainty in the independent variable X_n.

For wet coil operation, where $\Delta i = i_{ra} - i_{sa}$ according to Equation (6.4), $i_{ra} = f(T_{ra}, \phi_{ra})$ and $i_{sa} = f(T_{sa}, T_{dp})$. Thus, $\Delta i = f\left(T_{ra}, \phi_{ra}, T_{sa}, T_{dp}\right)$, which can be computed from the physics-based model

$$\Delta i = 0.24(T_{ra} - T_{sa}) + \frac{(k_1 + k_2 T_{ra})\phi_{ra}}{\dfrac{k_3 T_{raK}^{8.2}}{\exp\left(k_4 + k_5 T_{raK} - \dfrac{k_6}{T_{raK}}\right)} - \phi} - \frac{k_1 + k_2 T_{sa}}{\dfrac{k_3 T_{dpK}^{8.2}}{\exp\left(k_4 + k_5 T_{dpK} - \dfrac{k_6}{T_{dpK}}\right)} - \phi} \qquad (6.9)$$

where $k_1 - k_6$ are constants, while T_{raK} and T_{dpK} are the RA and dew-point temperatures in degrees Kelvin. The uncertainty for Δi in wet coil operation, based on Equation (6.8), then becomes:

$$\delta(\Delta i) = \sqrt{\left(\frac{\partial(\Delta i)}{\partial T_{ra}}\delta T_{ra}\right)^2 + \left(\frac{\partial(\Delta i)}{\partial \phi_{ra}}\delta \phi_{ra}\right)^2 + \left(\frac{\partial(\Delta i)}{\partial T_{sa}}\delta T_{sa}\right)^2 + \left(\frac{\partial(\Delta i)}{\partial T_{dp}}\delta T_{dp}\right)^2} \qquad (6.10)$$

where the uncertainties for $T_{ra}, \phi_{ra},$ and T_{sa} are provided by the manufacturer, while that for T_{dp} is assumed. Their values are shown in Table 6.1.

For dry coil operation, $\Delta i = c_{p_{sa}}(T_{ra} - T_{sa})$ according to Equation (6.5). Thus, $\Delta i = f(T_{ra}, T_{sa})$ since $c_{p_{sa}} = 0.244\ Btu/lbma$ is a constant. The uncertainty for Δi for dry coil operation, based on Equation (6.8), can therefore be written as:

$$\delta(\Delta i) = \sqrt{\left(\frac{\partial(\Delta i)}{\partial T_{ra}}\delta T_{ra}\right)^2 + \left(\frac{\partial(\Delta i)}{\partial T_{sa}}\delta T_{sa}\right)^2} \qquad (6.11)$$

The results for the uncertainties computed for both wet coil and dry coil operation are shown in Table 6.1.

Table 6.16. Uncertainties in Sensor Measurements

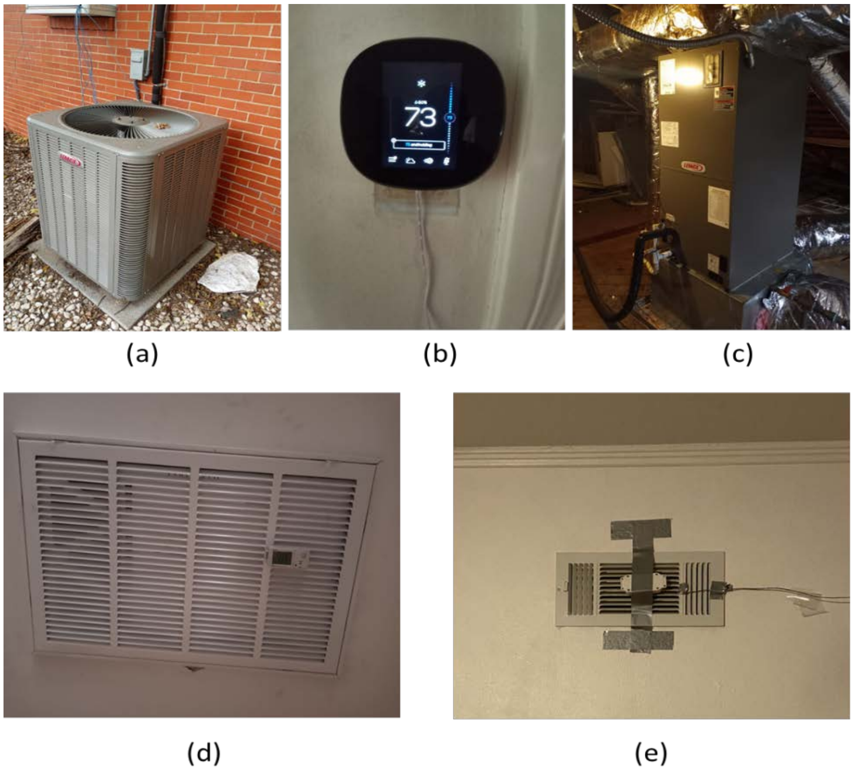
Variables	**Inputs**				**Outputs**	
					Wet	**Dry**
	$\boldsymbol{T_{ra}}$, $\boldsymbol{°F(°C)}$	$\boldsymbol{\phi_{ra}}$ (%)	$\boldsymbol{T_{sa}}$, $\boldsymbol{°F(°C)}$	$\boldsymbol{T_{dp}}$, $\boldsymbol{°F(°C)}$	$\boldsymbol{\Delta i}$ ($\boldsymbol{Btu/lbm}$)	$\boldsymbol{\Delta i}$ ($\boldsymbol{Btu/lbm}$)
Uncertainty	$\pm 1.0(0.6)$	± 3	$\pm 1.0(0.6)$	$\pm 2.0(1.2)$	± 1.2	± 0.34

Next, to ensure that there is no false fault alarm resulting from uncertainties, a ±1.2 Btu/lbm deviation threshold, $\overline{\delta_{\Delta i}}$, is used for wet coil operation and a ±0.34 Btu/lbm deviation is used for dry coil operation. Also, a duration threshold, $\bar{\tau}$, of 24 hours is chosen. The choice of 24 hours as the duration threshold is made based on the consideration that, presumably, if due to severe fault conditions, the AC runs for an entire day during a hot summer, such a fault can be detected by the end of the day. If a homeowner’s preference and the weather conditions are such that the AC does not run throughout the day, early fault detection is still possible. For instance, if AC runs for only 4 hours per day on average, a fault can be detected in 6–7 days’ time.

For the fan power measurement, the smart meter used has an accuracy of about $\pm 2\%$ (Emporia 2021). Factoring in the uncertainties discussed in Section 6.3.3, a threshold of 10% is thus used in Phase 1, while a 5% threshold is used in Phase 2 for the AFDD with this method. A bigger threshold is used for Phase 1 because of the notable uncertainties that are present in the measured data, which are most likely due to a different installation configuration in an actual home, rather than the one generating the EPT data in a manufacturer’s test bed. However, for Phase 2, since operation data from Phase 1 is used as a baseline, uncertainties generated by a different installation configuration would be consistent in both the training and testing datasets, so a smaller threshold is acceptable. More so, a larger threshold would lead to loss of AFDD potential to detect less severe degradation faults.

6.5 Validation of the Proposed FDD Algorithm in the Norman Test House

The Norman Test House has a 3.5-ton heat pump. For proof-of-concept purposes, a set of portable data loggers were installed and used to collect return and supply air temperature and humidity inside the test AC unit, as shown in Figure 6.4.


(a) (b) (c)

(d) (e)

Figure 6.4. Norman Test House and data acquisition devices: (a) outdoor unit, (b) smart thermostat, (c) indoor unit, (d) RA hobo logger, and (e) SA hobo logger.

First, the model described in Equation (6.2) was trained using EPT data from the AC manufacturer. The EPT data for the Norman Test House is shown in Table 6.2. The model parameters are obtained using the least-squares method with a maximum absolute error (MAE) of 0.002 and a mean-square error (MSE) of 6.318e-06, shown in Table 6.3.

Table 6.17. Data From EPT of Test System

Model N0	**14HPX-042-230-17 – CBX26UH-042**						
Entering wet bulb temperature	Total air volume	Outdoor air temperature entering outdoor coil					
		65 °F	75 °F	85 °F	95 °F	105 °F	115 °F
	cfm	Total cooling capacity kBtuh					
59 °F	1450	47	45	43	41	39	36.6
63 °F	1450	47	45	43	41	39	36.6
67 °F	1450	50	48	46	43.5	41	38.5
71 °F	1450	52.5	50.5	48.5	46	43.5	41

Table 6.18. Parameters of Cooling Capacity Model

Parameter	a_1	a_2	a_3	a_4	a_5	a_6
Value	4.0015350	-0.0972891	0.0008632	0.0011720	-0.0000128	-0.0000568

6.5.1 Validation of the Trained Model Using Normal Operation Data

Data under normal operating conditions were collected from October 7–12, 2021, for validating the model. For the training, validation, and testing of the AFDD algorithm, only times when the AC was on were considered. Also, because steady-state operation is important in a general FDD process (Rogers, Guo, and Rasmussen 2019), the first 4 to 5 minutes and the last 2 minutes of each on-cycle were filtered out. Only about 5 minutes of AC start data for each cycle were removed to ensure that sufficient data were collected for this proof-of-concept phase. However, for implementation and real-life application, it is recommended to remove data from the first 10 minutes, which is the typical time it takes most residential units to get to steady state.

First, the SA enthalpy calculation method was validated. For verification purposes, the actual SA dew point was measured using the data logger in the test. The estimated SA dew point was calculated using Equation (6.6). A comparison of the measured SA dew point with the estimated SA dew point is shown in Figure 6.5 along with the measured RA dew point temperature. The data covers the period of normal operation as well as the period when a fault is introduced. The testing data show a typical wet coil operation, whereby the measured RA dew point was always higher than the measured SA dew point, and the estimated SA dew point matched the measured SA dew point within the ±2˚F error. This result thus validates the proposed method for obtaining the SA enthalpy from just temperature measurement.

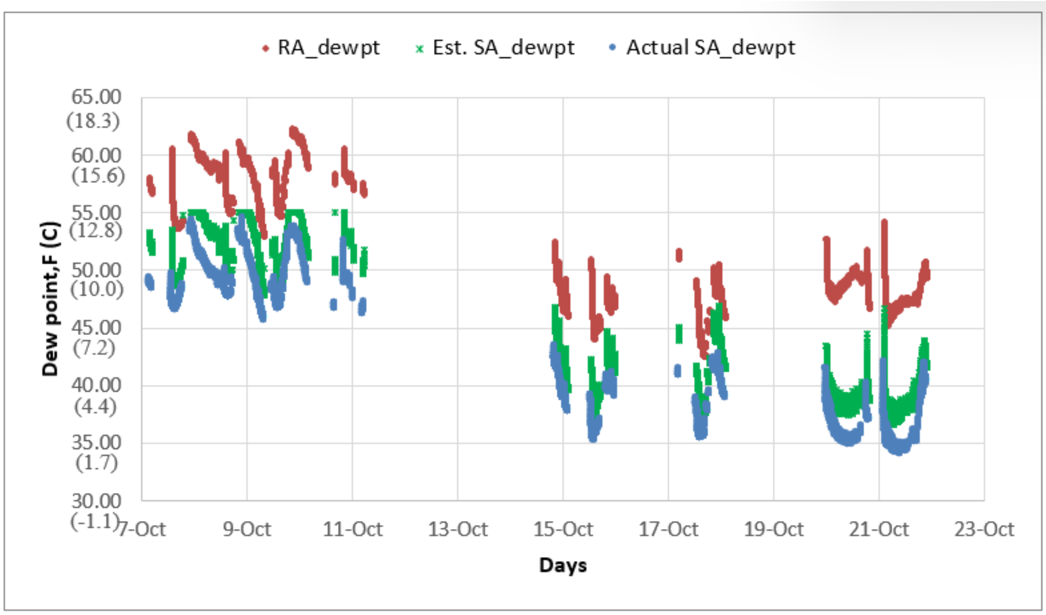


Figure 6.5. Comparison of measured and estimated SA dew points

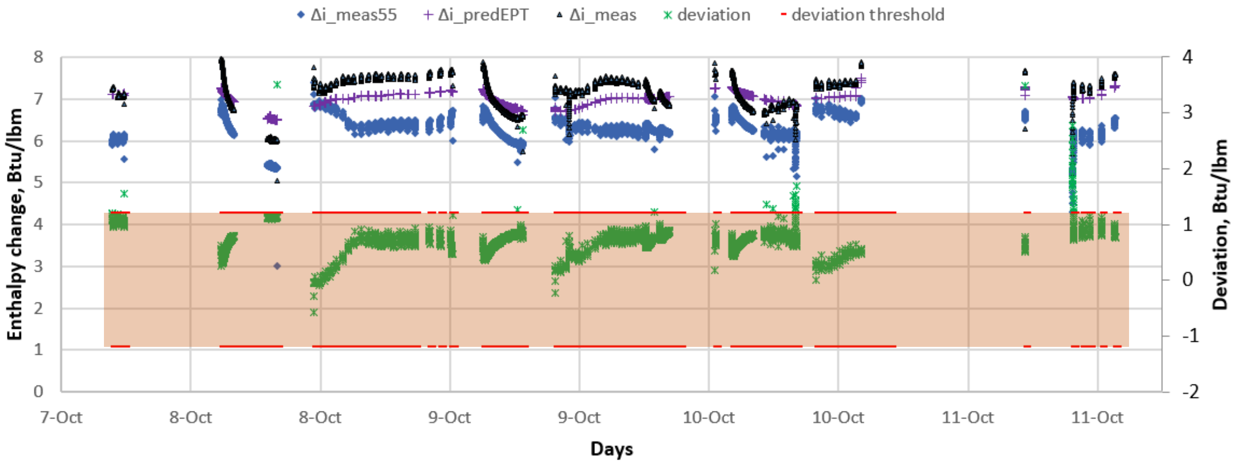


Figure 6.6. Validation of model trained with rated data

Figure 6.6 shows the validation results. The validation was done mainly with just the return air temperature, return air relative humidity, and SAT data collected with the hobo loggers. Equation (6.6) is used to calculate the SADP, which is then combined with the SAT to obtain the SA enthalpy for the calculation of Δi_{meas55}. The quantity $\Delta i_{predEPT}$ is the enthalpy change predicted using the model trained with EPT data. The deviation $\delta_{\Delta i}$

is the difference between $\Delta i_{predEPT}$ and Δi_{mea55} (Equation (6.1)). From the figure, one can see that for the most part there is a close match between the actual enthalpy difference and the predicted enthalpy difference, and the deviation stays within the threshold band except for a few outliers. Using the CUSUM control strategy, the proposed AFDD tool can safely infer that such outliers are not necessarily because of a fault but most likely due to disturbances/noises in the system. Meanwhile, the reason for the slightly significant disparity (which is, however, still below the threshold for the most part) is due mainly to the estimates used, especially for SADP. A comparison of Δi_{mea55} with Δi_{mea}, which is computed using actual SADP from the hobo logger shows a closer match with $\Delta i_{predEPT}$. A deviation threshold, $\overline{\delta_{\Delta i}}$ = $\pm$1.2 Btu/lbm (2.79 kJ/kg), is chosen based on the uncertainty propagation analysis. Also, to check for model uncertainty, the three-sigma rule is applied. However, to eliminate other factors such as the SADP estimate contributing to the calculated model uncertainty, Δi_{mea} and $\Delta i_{predEPT}$ are used. With the results in Figure 6.6, for 68% of the deviation, $\delta_{\Delta i}$ is between 0.07 Btu/lbm (0.2 kJ/kg) and 0.32 Btu/lbm (0.7 kJ/kg); for 95%, $\delta_{\Delta i}$ < 0.45 Btu/lbm (1.0 kJ/kg); and for 99%, $\delta_{\Delta i}$ < 0.58 Btu/lbm (1.3 kJ/kg). So, for a 95% confidence level, a model uncertainty of 0.45 Btu/lbm is obtained, which is well below the value of $\pm$1.2 Btu/lbm from the uncertainty propagation analysis. Therefore, $\pm$1.2 Btu/lbm is a safe threshold to ensure that the algorithm produces no false fault alarms.

6.5.2 Validation of the Effectiveness of Fault Detection Using Operational Data With an Introduced Fault

To test the effectiveness of the algorithm in fault detection, an AFR fault is considered as an initial case study. The AFR fault is considered because it is one of the dominant faults in residential HVAC systems. Also, it is relatively easy to create both technically and cost-wise. The AFR is introduced purposely by partially blocking the return air duct. Again, for proof-of-concept purposes, and from flow rate reduction simulated in previous studies, a 30% drop in the airflow is chosen for the tests. The tests were conducted from October 20–22, 2021. The results shown in Figure 6.7 indicate a clear disparity between the actual enthalpy difference and the predicted no-fault enthalpy difference, leading to significant deviations, which mostly fall outside the threshold band (brown shaded area). Furthermore, in line with the physics of heat transfer, where reducing airflow would increase Δi, the measured enthalpy difference is seen to be larger than the predicted enthalpy difference, i.e., an enthalpy change rises (so that the deviations are negative). As will be seen later for low charge fault, the reverse is the case. This way, two common faults in residential HVAC systems can be successfully detected and diagnosed. As this is primarily for proof of concept, the AC only ran for about 8 hours during the 2 days. So, the CUSUM algorithm could not be implemented for this test considering that the duration threshold is 24 hours. However, the large deviations demonstrate the potential of the algorithm to detect faults. In the field tests, the CUSUM algorithm was implemented. Meanwhile, Figure 6.8 reveals about a 5%–15% increase in energy consumption (kWh), with an average increase of about 9% due to the 30% low indoor airflow fault, thus showing the energy and cost savings potential of the proposed AFDD algorithm once the 30% AFR fault is detected and corrected.

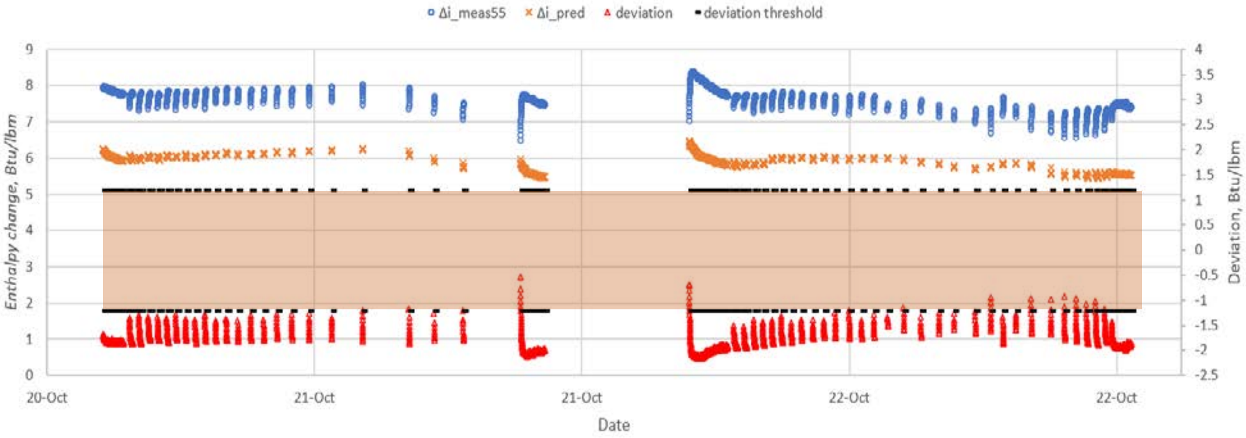


Figure 6.7. Detection of low indoor airflow fault

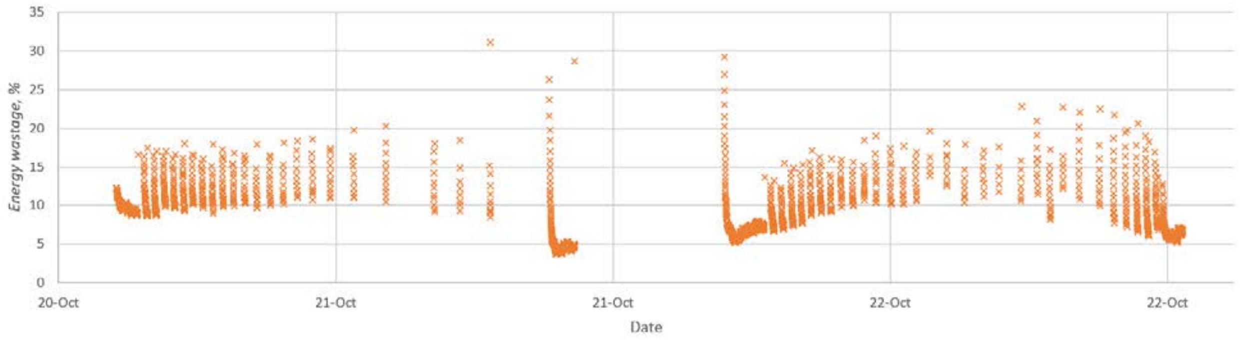


Figure 6.8. Energy consumption increase due to low indoor airflow fault

Figure 6.9 shows the results for the same AFR fault check conducted using indoor power consumption. The results show a 12%–16% drop in the indoor power compared with the predicted indoor power using measured data (blue markers in Figure 6.9). As seen from the plot, the deviation stays clearly outside the threshold band for this case. This suggests that the indoor power is more sensitive than Δi for the detection of low indoor airflow faults. This is also because the indoor power requires none of the assumptions introduced for the Δi method, except that it is only limited to detection of incorrect airflow faults and by a higher cost than with Δi. Also, as noted in Section 6.3.3, other factors that could potentially affect the indoor power consumption turn out to have minimal impact, hence the slight variation observed in the measured indoor power consumption, IP.

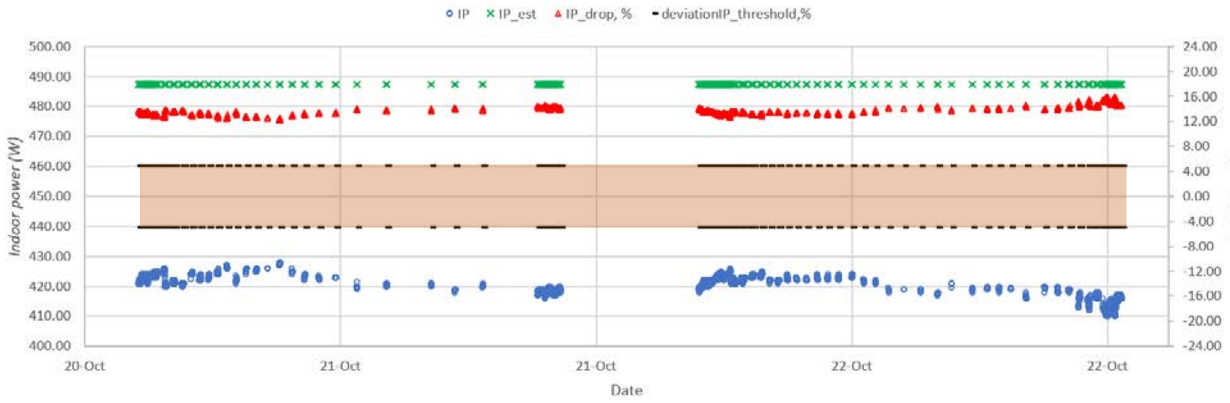


Figure 6.9. Detection of low indoor airflow using indoor fan power

6.6 Validation of the Effectiveness of Fault Detection Using the Miami House

For further validation of the AFDD algorithm, two other test homes were used, one of which is in a wet climate (humid) region while the other is in a dry climate region. Such diversity is considered to further investigate the robustness of the proposed AFDD algorithm and to discover possible limitations in its applications. This section is focused on the tests conducted in the Miami house experiencing a humid climate. A one- to three-year-old 3-ton AC unit, installed in a single-family home in Miami, Florida, was used in this test. Again, for proof-of-concept purposes, a set of portable data loggers was installed to collect return and supply air temperatures and humidity inside the test AC unit. The weather data are obtained from a local weather station. The design parameters of the test AC unit in Miami are shown in Table 6.4.

Table 6.19. EPT Data for Miami Test House

Model N0		4TWA3036A3/4 + 4TEC3F36A1					
Entering wet bulb temperature	Total air volume	Outdoor air temperature entering outdoor coil					
		65 °F	75 °F	85 °F	95 °F	105 °F	115 °F
		Total cooling capacity					
	cfm	kBtuh					
59 °F	1200	37.5	35	32.5	30.8	29.9	27.3
63 °F	1200	38.9	36.4	33.9	32.1	31.2	28.5
67 °F	1200	41.6	39.1	36.6	34.6	33.6	30.7
71 °F	1200	44.5	42	39.5	37.4	36.3	33.1

Applying Method 1 to the Miami house, the model in Equation (6.2) was first trained with the EPT data summarized in Table 6.4. The trained model was then used to predict the enthalpy change and compare it with the actual enthalpy change during normal operation in a bid to validate the model. For this house, the AC was only run for a short period of time per day. Hence, to aid visualization of the results, the plots are given in hours rather than days. However, the data for this house were collected from September 1–27, 2021. Two AFR faults were purposely created, one with a 12% flow rate reduction and the other with a 21% flow rate reduction. To alleviate the impact on the test unit, the AFR faults were created by adjusting the fan speed setting instead of blocking filter area since a multispeed fan is installed in the test unit. Due to the limitations of the system, we could not reduce the flow rate beyond 21%. Among the testing days, the 21% flow rate reduction fault was generated from September 1–September 5, 2021, followed by normal flow rate operation from September 5–September 12, followed by the 12% flow rate reduction fault from September 15–September 27, 2021.

Figure 6.10 shows the model validation results with the EPT data for Phase 1. The results compare the enthalpy change between the one predicted using the EPT data and the one measured under normal operation. The results reveal that the difference between the predicted and actual enthalpy change across the evaporator is below 0.5 Btu/lbm and stays mostly within the threshold band. There are, however, a few exceptions where it goes below the lower bound, mainly because of the impact of transient data. The presence of transient data for this dataset is likely due to the fact that only the first 2 minutes of the start of each AC run cycle are filtered out, unlike the 4 minutes adopted for the Norman test house. The first 2 minutes of data were disregarded due to the short AC run times for this house.

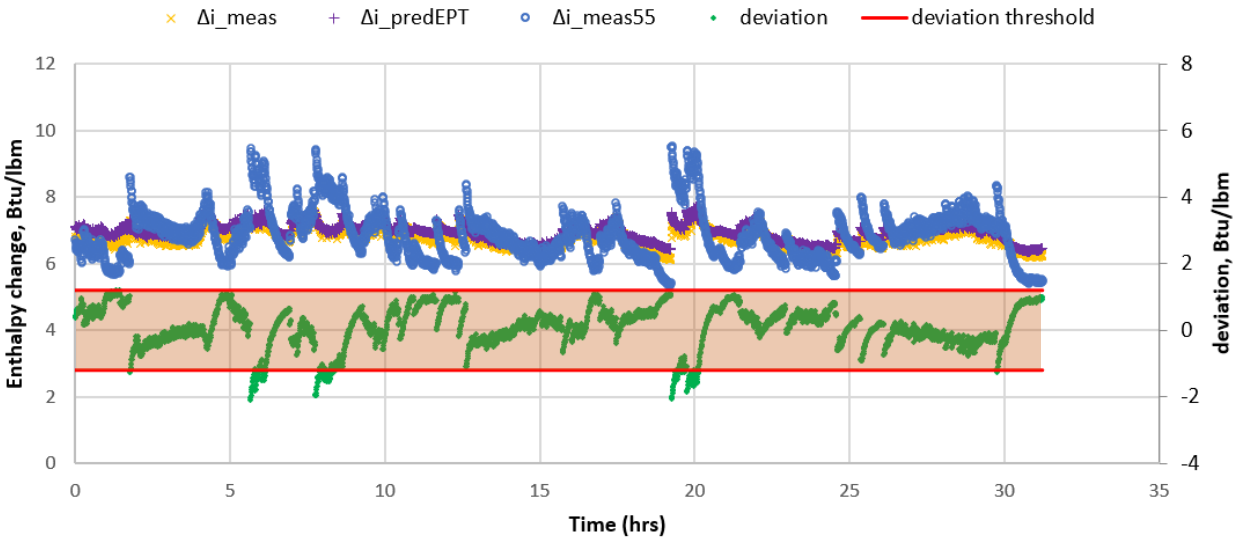


Figure 6.10. Validation of model trained with rated data for the Miami house

Next, for Phase 2, the model in Equation (6.2) was retrained with the normal operation data from September 5–September 12, 2021. For this Phase 2, to quantify the impact of the assumptions we made on SADP, we verify the algorithm through two steps, one using direct SADP measurements and the other using estimated SADP obtained based on the engineering assumptions.

Step One – Using measured SADP data to implement the AFDD analysis

The retrained model was used to predict the enthalpy change for the different cases (i.e., normal, 12% airflow reduction and 21% airflow reduction), which was then compared with the measured enthalpy change. Because the low flow rate fault generates a larger enthalpy change than normal operation, the expected deviation defined by Equation (6.1) should be below the lower bound of the threshold band. The results in Figure 6.11 reveal that the deviations show a distinctive pattern with the two faults (shaded area) and with normal operation. With the 21% flow rate reduction fault, most of the deviations are below the threshold lower bound (-1.2 Btu/lbm), i.e., 1,111 sample points among 1,363 points, equivalent to 81.5%. With the 12% flow rate reduction fault, although it is obvious that the deviations are between -0.4 Btu/lbm and -0.8 btu/lbm, much higher than normal operation, none of the deviations exceed the threshold -1.2 Btu/lbm. Therefore, the 12% flow rate reduction fault cannot be detected by the threshold of ±1.2 Btu/lbm. However, this result is based on using the actual SADP measurements without needing the assumptions. To enhance the sensitivity of the AFDD method, if the SADP can be directly measured, it is reasonable to reduce the threshold so that the AFDD algorithm can detect faults with less severity.

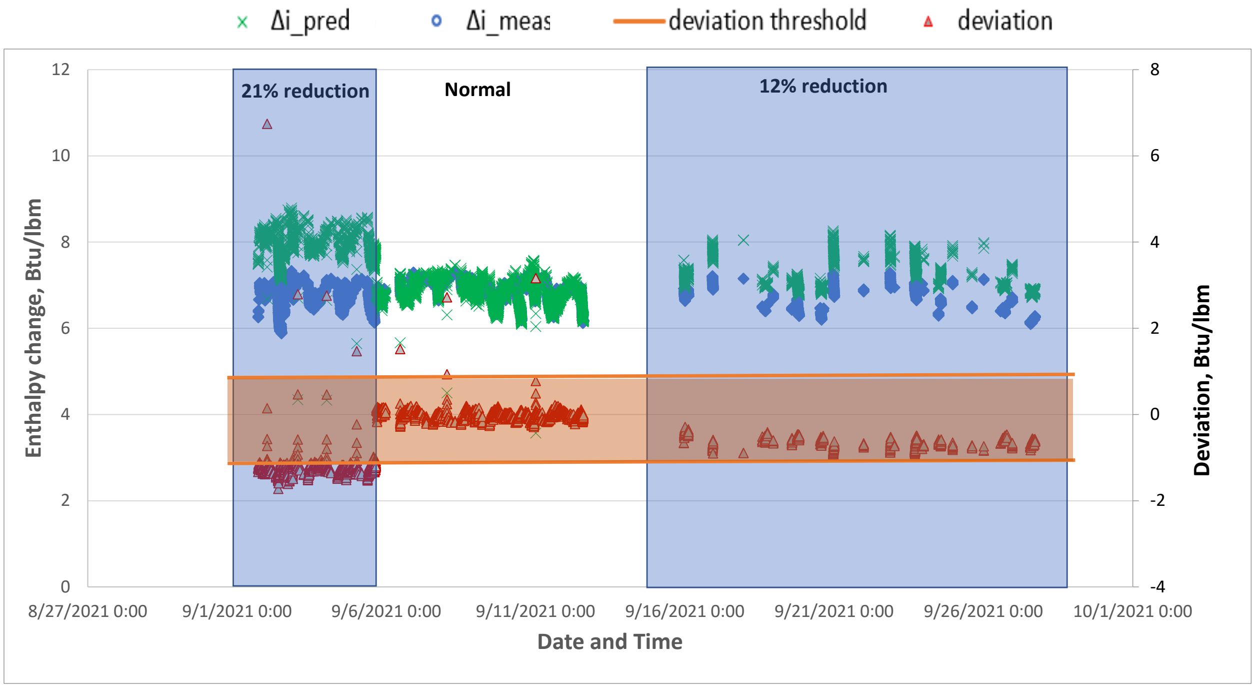


Figure 6.11. Test for the Miami house based on Method I and using actual SADP measurements

Step Two – Using estimated SADP data to implement the FDD analysis

Although the AFDD algorithm is shown to be effective when actual measured SADP is used, in field implementation the SADP is not available if only a smart thermostat and a node sensor are used. Therefore, the assumptions stated in Section 6.3.2 are adopted to estimate the SADP for enthalpy calculation of the supply air temperature. The enthalpy change model was retrained again using estimated SADP and then used to predict the enthalpy change. Compared with the results in Figure 6.11, the results in Figure 6.12 where the estimated SADP is used reveal that the deviations show more oscillations, although the different patterns for faulty and normal operations can still be observed. To be more specific, with the 21% flow rate reduction fault, there are 709 sample points exceeding the threshold of -1.2 Btu/lbm (52.0% over 1,363 total sample points). Compared with the results using actual measured SADP at 81.5%, the estimated SADP using the assumptions made the AFDD algorithm less sensitive to the fault. Consequently, it will take longer to detect the fault with the CUSUM algorithm, if the fault can be detected at all.

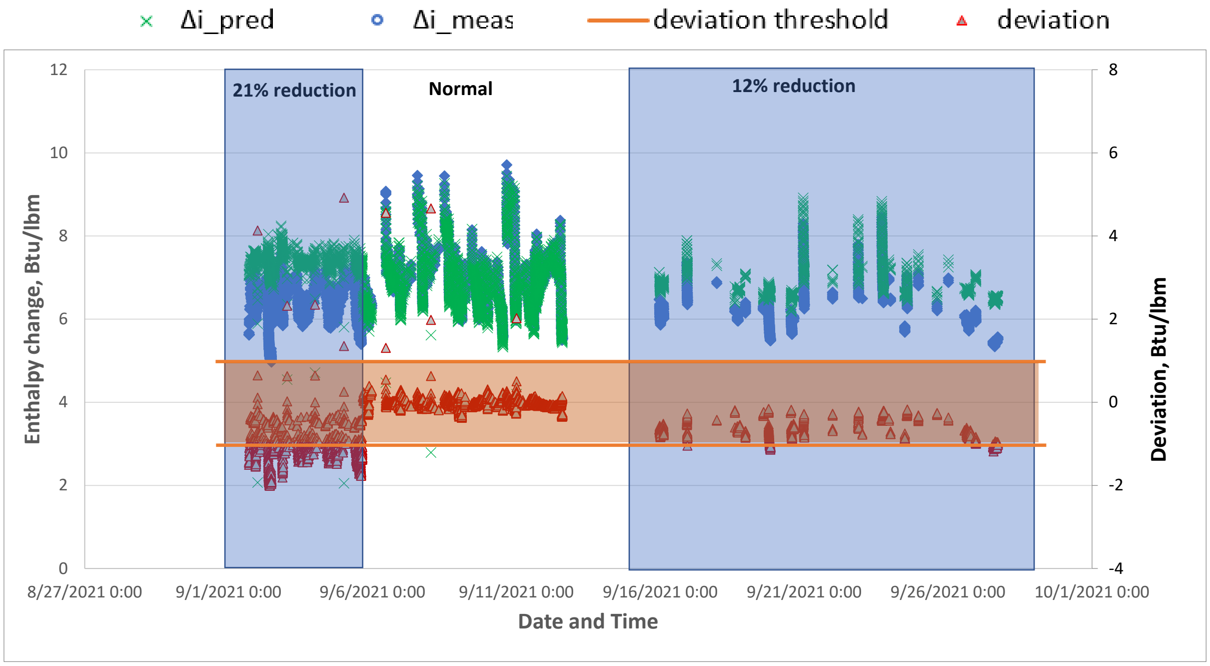


Figure 6.12. Test for the Miami House based on Method 1 and using estimated SADP

Applying Method 2 to the Miami House, Figure 6.13 shows a 16%–23% drop in the indoor power, with an average of 20% indoor fan power reduction introduced by the 12% airflow reduction fault. This shows the sensitivity of the indoor power usage to this specific fault. Again, due to the short run time of the AC during the fault tests, the CUSUM mechanism is not demonstrated. For the 21% indoor airflow reduction, a similar observation is made, with the drop in indoor power consumption ranging from 35% to 42%, at an average of 39%, shown in Figure 6.14. The fan power deviations in the Miami house provided much better sensitivity compared with the fan power test results in the Norman house. This is because of the different ways to create the flow rate faults: blocking filter area in the Norman house and adjusting fan speed in the Miami house. In the Miami house, higher fan power reduction (close to cubic relationship to the flow rate) is observed, which complies with the fan laws.

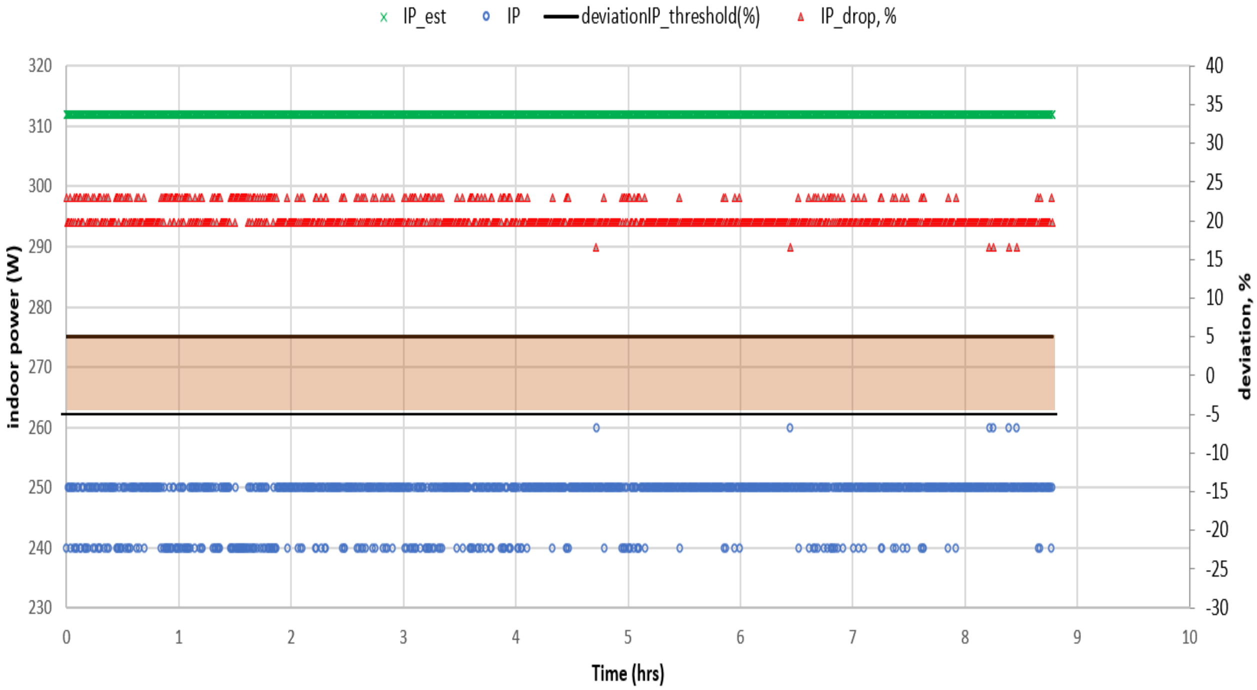


Figure 6.13. 12% indoor airflow reduction detected using Method 2 for the Miami house

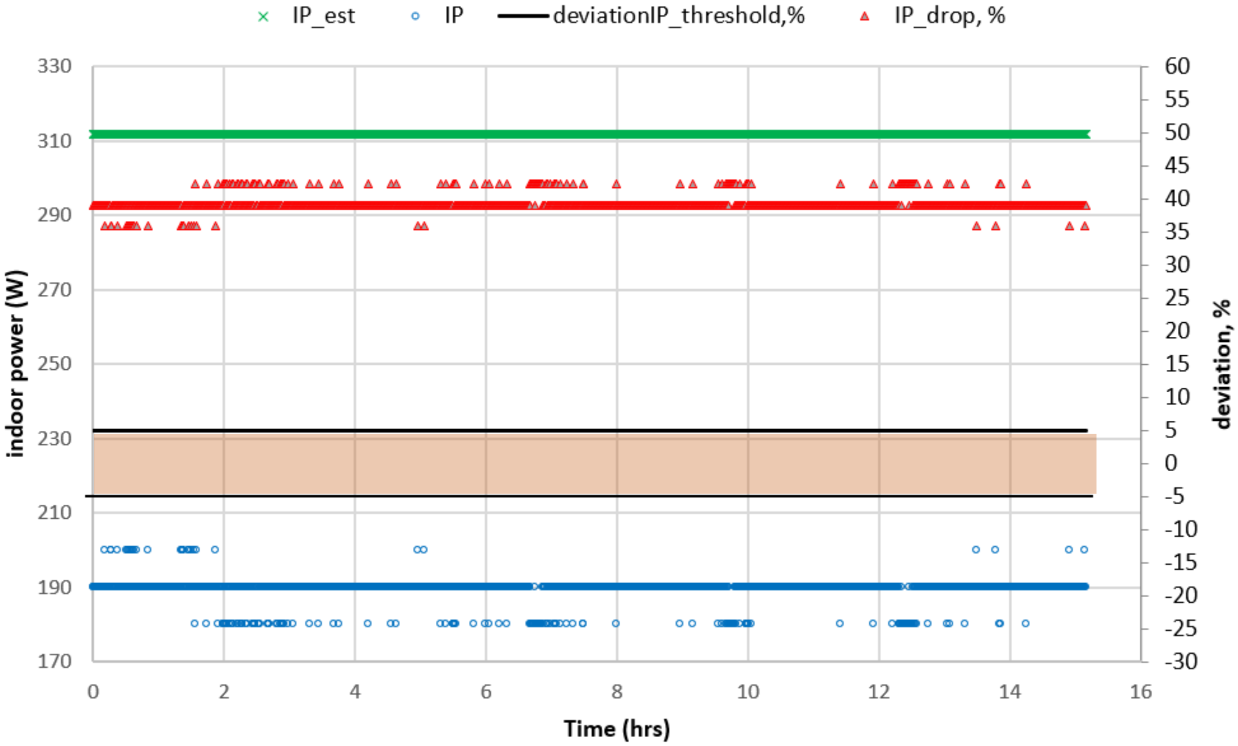


Figure 6.14. 21% indoor airflow reduction detected with Method 2 for the Miami house

6.7 Validation of the Effectiveness of Fault Detection Using PNNL House B

For a dry climate, a 2.5-ton AC unit installed in PNNL House B is used to test the AFDD algorithm. Phase 1 in this test is not included because the EPT data are collected when the AC unit operates under wet conditions. Therefore, Phase 1 is not applicable to PNNL House B. For Phase 2, both low flow rate fault and low charge fault were generated individually and simultaneously to test the AFDD algorithm. The test period was from September 4 to October 12, 2021. It started with creating a low airflow rate fault by blocking one filter area from September 4 to September 16, 2021. Because the system does not have a flow station, flow rate measurements were conducted three times using a handheld device during the test to ensure a proper amount of flow reduction. Different flow rates were obtained, at 33.7% reduction measured on September 4, 2021; at 29.4% reduction measured on September 7, 2021; and at 23.2% reduction measured on September 7, 2021. The different measurements might be caused by measurement errors and flow dynamic changes generated by other possible interruptions at PNNL because multiple teams were using the lab. From September 16 to September 18, 2021, a 30% flow rate reduction and a 30% undercharge of freon were both created to collect data under two simultaneous faults. However, the severity of the two faults caused the system to freeze, and those data were discarded. From

September 29 to October 5, 2021, after the unit recovered from being frozen, a single 30% undercharge fault was created for data collection. The test system was eventually set back to normal from October 5 to October 19, 2021, to collect normal operation data.

The enthalpy change model was trained using the normal operation data. Shown in Figure 6.15, the measured enthalpy change matches well with the predicted enthalpy change, and the average of the deviations between the two is 0.15 Btu/lbm. As shown in Figure 6.15, the resulting deviation is within the calculated threshold ±0.34 Btu/lbm most of the time.

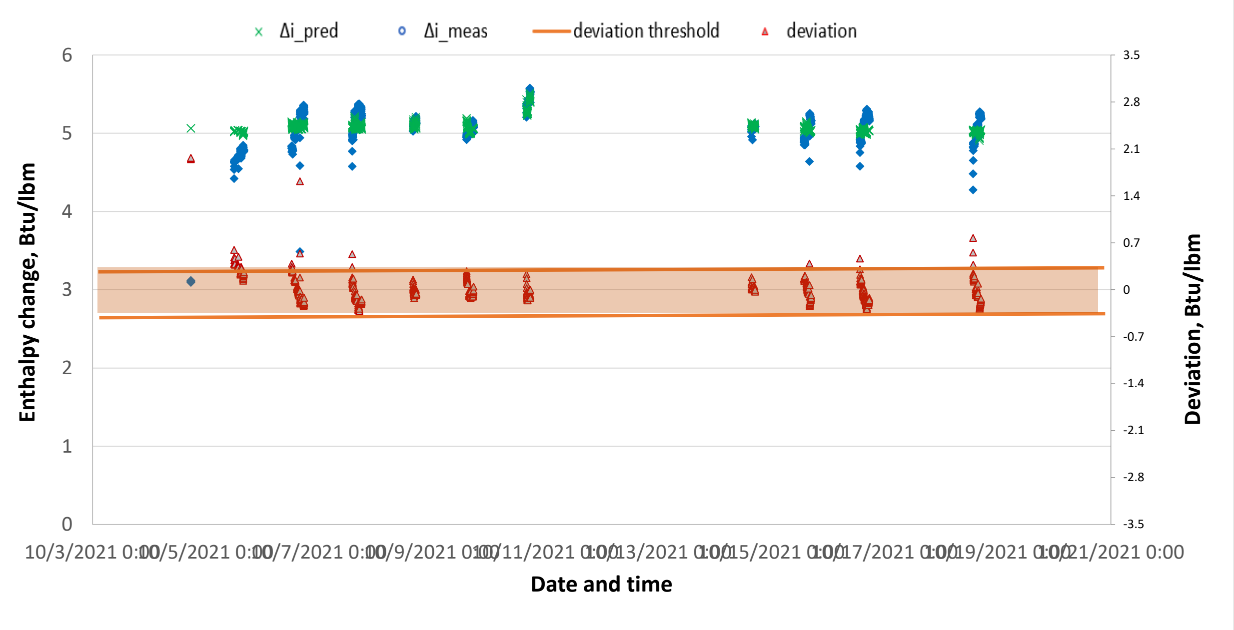


Figure 6.15. Comparison between estimated and measured enthalpy changes for PNNL House B (dry climate)

The data collected for the low flow rate fault are subsequently analyzed. As mentioned, at the beginning of the test the flow rate was reduced by 33.7%, but the airflow reduction continued to decrease as the test went on, which was confirmed by the handheld device. Because there is no permanently installed flow station, the continuous variation of the flow rate reduction is not recorded. However, from Figure 6.16, it is obvious that the measured enthalpy change is always higher than predicted enthalpy change except for a few spikes at the beginning of the test, i.e., the deviations (red triangles) are lower than the negative bound of the threshold, which is consistent with physical heat transfer laws. However, the deviation drastically reduced in the later part of the test. It might be caused by flow rate reduction varied throughout the test duration. This indicates that the AFDD algorithm is sensitive to the severity of the fault.

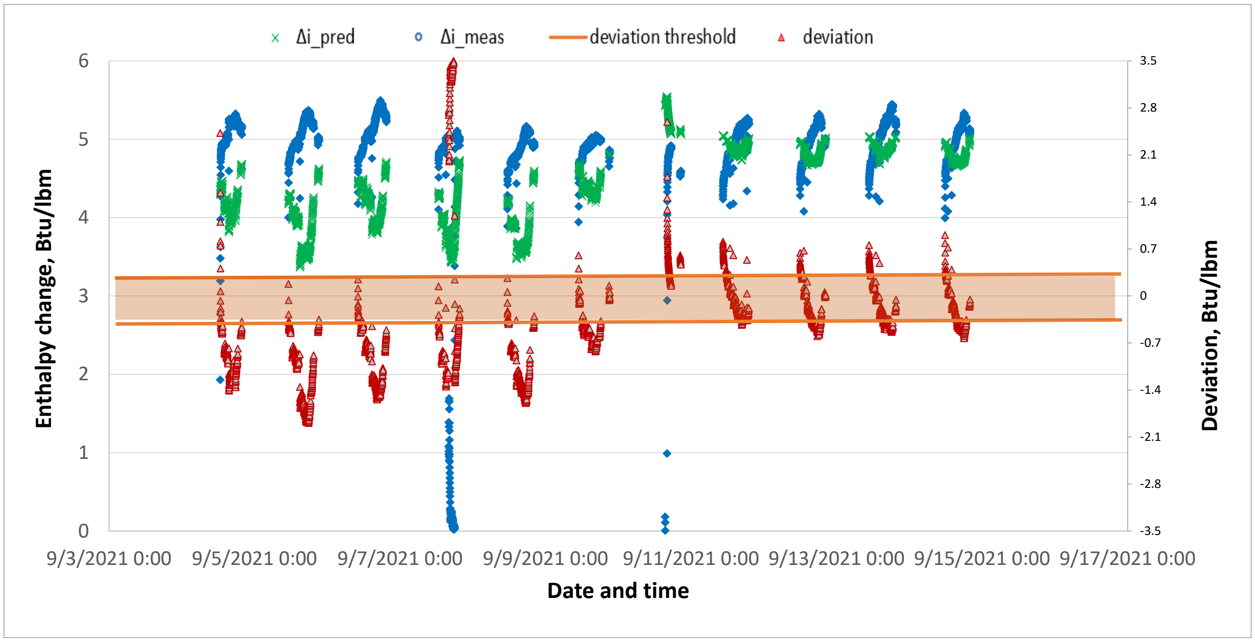


Figure 6.16. Test for airflow rate reduction fault in PNNL House B (dry climate)

Next, the data collected for the 30% undercharge fault are analyzed. As shown in Figure 6.17, the measured enthalpy change is consistently lower than the predicted enthalpy change, i.e., the deviations (red triangles) are well beyond the positive bound of the threshold band, which is consistent with the physics of the vapor compression cycle. All the deviation points exceeded 1 Btu/lbm, which is well above the threshold of 0.45 Btu/lbm. Therefore, the 30% undercharge fault is easily detected using enthalpy change across the evaporator for dry climate. The AFDD algorithm is more effective/sensitive to detect faults under a dry climate because there is no need for the SADP calculation, which generates errors due to using the assumptions.

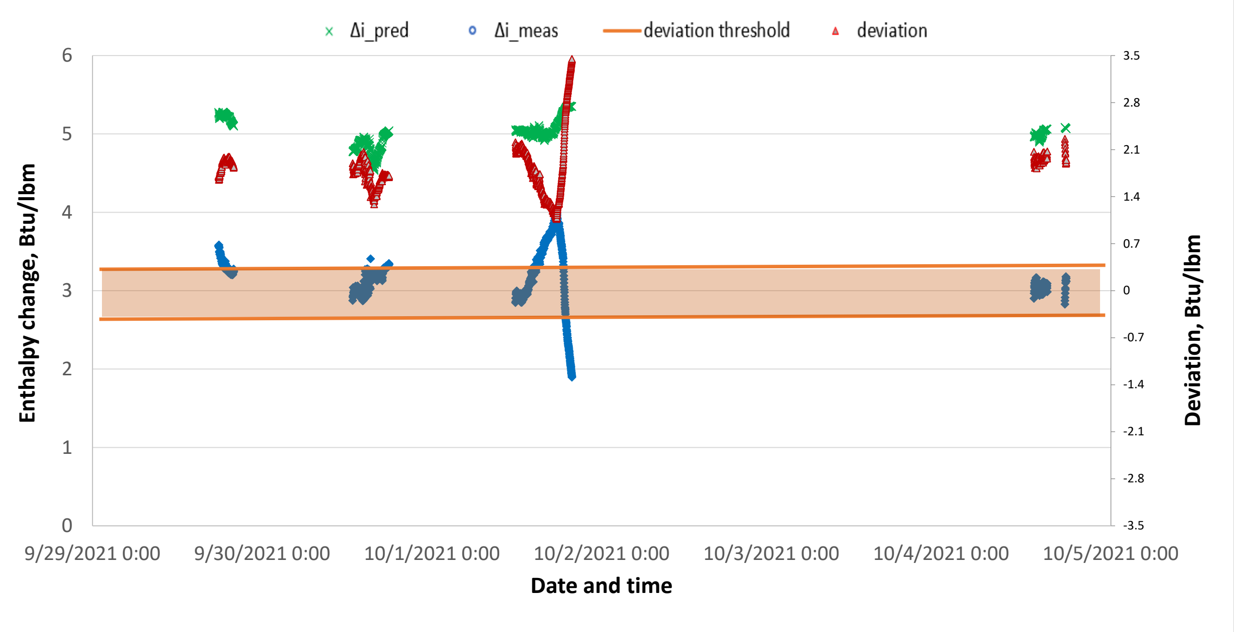


Figure 6.17. Test for 30% undercharge fault in PNNL House B (dry climate)

Finally, Method 2, which uses fan power measurements to detect an airflow fault, is validated using normal operation data as well as data containing a flow rate fault. In Figure 6.18, the blue circles show the measured fan power under normal operation conditions. It is very consistent with the estimated fan power except for a few transient points when the unit started and stopped. Therefore, the deviation percentage (red triangles) is close to zero and stays within the threshold of ±3%. The test results for using fan power to detect low airflow faults are shown in Figure 6.19. It is observed that fan power reduction for the 30% flow rate reduction is not consistent for a period of time. The inconsistency might be caused by the fact that the fan in the test system is driven by a multispeed motor. When the flow rate is reduced by blocking the filters, the motor might have the capability to adjust its speed to accommodate for the high resistance in the system created by the filter blockage. However, the project team is not able to verify this suspicion due to the lack of access to the test unit. Nevertheless, it is confirmed that the test unit is driven by a four-speed fan through checking the manufacturer specifications.

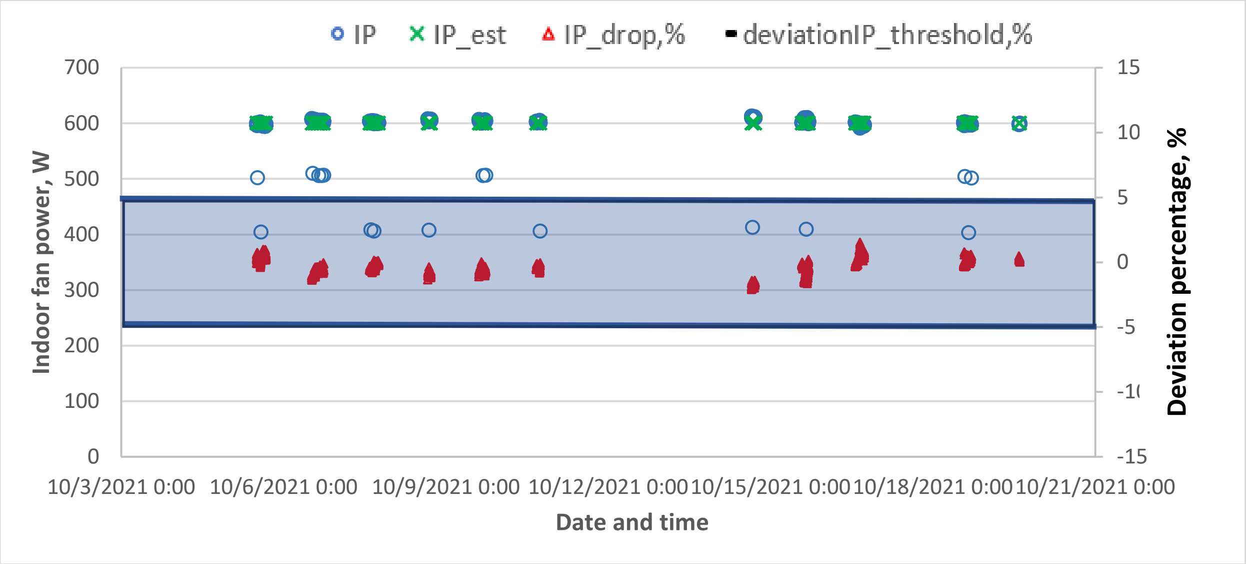


Figure 6.18. Measured fan power and predicted fan power for normal operation (dry climate)

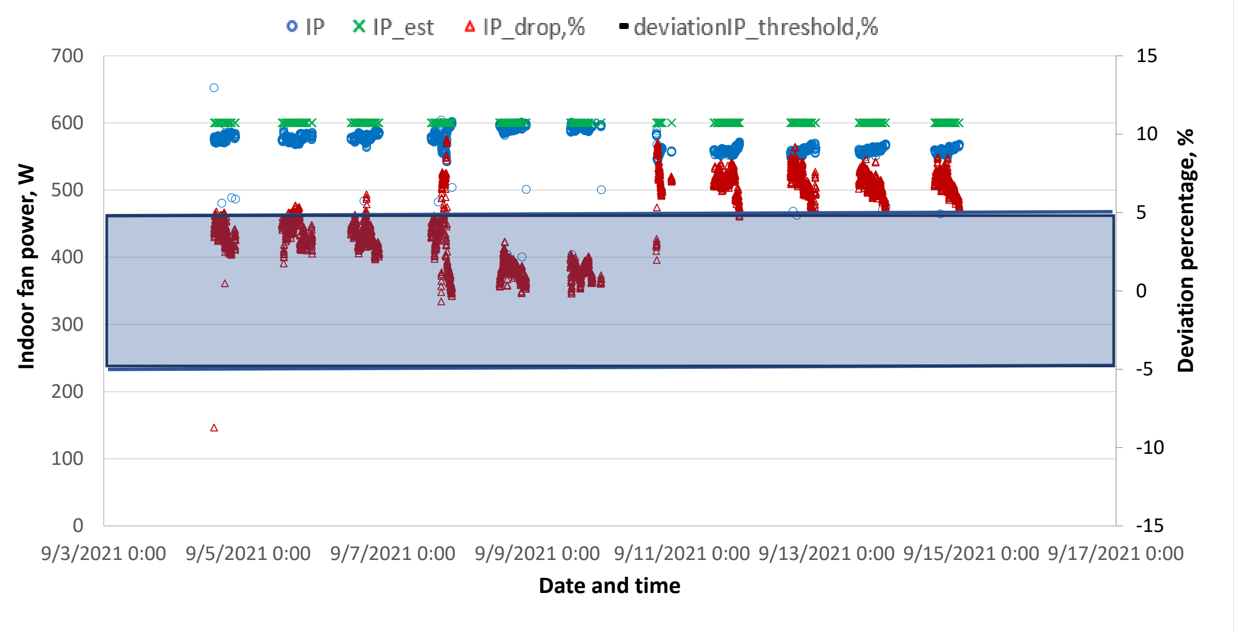
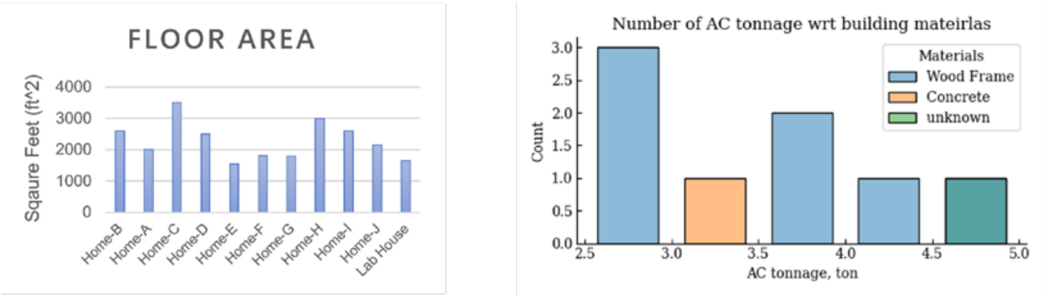


Figure 6.19. Fan power analysis for low flow rate fault in PNNL House B (dry climate)

7 Field Test

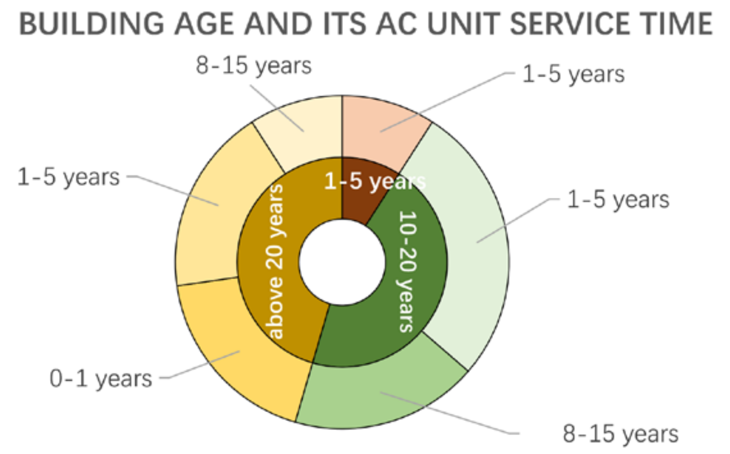
7.1 Introduction of Participating Test Houses

Although 10 homes were required for field tests to complete the project, we recruited 11 homeowners to account for unexpected circumstances. There were 10 single-family homes in Norman, Oklahoma, and 1 single-family home in Miami, Florida. To ensure that a diverse set of test houses were chosen, a survey was done for each home to gather building characteristics including their locations, floor areas, building ages, materials for the structure of the buildings, year of the AC units, and whether the homeowners have signed up for a TOU rate. Figure 7.1 shows the distribution of floor areas (subplot (a)), AC tonnages (subplot (b)), building ages, and year of the AC units (subplot (c)) for the 11 houses. Except for the lab house located near the University of Oklahoma, each test house was labeled alphabetically depending on when it was registered with our project. As can be seen from Figure 7.1, the test houses are diversely distributed in terms of floor areas (from 1,500 ft^2 to 2,500 ft^2) and building ages. Although only one new house (less than 5 years old) was selected, the year of the AC units inside the other houses (10–20 years or above) vary from 0–1 years to 8–15 years.



(a) Floor size of each building

(b) AC tonnage distribution with respect to building materials



(c) Building ages (inner) and its AC unit service time (outer) information

Figure 7.1. Information about the eleven (11) test houses

Table 7.1 shows detailed information for all the test houses. During the test period, only two participants (Home C and Home J) were enrolled in the Smart Hour programs. For those who had not registered for the TOU rate, we carried out a price-based control to test the effectiveness of the MPC-based optimization algorithm, assuming that the houses had registered for the TOU rate. The Smart Hour programs from local utility services inspired the referenced price signals used in the real control tests.

Table 7.1. Test House Information and TOU Enrollment

Home	Size (sq. ft.)	Building Age	Material of House	Year of AC	# AC units	AC tonnage	TOU
Home A	2,400	10–20 years	Wood Frame	1–5 years	1	5	-
Home B	2,600	10–20 years	Concrete	1–5 years	2	3 (up)/4 (down)	-
Home C	3,500	above 20 years	Wood Frame	0–1 years	2	2.5(up)/4(down)	Yes
Home D	2,500	10–20 years	Wood Frame	8–15 years	1	3 .5	-
Home E	1,540	1–5 years	Wood Frame	1–5 years	1	2.5	-
Home F*	1,812	10–20 years	Wood Frame	1–5 years	1	N/A	-
Home G	1,800	above 20 years	Wood Frame	1–5 years	1	4	-
Home H**	3,000	above 20 years	Wood Frame	8–15 years	1	N/A	-
Home I	2,600	above 20 years	unknown	0–1 years	1	5	-
Home J	2,150	10–20 years	Wood Frame	8–15 years	1	2.5	Yes
Lab House	1,658	above 20 years	Wood Frame	1–3 years	1	3.5	-

* The house was sold in June 2022. We could not conduct the MPC test with the new homeowner.

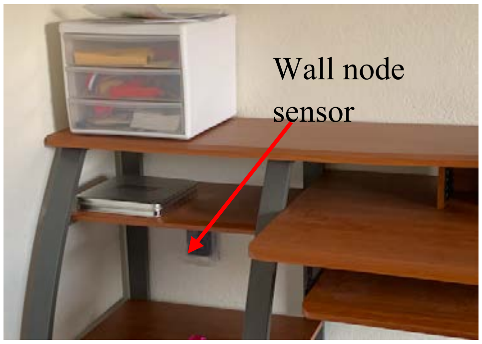
** The homeowner did not disclose that the house had two units, one with a Nest thermostat and another with a traditional thermostat. The homeowner only agreed to replace the traditional thermostat after an ecobee thermostat was installed. Therefore, no further tests were conducted for this house.

To prepare for the field tests, we installed ecobee thermostats at all the 11 test houses for data collection and ecobee node sensors for wall surface temperature measurements, which facilitate the learning model identification and verification and MPC-based optimization algorithm. In addition, we selected five houses (Home A, Home B, Home I, Home J, and the Norman Test house) for smart meter and additional node sensor installations to collect indoor and outdoor power measurements and supply air temperature measurements at one of the closest diffusers to facilitate the AFDD test. Moreover, we constructed a cloud-based data management and control system to access all the data collected with the ecobee thermostats and smart meters. Through the ecobee API, the participating AC units can be directly controlled using the MPC algorithm for real-time control purposes. However, the field tests were conducted only for nine homes, after consulting with the DOE project management team, for unexpected reasons. As indicated in Table 7.1, two houses were excluded from our analysis. One house (Home F) was sold in June 2022, six months after we set up our experiments. The other house (Home H) was excluded because the homeowner did not disclose that the house had two units, with one of them replaced with a Nest smart thermostat. The homeowner did not want to replace the Nest thermostat with ecobee.

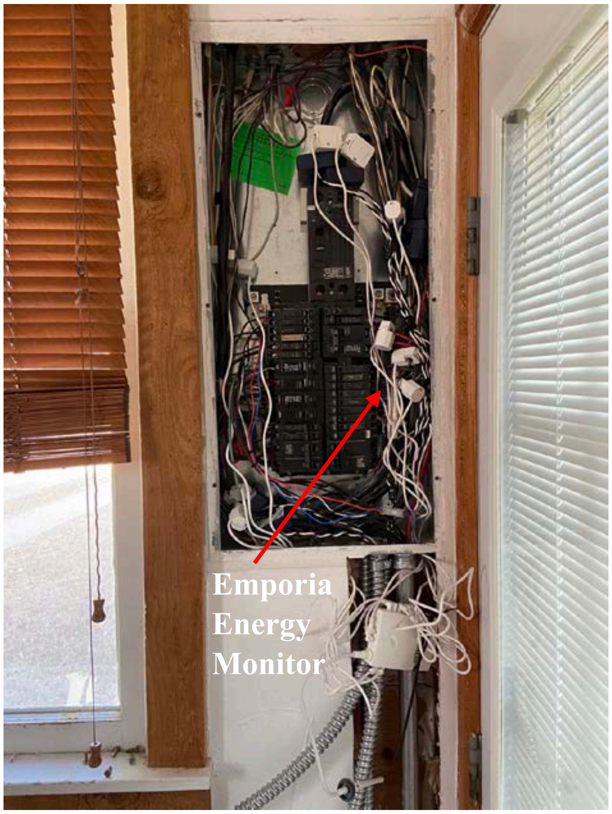
Figure 7.2 shows the temperature and power sensor installed in one of the houses. Subplot (a) shows an ecobee thermostat, which recorded indoor air temperature, HVAC on/off signal, the user's HVAC set point, and their schedule setting. Subplot (b) shows the Emporia energy monitor installed in the Norman Test House. The energy monitor measures the total energy usage of each individual circuit inside the electric panel. Thus, it is possible to distinguish between the power consumed by the HVAC's indoor and outdoor units. Subplot (c) shows the wall sensor used by the MPC algorithm. We put the node sensor that came with the ecobee thermostat in a plastic box and hung it on the wall.

The wall surface temperature was an important measurement used by the MPC algorithm. In the home thermal model described in Equations (3.7) and (3.8), the wall surface temperature was one of the states of the building and reflected the heat transfer between the outdoor environment and indoor air. As discussed in Sections 4.3.1 and 4.3.2, the sensor location was a key factor in making the model perform well. Thus, we mounted the wall sensor so that there was no direct solar impact, no wind impact, and no shading impact. We selected north-facing walls because the solar impact from the sun was lower compared with south- and west-facing walls. We also chose east-facing walls for some of the test houses depending on their structure. We tried to avoid surface walls near a diffuser, the kitchen, and the windows to not create disturbances.

(a) Smart thermostat in one residential house



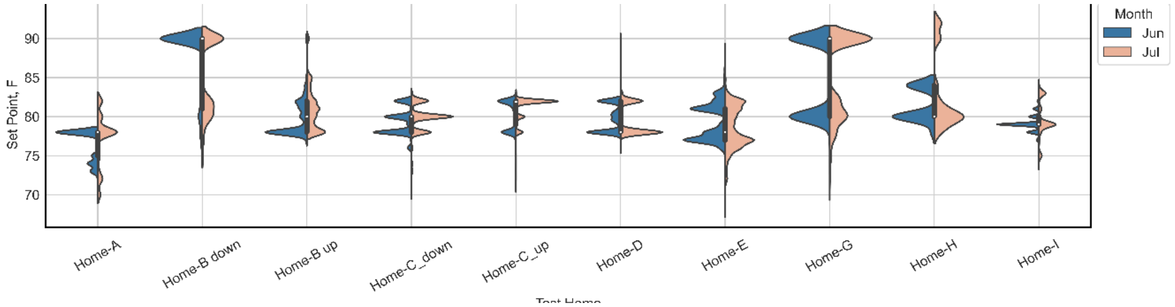
(c) Wall node sensor in another residential house



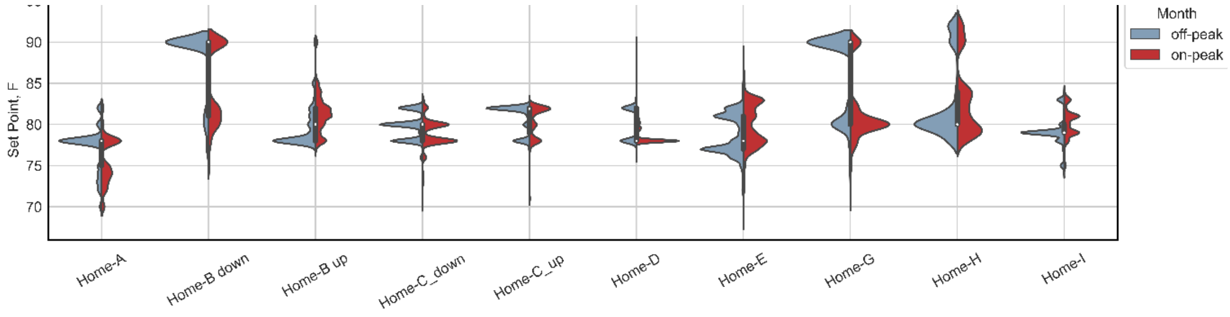
(b) The Emporia Vue which monitor the power

Figure 7.2. Installed smart thermostat, node sensor (surface temperature), and smart meter in one test home

Based on the data collected through the cloud-based system, the energy consumption of each home's AC unit was analyzed and found to depend on many factors, including the floor area, the building age, the AC unit’s capacity, and the materials used in building construction. A user’s energy awareness is also an important factor that might affect the electricity cost of a house. Subplot (a) of Figure 7.3 illustrates the distribution of the AC set points for eight of the test houses in June (blue) and July (orange). Note that some houses chose to turn off their AC completely when it was not needed, and as a result the thermostat would not update the set point during the AC-off period. In this case, we adjusted the set point to 90°F, as shown in the figure. It can be observed that for some houses, the AC set point preferences changed seasonally. Indeed, Home B (upper unit) preferred a lower set point on a cool summer day in June and a higher one on a hotter day in July. In contrast, Home E preferred a higher set point in June and kept lower set points when it got hot outside. Home A and Home I maintained a constant set point in June but experienced frequent changes between high and low set points in July. The ACs on the main floor of Home B and Home G were turned off in June and July more frequently than in other houses. Finally, the ACs on the main floors of Home C and Home D had consistent set point preferences during June and July. In addition to seasonal variations, users also changed their set points at different times of the day. Subplot (b) in Figure 7.3 is a violin plot showing the set point distribution for off-peak hours (dark blue) and on-peak hours (red) in June and July. During off-peak hours, Home B's upper unit and Home E slightly lowered their set points, while during on-peak hours, their set points were raised higher. In these cases, the concept of pre-cooling was used, which might indicate a better energy awareness of the homeowners. For some other houses such as Home D, the homeowner preferred to use a slightly lower set point during on-peak hours to maintain thermal comfort, which might lead to a higher electricity cost. In this study, although price-based control was performed to achieve the lowest electricity cost, we still tried to keep the indoor air temperature within a given temperature range to maintain thermal comfort during the MPC test. The upper and lower temperature limits were determined based on our observations of user thermal preferences (summarized in Figure 7.3) and users' requests for acceptable temperatures. Note that the Norman Test House, Home F, and Home J are not shown in Figure 7.3. The lab house was unoccupied and different tests were conducted in the house during June and July, during which the thermostat set points were chosen to make the HVAC system run longer. Therefore, its set points did not represent the preference of real users. Lastly, Home F was sold in June, while Home J suffered data loss in June due to Wi-Fi issues.



(a) Set point distribution for eight houses in different months



(b) Set point distribution for eight houses in on-peak and off-peak hours

Figure 7.3. Set point distribution in eight different test houses

7.2 Implemented MPC-Based Optimization Agent and Its Test Results

Figure 7.4 illustrates the framework of the MPC agent implemented in this study. At each 5-minute time step, the MPC agent receives time-varying (blue) and constant (yellow) inputs, and temperature measurements (red arrow), as shown in Figure 7.4. The home thermal model (2R2C model) first receives forecasts of the disturbances (wind speed, solar radiation, outdoor air temperature) over a prediction horizon of 12 hours, model parameters identified from the most recent data, and the current building states including the indoor air and wall surface temperature measurements. The home thermal model then predicts the future building states (indoor air and wall surface temperatures) and interacts with the optimization model. In a parallel fashion, the optimization model receives the TOU rate, heat pump energy usage model, and thermal comfort range, through which a cost function and constraints are generated. Gurobi (Gurobi Optimization 2021) is then used to solve the optimization problem to generate a sequence of optimal on/off decisions for the HVAC system over the prediction horizon. However, as is typical with MPC algorithms, only the first element of the sequence of optimal decisions, whether it is on or off, is physically implemented.

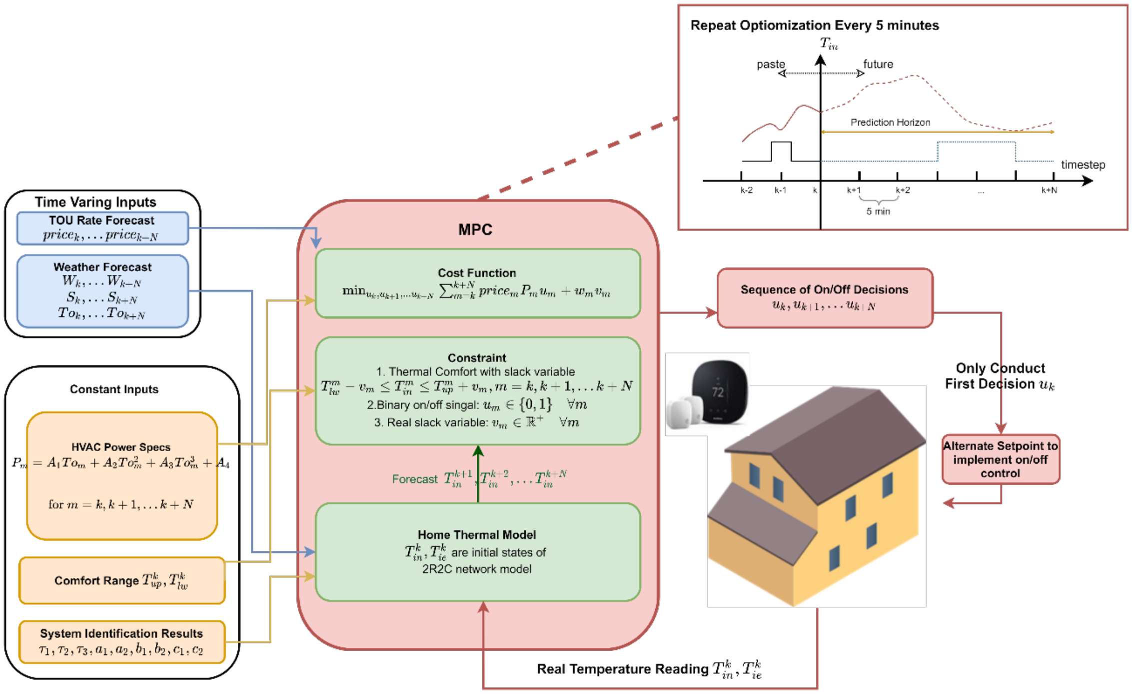


Figure 7.4. MPC controller schematic

7.2.1 Implemented MPC-Based Optimization Agent

The optimization problem addressed by the MPC agent at each time slot k is stated below:

$$\min_{u_k,u_{k+1},\dots u_{k+N}} \sum_{m=k}^{k+N} price_m \, P_m u_m + w \, v_m \tag{7.25}$$

$$T_{lw}^m - v_m \le T_{in}^m \le T_{up}^m + v_m \quad \text{for } m = k, k+1, \dots, k+N \tag{7.26}$$

$$\frac{T_{ie}^{m+1} - T_{ie}^m}{\Delta t} = \frac{1}{\tau_1}(T_o^m - T_{ie}^m) + \frac{1}{\tau_2}(T_{in}^m - T_{ie}^m) \quad \text{for } m = k, k+1, \dots, k+N \tag{7.27}$$

$$\begin{aligned}\frac{T_{in}^{m+1} - T_{in}^m}{\Delta \mathrm{t}} &= \frac{1}{\tau_3}(T_{ie}^m - T_{in}^m) + \frac{1}{\tau_3}\left(a_1 S^m + a_2 S^{m^2}\right) \\ &+ \frac{1}{\tau_3}(T_o^m - T_{in}^m)\left(b_1 W^m + b_2 W^{m^2}\right) + \frac{1}{\tau_3}\left(c_1 T_o^m + c_2 T_o^{m^2}\right)u^m \quad \text{for } m \\ &= k, k+1, \dots, k+N\end{aligned} \tag{7.28}$$

where T_{in} is the indoor air temperature in °F, T_{ie} is the interior wall surface temperature in °F, T_o is the outdoor air temperature in °F, W is the wind speed in mph, S is the solar radiation in $\frac{kW}{m^2}$, and u is the HVAC on/off signal. Note that there are nine unknown parameters in the model: τ_1, τ_3 are time constants of the house envelope and indoor air associated with the thermal resistance R and thermal capacity C of the house; τ_2 is a coefficient associated with the thermal resistance of air; a_1, a_2 are coefficients representing the solar radiation's contribution to indoor air; b_1, b_2 are coefficients representing the wind's contribution to indoor air; and c_1, c_2 are coefficients representing the ACs output to the room. The prediction horizon in this study is 12 hours, and the time step is 5 minutes, so $N = 144$. The variable $price$ is the utility rate determined based on the local utility company's Smart Hour program. In a vapor compression system, the evaporator inlet condition and condenser inlet condition are the most important factors that affect the power consumption of the air-source heat pump or outdoor unit. For a split system, the outdoor air condition is measured at the condenser inlet and the indoor air condition is measured at the evaporator inlet. The variation of the outdoor air in a cooling season is much larger than that of the indoor air. Hence, to maintain linearity of the cost function, we assume that the power consumption of the outdoor unit depends only on the outdoor air temperature through:

$$P_m = A_1 T_{om} + A_2 T_{om}^2 + A_3 T_{om}^3 + A_4 + P_{indoor} \text{ for } m = k, k+1, \dots, k+N \tag{7.29}$$

Lastly, the power consumption of the indoor unit is relatively constant, and thus is assumed to be so in the study.

In this study, only five houses are equipped with power measurement devices. Polynomial regression fit is used to find the unknown parameters A_1, A_2, A_3, A_4 for those houses. For the rest of the houses, a constant approximation of the total AC outdoor and indoor power consumption is selected based on their heat pump information.

Every 5 minutes, the lower and upper bounds of the indoor air temperature are sent to the MPC agent. The MPC agent then solves the optimization problem to generate the optimal HVAC on/off control signal. In general, the optimization problem might be infeasible because the initial and subsequent indoor air temperature appearing in some of the constraints depend on sensor reading, which might be noisy and biased. In fact, even if the temperature bounds—treated as hard constraints in $T^m_{lower} \leq T^m_{in} \leq T^m_{upper}$—are satisfied for every sampling instant $k, k+1, \dots, k+N$, it is still possible for the temperature to fluctuate beyond its lower or upper bound between two consecutive sampling instants that are 5 minutes apart. To address these issues, we "soften" the constraint by using a non-negative real slack variable v^m to represent the distance from the current temperature to its upper (or lower) bound when the temperature was beyond the given range. In the cost, $\mathrm{wv^m}$ was added, which only imposed a penalty when the temperature was higher than the upper bound or lower than the lower bound. In other words, the temperature constraints were relaxed only when necessary. By relaxing these constraints, the MPC agent no longer had to deal with an infeasible optimization problem which might cause the system to crash.

7.2.2 Experimental Setup and Data Management

In this section, we describe the MPC pre-cooling test setup for all the test houses. For each MPC test, the time step was 5 minutes (the finest granularity in the thermostat), and the prediction horizon was 12 hours. In addition to the control algorithm itself, one of the bottlenecks in implementing MPC in residential buildings is the financial burden created by the need for data communication between the building and the controller. Fortunately, the proliferation of IoT devices and availability of open-source databases have facilitated the implementation of advanced control in residential buildings. In the rest of this section, we introduce the data source, data pipeline, and equipment installation for the tests.

We divided the datasets required for the tests into three categories based on their purposes: the system identification dataset, the dataset for real-time control, and the dataset for post-analysis. Separate Python modules were developed to collect and pre-process the data in each category. The open-source time-series database InfluxDB 2.0 (InfluxDB 2021) was used for data storage and communication. Grafana v7.4 (Grafana Labs 2021) was used for data display and user interface.

The system identification dataset was made up of historical data that included the system state dataset, weather dataset, and HVAC power consumption dataset. The system state dataset and weather dataset were used to train the home thermal model, while the HVAC power consumption dataset was used to train the regression coefficients for power estimation at each house. The states of the system included the indoor air temperature and HVAC on/off signal collected from the ecobee thermostat as well as the wall surface temperature collected from the ecobee node sensor. The historical outdoor air temperature, solar radiation, and wind speed were obtained from Mesonet in Oklahoma (Brock et al. 1995; McPherson et al. 2007) and other weather services (OpenWeather 2018). The HVAC power consumption dataset was obtained from the smart energy monitor if such a device was installed in the house.

The dataset for real-time control included the building temperature measurements, weather forecasts, users' preferred temperature ranges for comfort, and utility rates. During real-time control, every 5 minutes a building temperature measurement from ecobee and its node sensor was set as the initial state of the system while a 12-hours-ahead weather forecast was collected using the weather API. Weather forecasts, especially solar radiation forecasts, were relatively expensive to obtain for residential control tests, so these forecasts were only updated once every 6 hours. The temperature range for comfort was obtained from the homeowner's feedback through emails. Because the test houses in this project were located in different states and enrolled in different utility programs, three different rate structures were adopted: a TOU rate that followed Oklahoma Gas & Electric's (OG&E) Smart Hours (OG&E 2022) schedule, with on-peak hours starting at 2 p.m. and ending at 7 p.m., and off-peak hours lasting 19 hours; a rate that followed Oklahoma Electric Cooperative's (OEC) TOU rate (OEC 2022), for which the on-peak hours were from 3 p.m. to 7 p.m. and the off-peak hours lasted 20 hours; and a rate that followed Florida Power and Light's TOU rate (Florida Power & Light 2022), for which the on-peak hours were from 12 p.m. to 9 p.m. and the off-peak hours lasted 15 hours.

Apart from datasets for the days MPC tests were conducted (called MPC days), we also collected datasets for the days where MPC tests were not performed (called normal operation days) to enable post-analysis. These historical datasets included the system states of the buildings, weather information, and HVAC power consumption stored in InfluxDB. Additional information, such as the user's HVAC set point, HVAC system mode (i.e., whether it was in cooling mode, heating mode, or completely off), and other scheduled vacation time, was needed to enable performance comparison between MPC days and normal operation days.

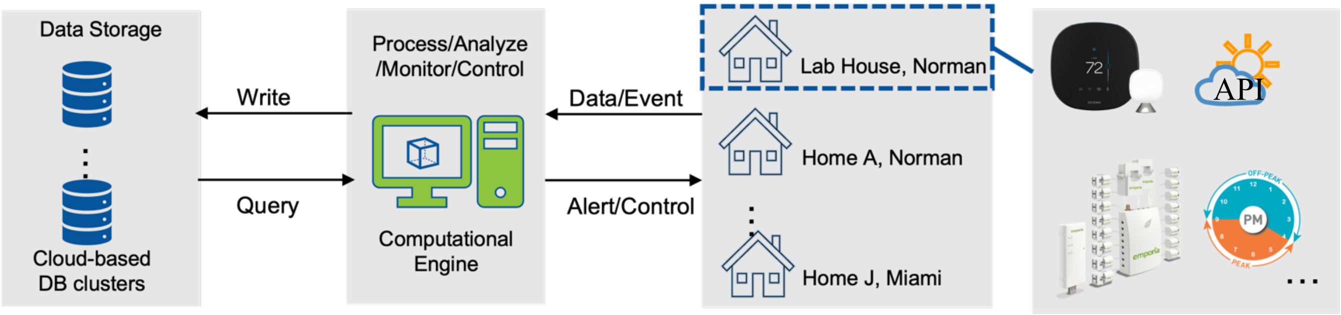


Figure 7.5. Data pipeline of MPC agent

7.2.3 Field Test Results

In this section, we describe the MPC test results for multiple real buildings. We first present the MPC test results for a single home, in which the HVAC system operation, indoor air temperature, and interior wall surface temperature were accounted for. We then compare the MPC test results across multiple houses by showing the cost savings, run time, and power consumption during both MPC days and normal operation days. We also discuss possible factors that could affect the results of MPC tests in different houses. In addition, a data mining technique was used to make a fair comparison for multiple houses.

7.2.3.1 MPC Test Results in a Single-Family House

Figure 7.6 shows seven consecutive days of MPC testing in the lab house in Norman, Oklahoma. The horizontal axis in each subplot of Figure 7.6 represents time from August 4 to August 10, 2022. The first subplot shows the building temperature measurements. The second subplot displays the total number of seconds the AC was on during each 5-minute time interval. The last subplot shows the actual local weather information. In the first subplot, the blue curve represents the indoor air temperature queried from the ecobee thermostat, while the red curve represents the wall surface temperature queried from the node sensor mounted on the north-facing wall. The two black solid lines represent the upper and lower bounds on indoor air temperature that were considered to be comfortable. The on-peak hours are shown as red-shaded areas. Note that most of the time, the indoor air temperature remained within the two bounds. This was because whenever the temperature exceeded either the upper or lower bound, the slack variable in the optimization problem imposed a penalty on the objective function. If this happened during on-peak hours, such a penalty might not be significant enough to cause the HVAC to turn on immediately because electricity cost contributed more to the objective function. As shown in the first subplot, the indoor air temperature sometimes went above the upper bound for a while during on-peak hours. The indoor air temperature also reached $83^{\circ}\mathrm{F}$ on August 10 due to the MPC controller going offline unexpectedly between 2:37 p.m. and 5:44 p.m. Compared to the indoor air temperature, the wall surface temperature changed relatively slowly during the test, possibly because the building acted as a large thermal mass. From Figure 7.6, it can be seen that the pre-cooling effect on each test day was quite obvious. Indeed, the AC ran hard before on-peak hours so that the indoor air temperature at the beginning of each on-peak period was usually around the lower bound. However, although AC operation time was reduced during on-peak hours due to the thermal mass of the building, the AC could not completely avoid operation during on-peak hours. This was because the temperature often reached the upper bound before each on-peak period ended. These results suggest that the MPC agent worked as intended.

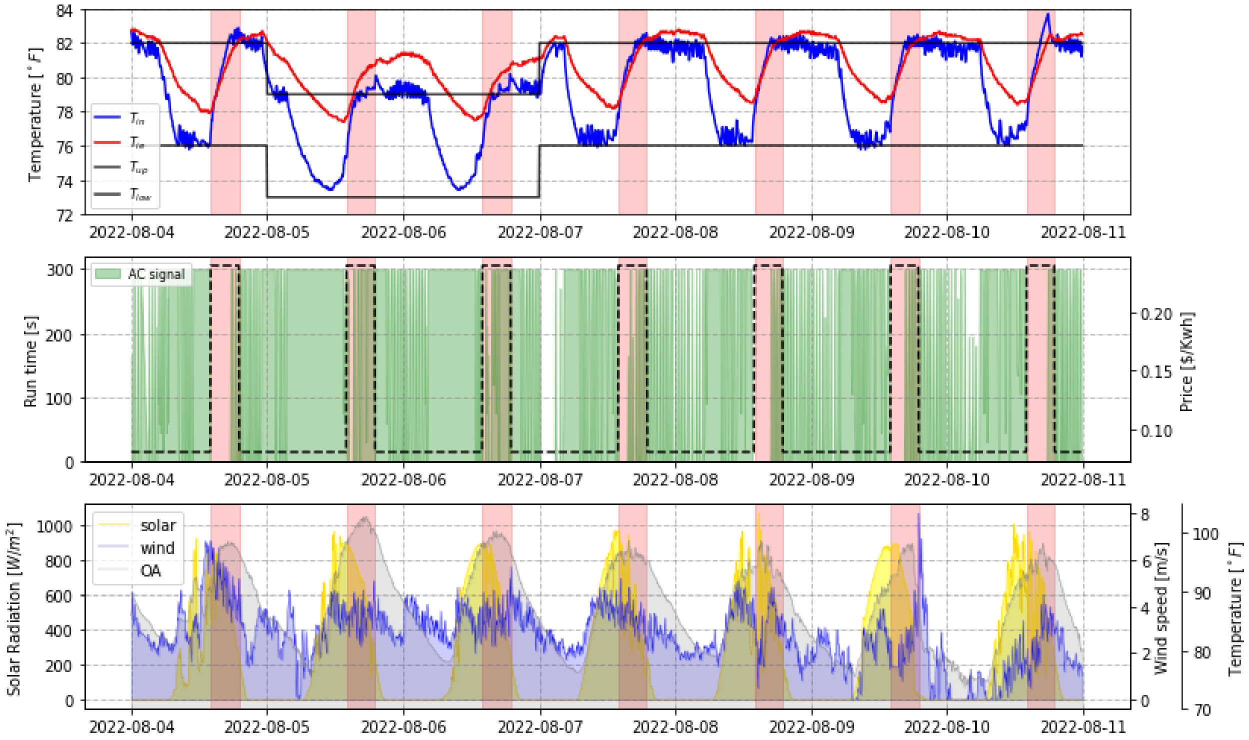


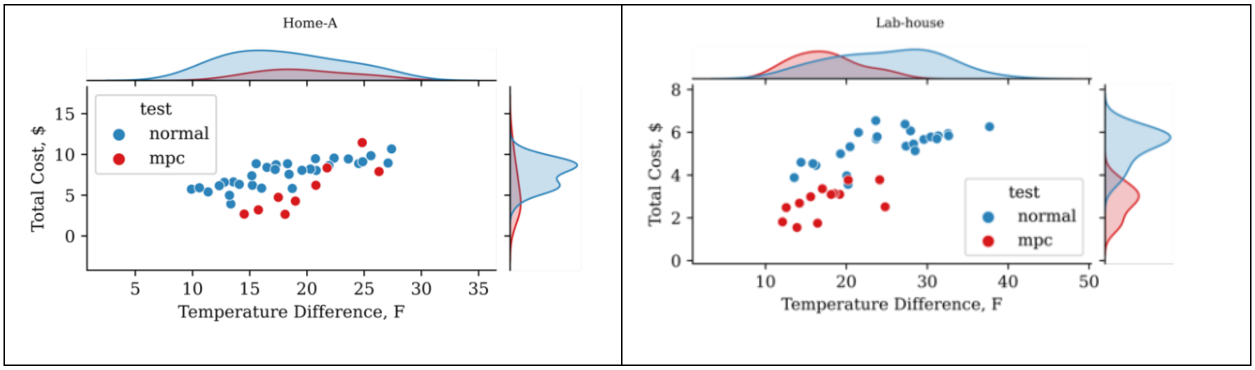
Figure 7.6. Real-time MPC test results and weather conditions over 7 consecutive days

7.2.3.2 MPC Test Results in Multiple Houses

As discussed in Section 7.1, not all houses participated in the MPC tests in summer 2022 due to some of the homeowners' personal requests. Home F was sold in June 2022, while Home H quit the MPC test because the homeowner preferred another smart thermostat. In addition, although we were able to conduct multiple MPC tests on Home B's main floor, we were not able to do the same on Home B's second floor because the homeowner wanted to work from home. From July to October 2022, we conducted MPC control tests at nine houses. The tests were conducted for different houses at different times over four months. Each test lasted 5 to 20 days, depending on the homeowner's schedule, except for the test at Home G, which was terminated prematurely per the homeowner's request.

7.2.3.2.1 Cost-Saving Analysis vs. Temperature Difference Between Indoor and Outdoor

Figure 7.7 shows the electricity cost versus the temperature difference for all test houses that ran the MPC tests. Each point on the figure represents the HVAC cost for one day, either with an MPC controller (red) or without an MPC controller (blue). In previous studies, researchers mostly focused on understanding the impact of a single variable (e.g., the indoor air temperature or the solar radiation) on the control results. In this study, we considered the indoor air temperature, which affected not only the power consumption of the outdoor unit and the building envelope heat transfer, but also the electricity cost. However, unlike previous studies, we accounted for the fact that the indoor air temperature also depended on the choices of set points on normal operation days and the choices of temperature bands on MPC days. Therefore, in our analysis we looked at the *temperature difference*, defined as the difference between the maximum daily outdoor air temperature and the daily average indoor air temperature, and used such a difference to determine the cooling cost for the MPC tests. The density curves for the temperature difference and the total daily electricity cost are shown on the top and right of each subplot, respectively. Certain homeowners chose higher set points throughout the day, which could result in very low electricity costs, especially on days when the outside temperature was not that high. In this study, days where the AC was mostly turned off, and days where the homeowners were known to be away from home for an extended period of time, were excluded. Additionally, days with cold or extreme weather conditions were excluded. One exception was Home D, which experienced COVID-19 in August 2022, and as a result the MPC test was postponed until late September, which had several cold days. For most of the houses, we observed a higher cost on both normal operation days and MPC days when the temperature difference was higher. In addition, for most tests, electricity costs on MPC days were lower than those on normal operation days under the same weather conditions. However, there was no clear evidence that showed that a larger temperature difference led to a higher percentage of cost savings whenever the MPC controller was used. For example, the subplot on the top right corner of the figure is a box plot showing the cost savings of Home A at different temperature differences. In this subplot, the cost difference between normal operation days (blue) and MPC days (red) decreased as the temperature difference grew.



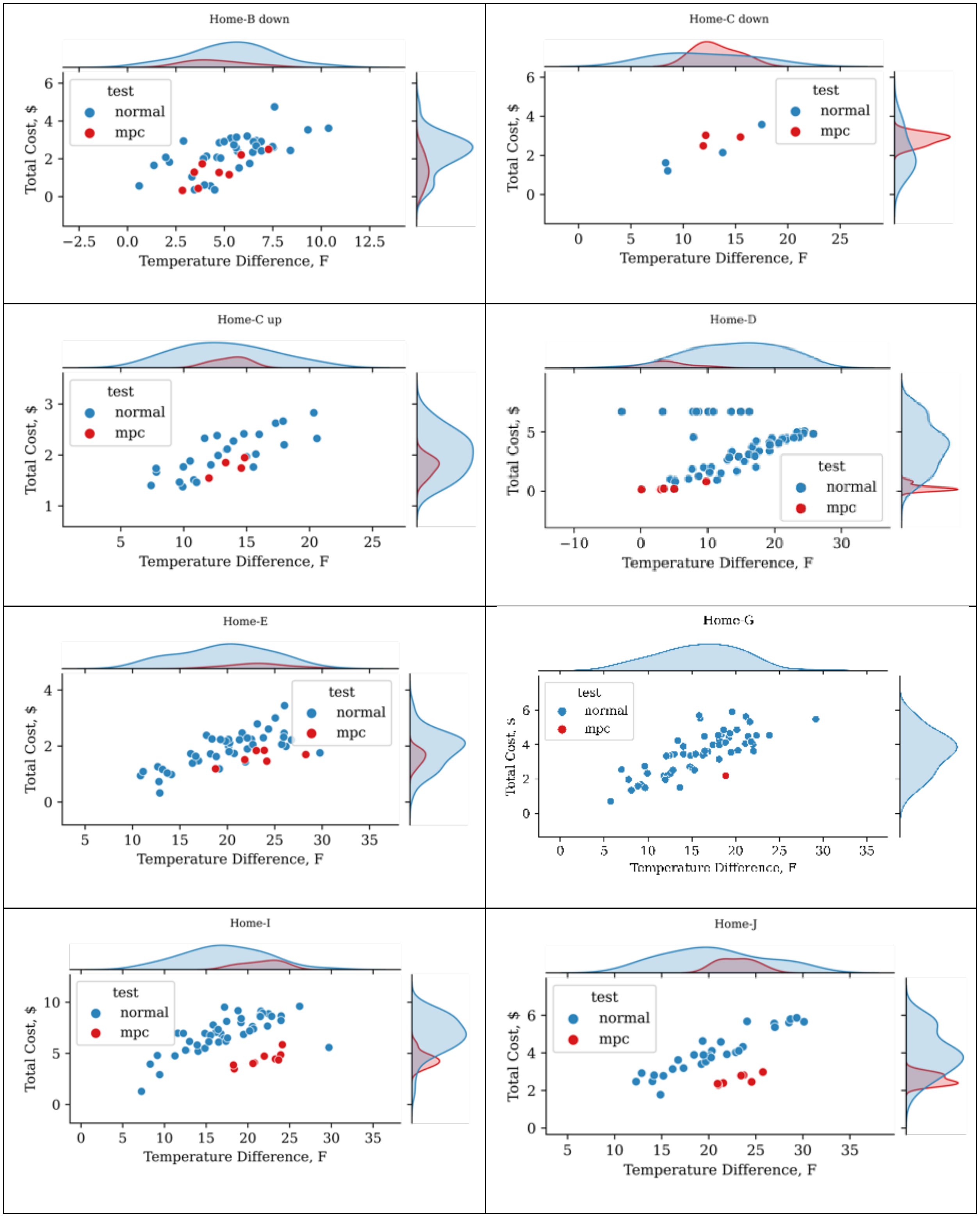
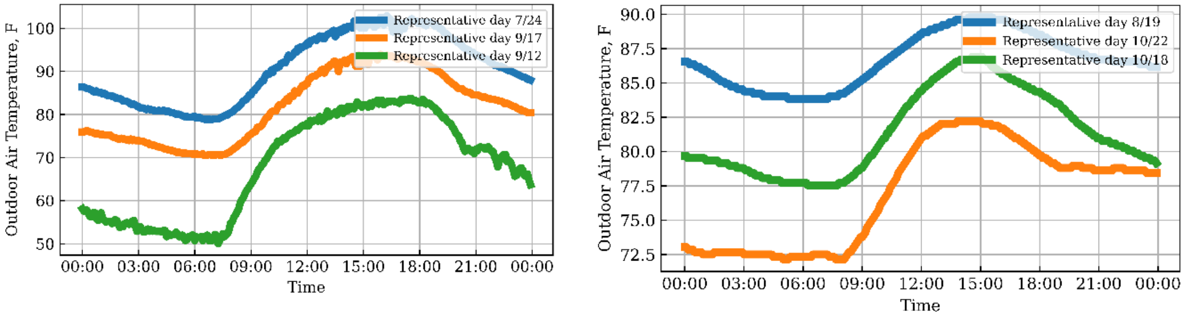


Figure 7.7. Total cost vs. temperature difference between outdoor and indoor when MPC was used and not used

7.2.3.2.2 Cost Reduction Calculation Through Weather Clustering

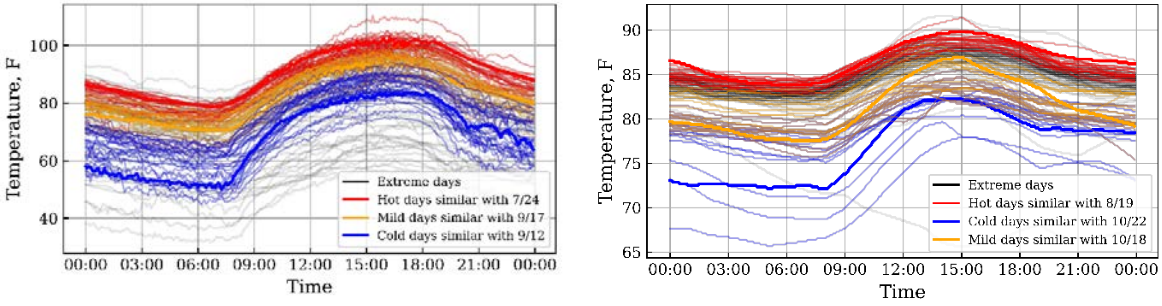
To evaluate the effectiveness of the MPC agent at different houses, we compared the electricity cost for both normal operation and MPC days under similar conditions. Weather conditions varied substantially and significantly affected the results. Thus, it would be unfair to claim that MPC saved more money when the daily cost of MPC on a colder day was significantly lower than the daily cost of normal operation on a hotter day. To more fairly evaluate the MPC agent, we first clustered Norman and Miami weather data by applying Matrix Profile-based algorithms from data mining applications (Yeh, Kavantzas, and Keogh 2017; Gharghabi et al. 2018; Imani et al. 2018). To cluster the long (about four months) time-series (time sequences rather than a single time point), and multidimensional (outdoor air temperature, solar, etc.) data, we used the snippets algorithm (Imani et al. 2018) to find the top three most representative days in the test period. We then selected the top *K* days similar to these representative days using our modified MASS algorithm (Mueen et al. 2022), which took into account solar radiation, outdoor air, and relative humidity. Through trial and error, we arrived at *K* equal to 25.

Figure 7.8 shows the weather clustering results for Norman ((a) and (b)) and Miami ((c) and (d)). The top two subplots ((a) and (c)) show the time-series snippets results, i.e., the top three representative days from July to October. From subplots (a) and (c), we see that the snippets algorithm indeed helped us find the hot, mild, and cold days. The bottom two subplots ((b) and (d)) show the time-series clusters found by the algorithm. In these subplots, each curve represents 24-hour weather data, and only the outdoor air temperature is shown. The solar radiation and relative humidity were also considered, but not shown in the clustering. The red, orange, and blue curves represent days associated with the hot, mild, and cold clusters, respectively. The gray curves represent the rest of the days, which were not part of any cluster and which are labeled as extreme days. For the Norman cluster, the extreme days included one day with a very hot morning and several days with relatively low temperatures. For the Miami cluster, the extreme days included one day with a relatively hot afternoon and several very cold days. For the Norman data, we see that the layer between each cluster is clear, which means that the data within each group were similar to one another. For the Miami data, however, the cold cluster is wide. The main reason is that Miami experienced a big thunderstorm followed by a short period of temperature rise and then a sudden temperature drop. This weather change started at the end of September and lasted until the second half of October. Compared with normal climate, included in the hot and mild clusters, the cold cluster included most of the abnormal weather change period. Thus, each day in the cold cluster has a large difference but similar shape.



(a) Norman representative days

(c) Miami representative days

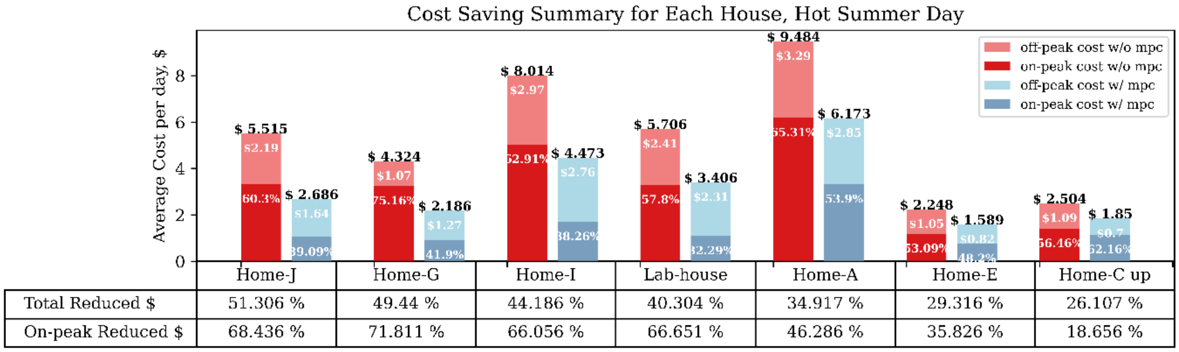


(b) Norman weather clusters

(d) Miami weather clusters

Figure 7.8. Weather clustering results in Norman and Miami from July to October

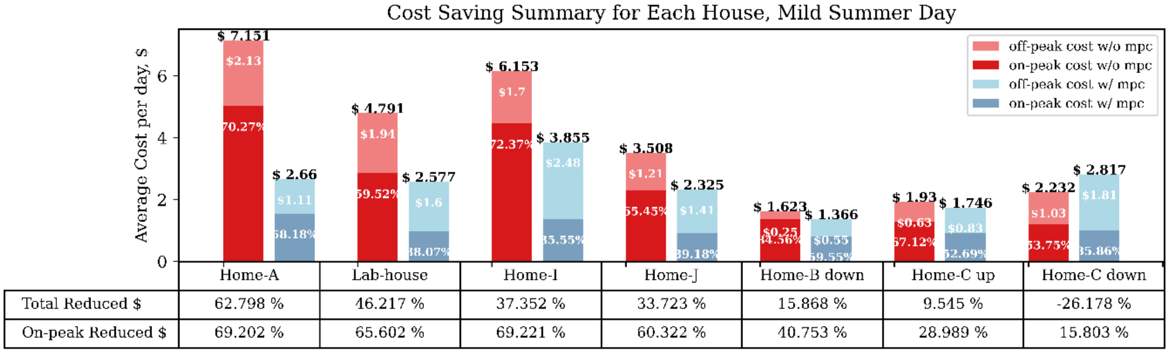
Figure 7.9 summarizes the cost savings during the hot summer test period established by the clustering algorithm. Notice that not all of the houses were tested during this period, and those that were not tested are excluded from the figure. The test houses in Figure 7.9 are listed in descending order of their total cost reduction percentage, defined as the total cost on normal operation days minus the total cost on MPC days and divided by the former. The dark red bars represent the average on-peak hour cost of hot days without MPC, and the light red bars on top of the dark red bars represent the average off-peak hour cost of hot days without MPC. The dark blue bars represent the average on-peak hour cost of hot days with MPC, and the light blue bars on top of the dark blue bars represent the average off-peak hour cost of hot days with MPC. The specific average off-peak costs for the hot days are listed inside the light red and light blue bars. The average total costs for the hot days, which are sums of the off-peak and on-peak hour costs, are listed on top of the bars. The average on-peak hour costs as a percentage of the total costs for the hot days are listed inside the dark red bars. To better visualize the impact of MPC, a table is added that compares the percentage reduction in total cost and on-peak cost with and without MPC. From Figure 7.9, we see that running the HVAC system with MPC indeed helped reduce homeowners' electricity bills to varying degrees during hot days. Across different houses, the average cost savings ranged from 51.306% to 26.107% on MPC days compared to normal operation days.



	Home-J	Home-G	Home-I	Lab-house	Home-A	Home-E	Home-C up
Total Reduced $	51.306 %	49.44 %	44.186 %	40.304 %	34.917 %	29.316 %	26.107 %
On-peak Reduced $	68.436 %	71.811 %	66.056 %	66.651 %	46.286 %	35.826 %	18.656 %

Figure 7.9. Cost savings summary for each house on hot summer days for normal operation days and MPC days

Figure 7.10 summarizes the cost savings during the mild summer days. The legend in Figure 7.9 also applies to this figure. MPC allowed homeowners to save money in most cases, with percentages saved ranging from 62.789% to 9.545%. For certain houses, however, the cost savings were not obvious (e.g., Home B and Home C).



	Home-A	Lab-house	Home-I	Home-J	Home-B down	Home-C up	Home-C down
Total Reduced $	62.798 %	46.217 %	37.352 %	33.723 %	15.868 %	9.545 %	-26.178 %
On-peak Reduced $	69.202 %	65.602 %	69.221 %	60.322 %	40.753 %	28.989 %	15.803 %

Figure 7.10. Cost savings summary for each house on mild summer days for normal operation days and MPC days

Table 7.2 presents detailed information on the cost savings and energy consumption for each test house with and without MPC during the hot summer days. The table includes daily averages of the total costs, costs during on-peak hours, energy consumption, and HVAC run time during on-peak hours. As can be seen from Table 7.2, MPC resulted in cost reductions for most test houses. Note that the maximum cost savings were achieved by avoiding frequent on/off AC operation during on-peak hours. In addition, the peak run time reduction was significant. Table 7.3 presents the same information for the mild summer days. Unsurprisingly, the cost savings due to using MPC under mild weather conditions were lower than that during hot summer days. Almost all test houses show encouraging test results, except for the test on the main floor of Home C. Because both the living room and bedrooms of Home C are on the main floor, the homeowner gave a 2-degree temperature band during sleep time, from 9:00 p.m. to 6:00 a.m., and a 3-degree temperature band the rest of the time. Due to such narrow temperature bands, MPC could not take full advantage of pre-cooling and, therefore, failed to cut electricity costs. Finally, Table 7.4 presents the same information for cold summer days. Note that despite the colder outdoor conditions, cost savings were still possible with MPC for the main floor of Home B and for Home D.

Although the above test results show promising cost reduction in multiple houses, a number of factors could have affected the amount saved. In this study, we found three factors that significantly affected the results: weather conditions, the AC's ability to cool the space in each house, and a user's energy awareness.

Table 7.2. Cost Savings and Energy Consumption Summary for Houses on Hot Summer Days

	Hot Summer Day							
Home	Normal Operation Days				MPC Days			
	Cost, $	Peak Cost, $	Energy, kWh	Peak run time, %	Cost, $	Peak Cost, $	Energy, kWh	Peak run time, %
Home A	9.48	6.19	66.93	93.84	6.17	3.33	49.43	55.06
Home C up	2.5	1.42	16.94	58.91	1.85	1.15	11.89	47.92
Home E	2.25	1.19	18.15	38.45	1.59	0.77	13.48	22.66
Home G	4.32	3.25	26.97	68.45	2.19	0.92	19.69	20.92
Home I	8.01	5.04	58.16	76.23	4.47	1.71	41.65	28.1
Home J	5.51	3.33	41.23	92.41	2.68	1.05	24.82	29.42
Lab house	5.71	3.3	43.84	89.88	3.41	1.1	33.41	32.86

Table 7.3. Cost Savings and Energy Consumption Summary for Houses on Mild Summer Days

Mild Summer Day								
Home	Normal Operation Days				MPC Days			
	Cost, $	Peak Cost, $	Energy, kWh	Peak run time, %	Cost, $	Peak Cost, $	Energy, kWh	Peak run time, %
Home A	7.15	5.03	47.51	80.54	2.66	1.54	20.35	24.96
Home B down	1.62	1.37	12.69	19.62	1.36	0.81	16.58	11.82
Home C down	2.23	1.2	15.46	31.01	2.82	1.01	22.51	26.11
Home C up	1.93	1.3	11.83	53.98	1.75	0.92	12.2	38.33
Home I	6.15	4.45	39.81	77.85	3.85	1.37	36.76	23.67
Home J	3.51	2.3	24.71	69.77	2.33	0.91	21.46	27.12
Lab house	4.79	2.85	36.13	90.15	2.58	0.98	24.04	30.39

Table 7.4. Cost Savings and Energy Consumption Summary for Houses on Cold Summer Days

Cold Summer Day								
Home	Normal Operation Days				MPC Days			
	Cost, $	Peak Cost, $	Energy, kWh	Peak run time, %	Cost, $	Peak Cost, $	Energy, kWh	Peak run time, %
Home B down	1.53	1.51	8.23	22.5	0.38	0.13	6.12	1.96
Home D	2.81	1.52	22.44	38.44	0.27	0.2	1.74	4.99

7.3 Implemented AFDD Agent and Its Test Results

For field testing of the AFDD agent, five residential AC units among the 11 recruited houses were selected, but only four were used for implementation purposes. The four selected homes are all located in Norman, Oklahoma. The fifth home was dropped because its AC uses an ECM motor and, as noted in Section 6.3.3, the proposed AFDD algorithm was not intended for such a system. All tests were conducted during the cooling season in 2022. Some details of the homes and their AC units can be found in

Table 7.5. The full-scale Norman Test House served as the main test home where we could create low charge faults for testing purposes.

Table 7.5. Specifications of Tested Homes

Home	**Square footage**	**Tonnage**		**AC age**	**Test period**
		Outdoor unit	**Indoor unit**		
Lab-house	1658	3.5	4	0 – 1 years	July 14 – Sept. 20
Home A	2100	5	5	4 years	May 14 – Sept. 26
Home I	2450	5	5	8 – 15 years	May 14 – Sept. 26
Home J	2150	2.5	3	8 – 15 years	July 14 – Sept. 26

A smart thermostat, a node sensor for supply air temperature (SAT) measurement at a diffuser, and a smart power meter were installed in each of the four test homes. Because the node sensor came with the thermostat, it was linked to the latter during installation so that the SAT measurement can be accessed from the same database where the thermostat data were stored. Outdoor air temperature (T_{odb}) and humidity (ϕ_{oa}) measurements were obtained from Mesonet's database, which was shown to be reliable in a previous study (Wang, Tang, and Song 2020). The AFDD agent was built with the Python programming language, and a two-way communication platform was developed to query data from the smart thermostat database. Before using the data for training and/or by the AFDD, the on/off signal from the thermostat was used to filter out the transient data, defined as data from the first 10 minutes of each AC on cycle. Also, the heating signal from the thermostat was used to filter out noise in the data due to switching from cooling to heating mode by homeowners, either mistakenly or deliberately. Further disturbances such as internet disruptions and power outages leading to the thermostat going offline and not storing data were also eliminated when collecting data. The outdoor air dew point was then computed using CoolProp and used to divide the data into wet coil data and dry coil data, as described in Section 656.3. The appropriate dataset was subsequently used to carry out the AFDD test based on the algorithm described in Sections 6.3.2 and 6.3.3.

The implementation was done between May and September 2022, and was divided into two phases, with Phase 1 focused on checking for installation faults and Phase 2 focused on checking for degradation faults. Both Method 1 and Method 2, previously discussed, were also tested since the four homes were each equipped with a smart power meter. For Method 1, where enthalpy change is used as a fault detection index, baseline data of enthalpy measurements were utilized to train the model in Equation (6.3). On the other hand, for Method 2, an average of the indoor power consumption was computed and used as the indoor power baseline for Phase 2. The baseline data used for Phase 2 were collected while testing for installation faults in Phase 1 was taking place. Thus, Phase 2 did not begin until Phase 1 completed (regardless of whether a fault was detected).

7.3.1 Phase 1: Real-Time System Monitoring for Installation Faults

For this Phase 1, applying Method 1, the EPT data of enthalpy change were gathered and used as baseline data for training the model. Figure 7.11 shows the results generated by the AFDD agent in real time for the AC in the four test homes.

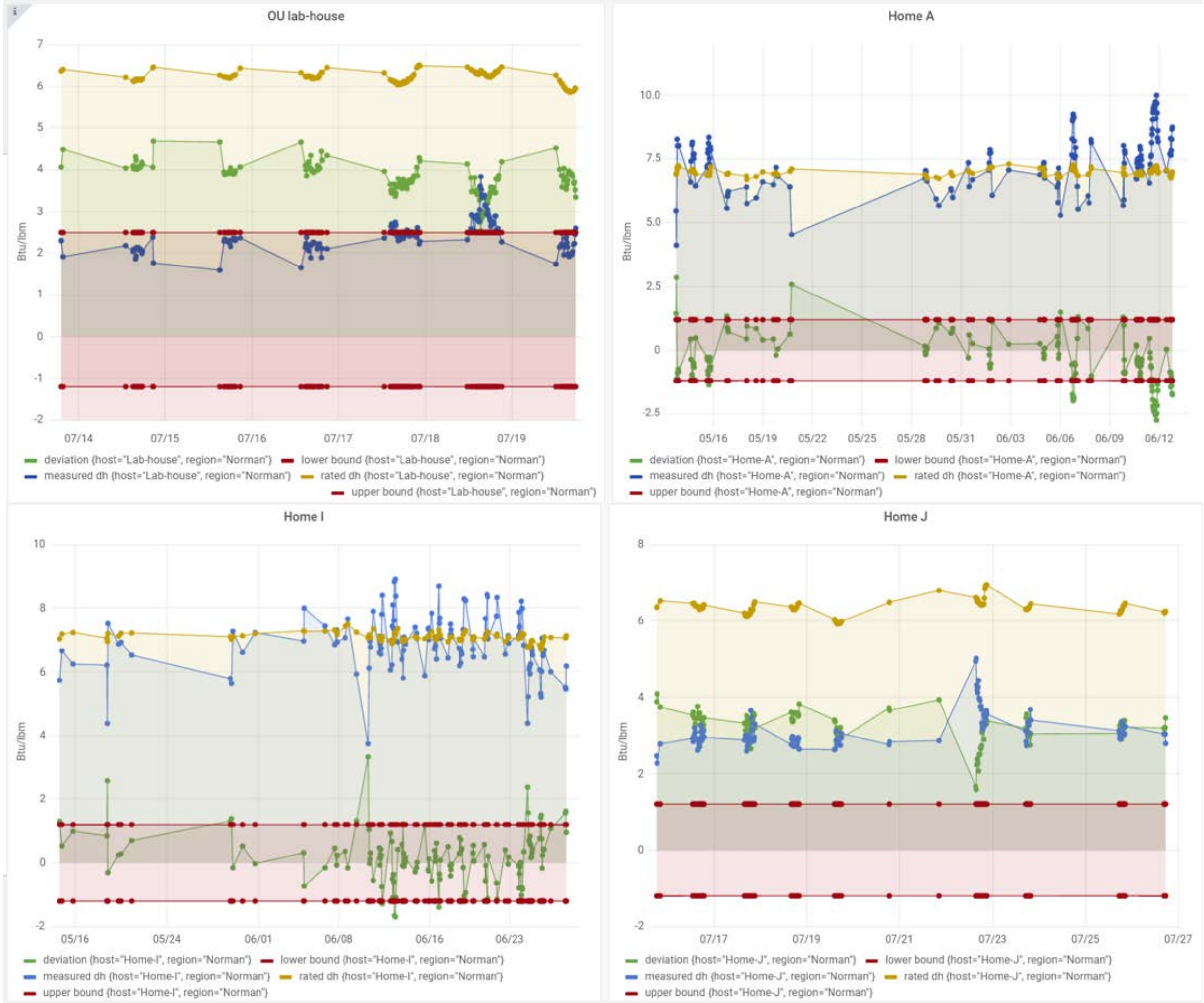


Figure 7.11. Real-time monitoring of enthalpy change across the evaporator for four test homes for Phase 1

For the Norman Test House, the results indicated that there was a large mismatch between the expected enthalpy change (labeled as rated dH in the figure) across the evaporator and the measured enthalpy change (labeled as measured dH in the figure). This mismatch is suspected to be mainly due to the indoor/outdoor unit mismatch, as can be seen from Table 7.5. With the indoor unit being bigger than the outdoor unit, the EPT rated indoor airflow rate, which was based on the outdoor unit size, would be less than the actual indoor airflow rate, thus leading to a lower enthalpy change across the evaporator. Similarly, Home J showed a large disparity between the measured dH and rated dH, thus resulting in deviations exceeding the threshold. For Home A, Figure 7.11 shows that deviations of the measured dH (i.e., enthalpy change) were close to the rated dH, with the deviations occasionally exceeding the threshold. But because such large deviations were not necessarily due to a fault and could be due to possible disturbances in the system such as transients, etc., the CUSUM algorithm was able to confirm that such deviations were in fact not caused by a fault. Thus, a possible false fault alarm was forestalled. Home I, again from Figure 7.11, had a similar trend as Home A. However, results for Home I reveal that its AC unit had a closer match to rating conditions than Home A. Furthermore, it can be observed from Figure 7.11 and subsequent figures that there were occasional large omissions of several days. For instance, for Home A in Figure 7.11, there was a jump from May 20, 2022, to May 28, 2022. This was due to disturbances from either a lack of internet connectivity or a power outage. As the proposed AFDD algorithm is IoT-based, whenever there is an internet or power interruption it disrupts the data collection process and leads to missing data, sometimes for long periods. To prevent such missing data from hurting the accuracy of the AFDD algorithm, such time periods are omitted during the AFDD process. This explains why the results shown have time gaps.

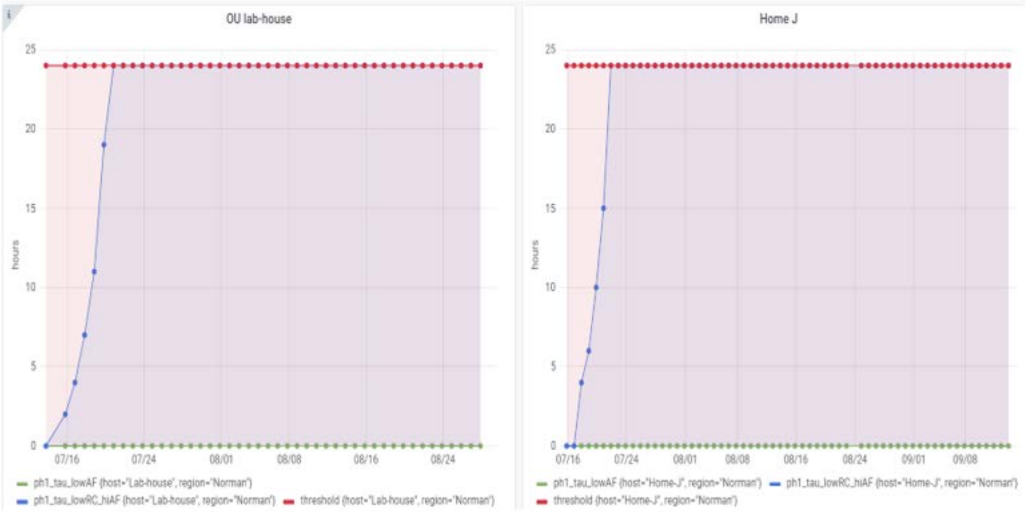


Figure 7.12. Detection of installation issues for two of the selected homes using Method 2

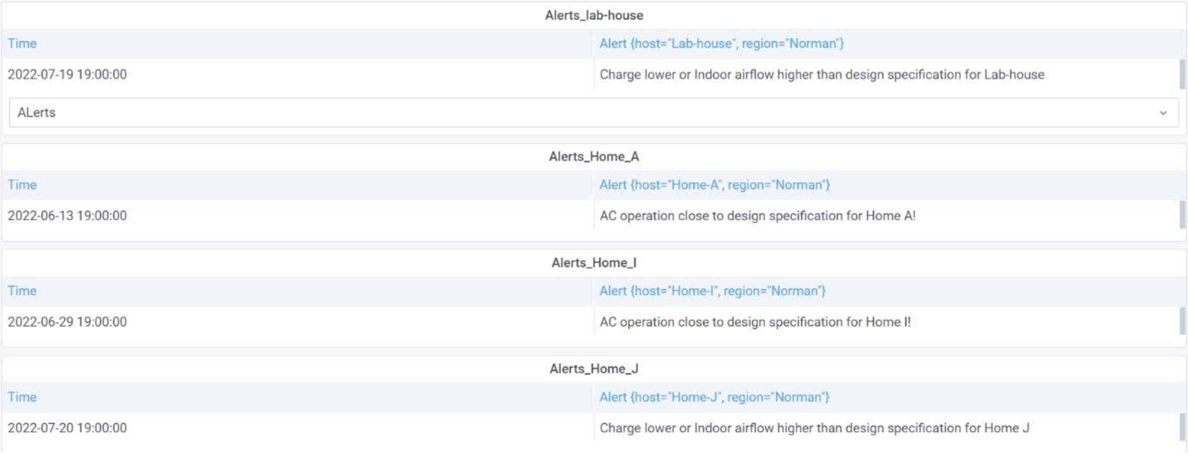
Figure 7.12 shows results from applying the CUSUM algorithm to the observed deviations in Figure 7.11. For the Norman Test House, due to sustained deviations that exceeded the deviation threshold, the variable associated with either the low charge or the high indoor airflow (ph1_tau_lowRC_hiAF) is seen to steadily rise until it reached the duration threshold about 7 days after the start of the installation check. Similar results were obtained for Home J, where the variable ph1_tau_lowRC_hiAF also had a steady rise, which led to the detection of a possible installation issue (suspected to be oversupply of indoor air) about 6 days after the commencement of the installation check. The observations for this home are similar to those for the Norman Test House. Like the latter, the AC unit installed in Home J also had the issue of indoor/outdoor unit mismatch, with the indoor unit being larger than the outdoor unit, as seen in Table 7.5. The unit mismatch could therefore be a possible cause of the observed oversupply of indoor air in both homes. For Home A and Home I, as the deviations in Figure 7.11 stayed below the deviation threshold for the most part, results from the CUSUM algorithm were not included for these homes. As seen from Figure 7.12, the variable (ph1_tau_lowRC_hiAF) associated with this suspected oversupply of indoor air was also associated with a low refrigerant charge. This was easily decoupled by applying Method 2, which first used indoor power for the fault check before using the enthalpy change (Δi) if indoor power fell below the deviation threshold, as described in Section 6.3.3.

Figure 7.13. Real-time monitoring of indoor power usage for all four test homes for Phase 1

The results obtained (Figure 7.13) indicate that the Norman Test House and Home J both had indoor power measurements that exceeded the baseline indoor power. For the former, the discrepancy in the indoor power from rated values was above 60%, while for the latter it was between 40% and 50%. This therefore confirms the suspicion of oversupply of indoor airflow. For Home J, a steady decline in the indoor power measurement was also observed. This observation was suspected to be caused by an indoor airflow reduction due most likely to a dirty filter. This was found to be true from results shown in Phase 2. In addition, Home A also showed indoor power about 14% higher than the rated indoor power, possibly suggesting an oversupply of indoor airflow. However, there was no mismatch between indoor and outdoor units and there were no other obvious installation faults in Home A after inspection. It can therefore be concluded that the oversupply of the power to the supply fan might be caused by an inefficient motor, or there were other loads in the circuit other than the indoor power fan. For Home I, the indoor power matched well with the rated conditions, staying below the threshold for the most part, except for the few outliers triggered most likely by sudden disturbances in the system, which the CUSUM algorithm successfully handled to prevent false fault alarms.

While the unit mismatch at the Norman Test House and Home J sometimes happens in residential HVAC systems and offers advantages such as increased cooling capacity and ventilation, especially when the indoor unit is bigger than the outdoor unit, manufacturers usually recommend matching the units.

Figure 7.14 shows the alert sent to each home based on the results of the installation check carried out by the AFDD algorithm. As seen from the figure, Home A and Home I both reported system operations close to rating conditions, while the Norman Test House and Home J both reported possible higher indoor airflow rates than rating conditions.



Alerts_lab-house

Time	Alert {host="Lab-house", region="Norman"}
2022-07-19 19:00:00	Charge lower or Indoor airflow higher than design specification for Lab-house

Alerts_Home_A

Time	Alert {host="Home-A", region="Norman"}
2022-06-13 19:00:00	AC operation close to design specification for Home A!

Alerts_Home_I

Time	Alert {host="Home-I", region="Norman"}
2022-06-29 19:00:00	AC operation close to design specification for Home I!

Alerts_Home_J

Time	Alert {host="Home-J", region="Norman"}
2022-07-20 19:00:00	Charge lower or Indoor airflow higher than design specification for Home J

Figure 7.14. Alerts sent for all four test homes

7.3.2 Phase 2: Real-Time System Monitoring for Degradation Faults

Upon completion of Phase 1, Phase 2 commenced, with the model retrained using baseline data collected during Phase 1. Results for this Phase 2 for all 4 test homes are shown in Figure 7.15. For the Norman Test House, a 30% low charge fault was purposely introduced on July 29, 2022. After introduction of the fault, the deviation between the measured and predicted enthalpy changes across the coil began to rise above the deviation threshold upper bound, and a few times fell below the threshold due to reduced load conditions (mainly outdoor air temperature). For the other homes, because no faults were introduced, and monitoring was done only for a period of 2–4 months, as shown in Table 7.5, a close match can be seen between the measured and, predicted enthalpy changes across the coil. Thus, the deviations mostly stayed below the deviation threshold and no faults were detected by the algorithm within this period. Faults were not introduced in any of these other homes to not cause inconvenience to the occupants. However, as the AFDD algorithm runs in real time and continuously with minimal computation cost, it is believed that if there is a deterioration in the AC system performance over time or if there is an occurrence of either of the two common faults in the future, the AFDD algorithm would be able to detect it.

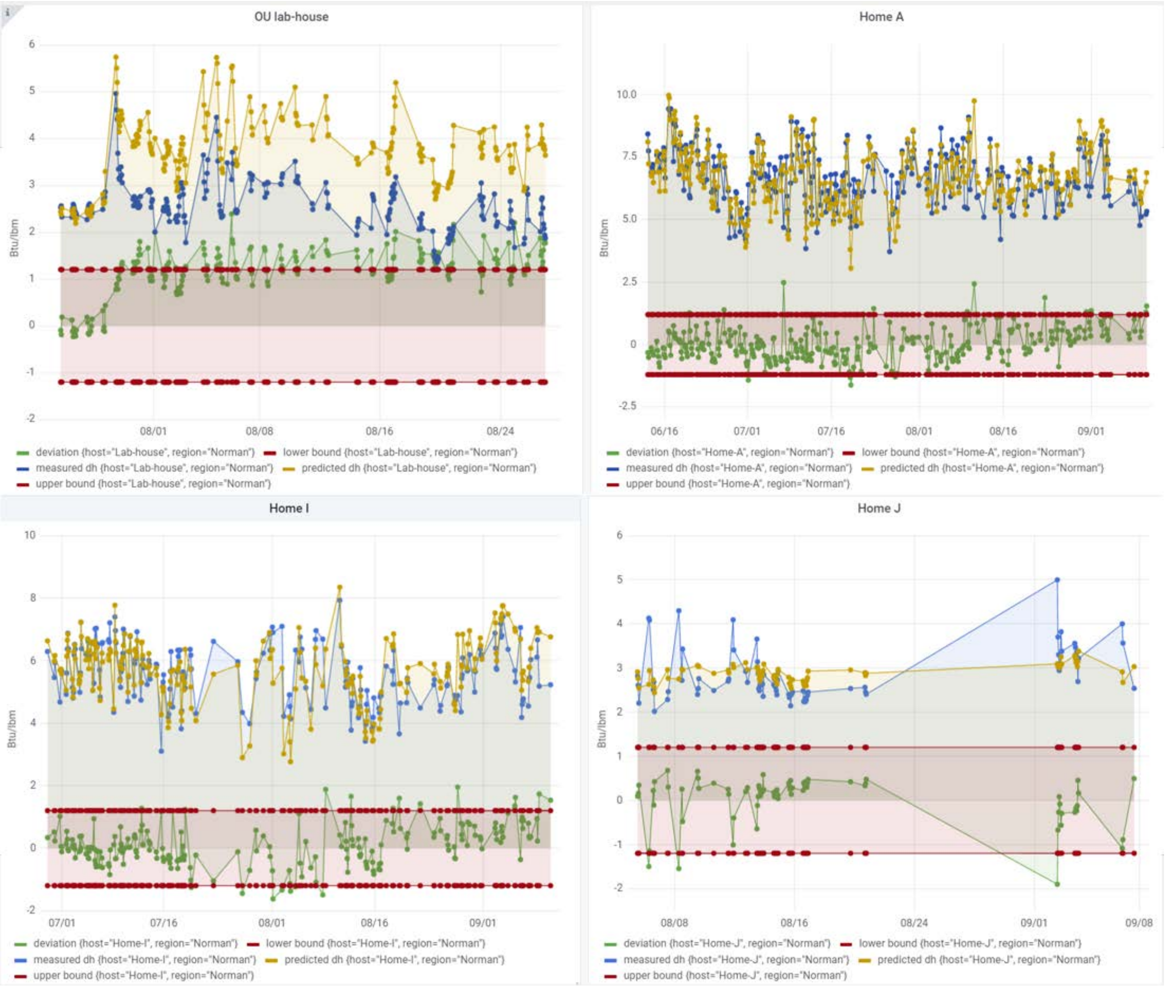


Figure 7.15. Real-time monitoring of enthalpy change across the evaporator for four test homes for Phase 2

Figure 7.16 shows results from the embedded CUSUM algorithm built into Method 1. As noted, due to the large deviations in the Norman Test House, the CUSUM algorithm was able to eventually detect the fault on August 27, 2022, which was about four weeks after the occurrence of the fault. Under higher load conditions, it is expected that the fault can be detected earlier. Also, as demonstrated in Section 6.6, if the actual SADP measurements were used, an earlier detection of the fault can be achieved. Figure 7.17 shows the alert that was sent from the AFDD algorithm for the detection of a low charge fault at the Norman Test House. This alert would also be sent to the homeowner's email to notify the homeowner of the presence of a fault in the system. Thus, using the proposed IoT-based AFDD algorithm, soft faults that would not have been discovered by homeowners can now be detected and corrected before they cause a major system breakdown.

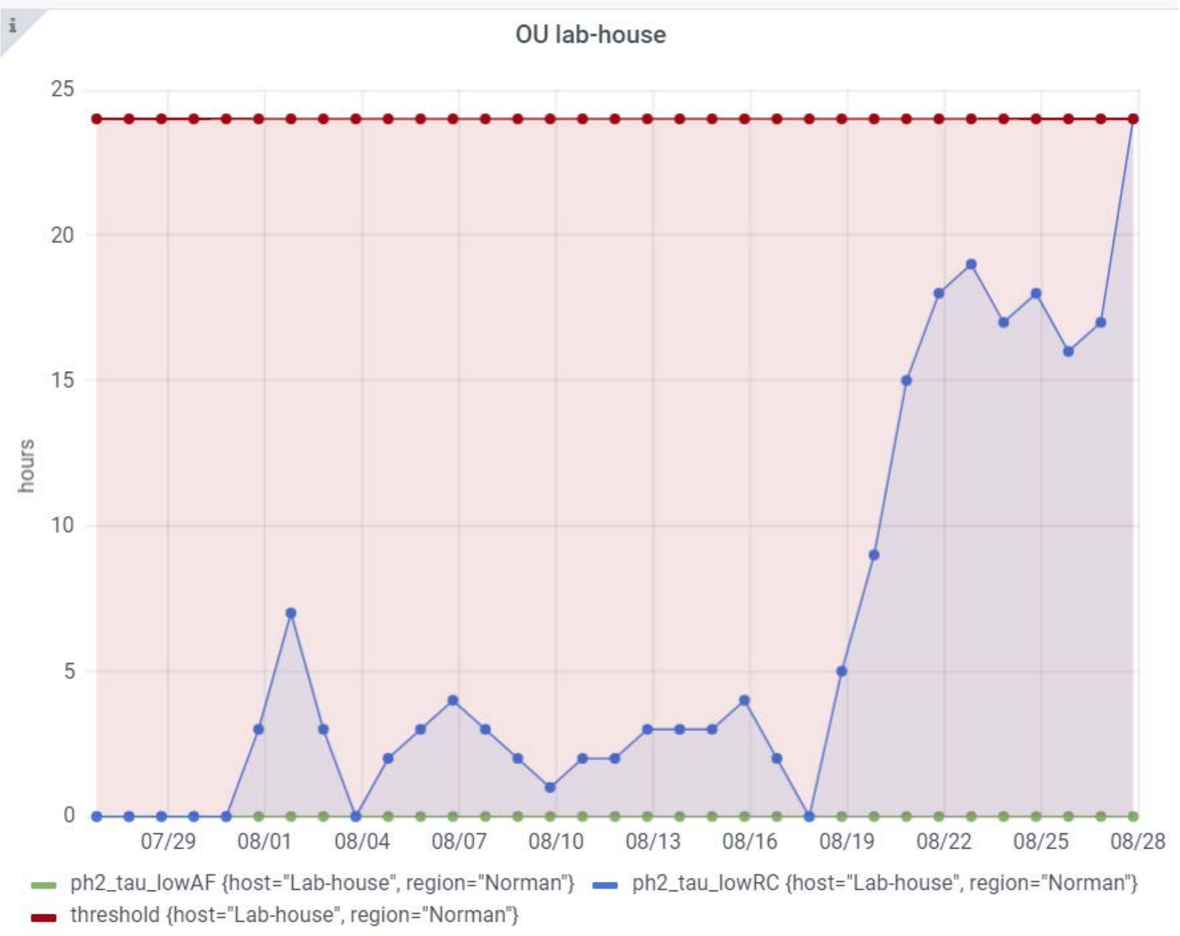
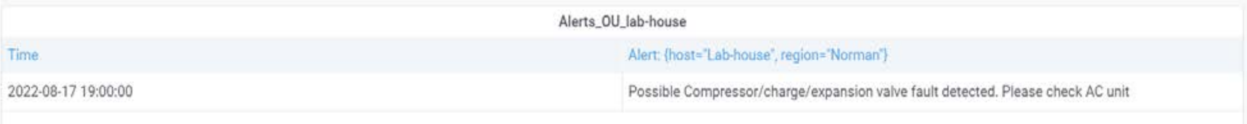


Figure 7.16. Detection of low charge fault for OU lab house



Alerts_OU_lab-house

Time	Alert: {host="Lab-house", region="Norman"}
2022-08-17 19:00:00	Possible Compressor/charge/expansion valve fault detected. Please check AC unit

Figure 7.17. Alert sent for detected low charge fault in OU lab house

Applying Method 2 to Phase 2, Figure 7.18 shows results for real-time monitoring of the indoor power measurements in the different units. Results for the Norman Test House reveal the inability of the indoor power to detect the low charge fault. This is in line with earlier observations that the indoor power was mainly influenced by airflow rates but independent of refrigerant charge levels. For Home A and Home I, as no fault was introduced, the results obtained were consistent with those shown in Figure 7.15, obtained from Method 1. However, Home I showed indoor power deviations (i.e., rise in indoor power above the baseline) closer to the lower threshold bound. This might have been caused by disturbances in the system and would be addressed by the CUSUM algorithm to forestall any possible false fault alarms, especially given that the deviations were not sustained. Home J on the other hand showed very few data points, a trend observed from Phase 1 specifically with this home. The paucity of data for this home was primarily due to frequent internet connectivity issues, thus leading to a lot of missing data. However, from the few data collected and analyzed, the results obtained show deviations that stayed clearly above the threshold band, with measured power being greater than estimated power. This observation is due to the impact of the suspected low airflow fault (mentioned in Section 7.3.1) on the baseline data collected during Phase 1. As the indoor power baseline was computed based on the average of the indoor power data collected during Phase 1, a decline in the power measurement led to a lower indoor power estimate at 560 W, as the unit was suspected to have a dirty filter when the AFDD was implemented. On August 17, 2022, the filter was replaced and the indoor power measurement returned to normal (630 W), as shown in Figure 7.19. This observation demonstrates the potential of the AFDD algorithm to detect simultaneous installation issues and degradation faults, as was the case for Home J, especially with Method 2. Method 1, which offers the advantage of a lower cost, was not able to detect this problem because the airflow rate changes were not large enough to be detected by Method 1. Also, the observations for Home J reveal the challenge of carrying out AFDD with a machine learning approach that relies on operation data for training the model, as was done in some previous studies. Unless such operation data can be guaranteed to be fault-free, it could lead to misdetection or false detection when carrying out AFDD. Therefore, it is suggested to carry out system commissioning before implementing the AFDD algorithm, if possible.

Figure 7.18. Real-time monitoring of indoor power usage for all four test homes for Phase 2

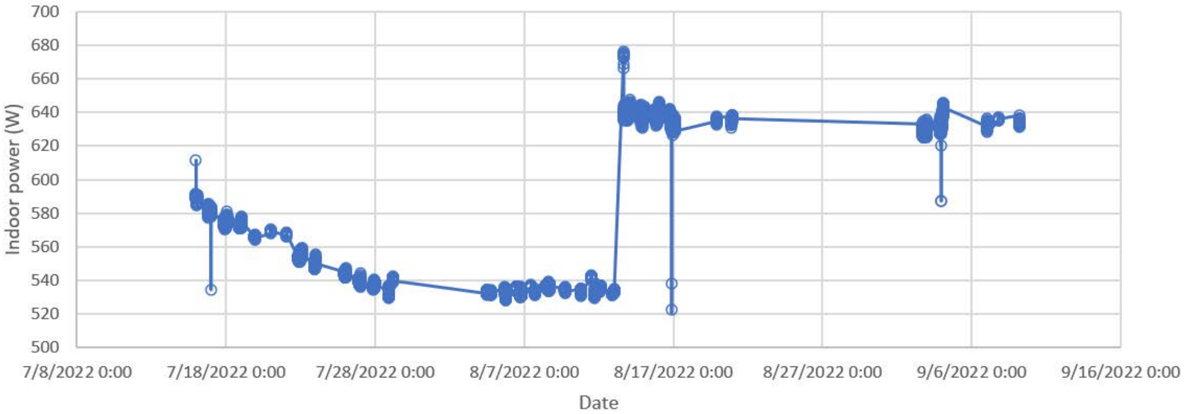


Figure 7.19. Indoor power measurements for Home J showing airflow rate degradation and subsequent filter replacement

8 Technology Transfer

Technology transfer activities have been carried out with a focus on technology dissemination through publications and technology commercialization through licensee identification. In addition, the project team has attended all webinars hosted by the Building America program to interact with other Building America teams for technology exchange.

8.1 Technology Dissemination

Results from this project have led to a total of eight publications, listed below, including six journal papers and two conference papers. Two additional journal papers are in preparation. In addition, the technology created in this project has been disseminated through community outreach. The Norman Test House has also been used as a living laboratory for community engagement to showcase the technology to local stakeholders including the state legislature, utility companies, residential communities, and tribal nations for future technology deployment. We have also engaged local newspapers and TV channels for technology dissemination. For instance, the technology was reported twice in the Journal Record, which is a daily business and legal newspaper based in Oklahoma City, Oklahoma, at the beginning of the project in August 2019 and at the conclusion of the project in December 2022.[1] Appendix D provides detailed coverage. In addition, Oklahoma News Channel 4 also reported on the project in December 2022.[2]

- Wang, J., Tang, C.Y., and Song, L. 2023. “Analysis of predicted mean vote-based model predictive control in building HVAC systems.” *Building and Environment*, in press.
- Li, D., Song, L., and Wang, G. 2022. “Energy and dehumidification performance investigation of different fan control modes of split residential air conditioners in hot and humid climates,” 2022 ASHRAE Summer Conference. *ASHRAE Transactions,* Vol. 128, Part 2.
- Wang, J., Tang, C.Y., and Song, L. 2022. “Analysis of pre-cooling optimization for residential buildings.” *Applied Energy,* Vol. 323: 119574.
- Wang, J., Jiang, Y., Tang, C.Y., and Song, L. 2022. “Development and validation of a second-order thermal network model for grid-interactive HVAC operation in residential buildings,” *Applied Energy*, Vol. 306 (B):118124.
- Wang, J., Tang, C.Y., and Song, L. 2020. “Design and analysis of optimal pre-cooling in residential buildings,” *Energy and Buildings*, Vol. 216 (1):109951.

---

[1] https://journalrecord.com/2022/11/23/old-house-becomes-center-of-energy-research-at-ou/

[2] https://kfor.com/news/local/ou-engineering-professor-converts-old-home-into-high-tech-lab-for-students

- Wang, J., Tang, C.Y., and Song, L. 2020. “Home envelope performance evaluation using a data driven method,” *ASHRAE Transactions*, Vol. 126 (1).
- Ejenakevwe, K. A., Wang, J., and Song, L. 2022. “Investigation of an IoT-Based approach for automated fault detection in residential HVAC systems,” 2022 ASHRAE Summer Conference, TO-22-C025, June 25–29, 2022, Toronto Canada.
- Ejenakevwe, K. A., and Song. L. 2021. “Review of fault detection and diagnosis studies on residential HVAC systems,” Proceedings of International Mechanical Engineering Congress and Exposition, IMECE2021-72745, November 1–5, 2021, virtual.

8.2 Technology Commercialization

Our proposed plan for commercialization involves licensing the technologies to potential licensees. Therefore, the first step is to file patents for the technologies so that the licensees’ rights can be protected. So far, with the Office of Technology Development at the University of Oklahoma, the following two patent applications have been filed, including one that was issued in March 2021:

- Monitoring system for residential HVAC systems, U.S. Patent No. 10,948,209, 2021; Inventors: L. Song and C.Y. Tang.
- System and method for residential HVAC control, U.S. Provisional Patent Filed in May 2021, Docket No. 5839.166; Inventors: L. Song, C.Y. Tang, J. Wang, and Y. Jiang.

For licensee recruitment, we have learned through the project that licensee candidates for the technologies need not be limited to smart thermostat manufacturers. This significantly increases the potential for successful commercialization because the technologies developed in this project—including the MPC agent and AFDD agent, which require additional measurements at wall surfaces, supply air, and/or power meters—can be deployed through third-party sensor manufacturers with much lower cost, better sensing capability, and higher accuracy. So far, we have identified two potential licensees, outlined below.

Using the cloud-based data management system developed in this project, we can easily retrieve data from different vendors of devices, such as ecobee’s thermostats and smart meters. We can then feed those data into an online platform using their APIs to conduct MPC-based optimization and AFDD with only labor cost for programmers. For example, the enthalpy change calculation used in the AFDD agent requires supply air temperature and humidity measurements. Working with ecobee as a partner in this project, we are limited to using ecobee’s node sensor (market price of $99 for two in one box) in field demonstration to collect the supply air temperature and estimate the supply air humidity measurement. This limitation reduces the sensitivity of the AFDD agent to more severe faults. Currently, we are working with a local startup company, Juniper Technology, to develop a Wi-Fi-enabled smart sensor series with temperature, humidity, and pressure measurements, along with different mounting configurations that enable easy installation on walls and insertion into AC ducts. The estimated cost of a complete product is $25 per sensor plus additional small data management sign-up fee if data management is desired by a buyer. A contract for prototype development is currently under review by the University of Oklahoma's purchasing department. If approved, the prototype development will be sponsored by an Oklahoma state fund. Therefore, Juniper Technology is identified as one of the two potential licensees.

The other potential licensee is our project partner, ecobee, who generously provided in-kind cost share by donating their time and the ecobee devices (20 thermostats and 40 node sensors). With ecobee, to provide technology updates and engagement, we organized two workshops with the ecobee data analysis group through Zoom during the project. One nondisclosure agreement has also been filed prior to the technology exchanges. Therefore, ecobee is identified as the other potential licensee for technology commercialization.

9 Conclusions

In this project, we successfully developed and extensively validated a learning-based home thermal model that enables an MPC-based optimization agent and an AFDD agent to run on top of it. The home thermal model was constructed using a two-node resistor-capacitor model and enriching it with solar and wind terms. To train the model parameters, two parameter identification methods—least squares and optimization—were proposed. The MPC-based optimization agent was designed based on the home thermal model and shown to be capable of optimizing residential HVAC operation, reducing operation costs while maintaining user comfort. In addition, the AFDD agent was designed and shown to be capable of detecting and diagnosing two prevalent residential AC faults, namely, the airflow reduction fault and incorrect refrigerant charge fault. The home thermal model, the MPC-based optimization agent, and the AFDD agent were all experimentally tested at the Norman Test House, Miami Test House, PNNL Test House A, and PNNL Test House B. They were further field tested at nine demonstration homes with real occupants.

To summarize the accomplishments of this project, we revisit the five project objectives and four research questions stated in Section 1 and briefly describe how they have been addressed and their broader implications.

- **Objective 1**: Validate the hypothesis that the home thermal model can be applied to accurately capture home thermal properties and predict space temperature dynamics.
- **Objective 2**: Validate the hypothesis that data collected from smart thermostats and smart meters can be applied to detect both the deficiencies in system design and construction, and faults during residency.
- **Objective 3**: Validate the hypothesis that real-time optimization can be achieved to balance space temperature and energy costs based on occupants' preferences, home thermal properties, weather forecasts, occupancy schedules, and TOU energy pricing.
- **Objective 4**: Demonstrate the cost and performance benefits of the technology at homes with different ages, sizes, and household incomes.
- **Objective 5**: Disseminate the technology through a public domain and/or on a website so that potential users, developers, and vendors can download pseudocode, publications, and presentations, and identify a minimum of one vendor as the technology licensee to commercialize the technology at the end of this funded project.
- **Question 1**: Can the thermal model effectively predict energy use of an HVAC system (extracted heat) with less than 15% error at 90% confidence?

- **Question 2**: Can the thermal model effectively predict the space air temperature within 2°F error at 90% confidence?
- **Question 3**: Can the MPC-based optimization agent be executed in real time with given weather forecasts and parameters identified automatically through data training?
- **Question 4**: To what extent (severity of AC faults) can the SYSTEM detect faults using data collected through smart thermostats?

Over the project's duration, we achieved each of these objectives and answered each of these questions. Objective 1 and Questions 1 and 2 on the effectiveness of the home thermal model are addressed in Sections 4.1, 4.2, and 4.3. We proved the hypothesis that the home thermal model can accurately capture home thermal properties and predict 12-hours-ahead space air temperature. Indeed, results from the Norman Test House and Miami Test House showed that a 1.96°F error at 95% confidence was attainable, surpassing the success criterion of 2°F error at 90% confidence. However, a 2.6°F error at 90% confidence was observed at PNNL Test House A, and an invalid model was obtained at PNNL Test House B. As explained in Section 4.3, these issues were caused by erroneous sensor measurements in the PNNL datasets. This suggests that sensor/data quality control is a must before engaging in parameter identification.

Objective 3 and Question 3 on performance of the MPC-based optimization agent are discussed in Sections 5.2 and 7.2. We established the hypothesis that an energy cost optimization problem—which takes into account home thermal properties, weather forecasts, users' preferred temperature ranges, occupancy schedules, and TOU energy pricing—can be formulated and solved in real time via an MPC-based optimization agent. The optimization problem is an integer linear program that may be solved using, for example, the CVX framework along with a GUROBI solver. The MPC-based optimization agent has been implemented on an online, cloud-based data management platform that collects data from smart thermostats and smart meters and enables remote, optimal control of AC units.

Objective 2 and Question 4 concerning viability of the fault detection methods are answered in Sections 6.5, 6.6, and 6.7. We have validated the hypothesis that data collected from smart thermostats and smart meters can be used by an AFDD agent to detect and diagnose airflow reduction faults and incorrect refrigerant charge faults in residential AC units. The AFDD agent does so using one of two methods. Method 1 compares actual enthalpy changes across an evaporator with baseline enthalpy changes and uses their differences to signal a possible fault. The method was found to be effective in detecting and diagnosing RCL and AFR faults when their severity reached 30%. However, when both faults occurred simultaneously, Method 2, which uses two fault indices—enthalpy changes and indoor fan power—is required. The method first uses indoor fan power measurements to diagnose AFR and then uses enthalpy changes to diagnose overcharge, undercharge, and occurrent faults with AFR. Through lab and field tests, the method was shown to be capable of catching less severe (around 15%) AFR faults.

Objective 4 concerning costs and benefits of the technology is discussed in Sections 7.1, 7.2, and 7.3. We illustrated the performance benefits of both the MPC-based optimization agent and AFDD agent by carrying out numerous field tests at homes with different ages, sizes, and household incomes. These homes included the Norman Test House, Miami Test House, PNNL Test House A, and PNNL Test House B, as well as nine demonstration homes with real occupants. The field tests have shown, for instance, that with the MPC-based optimization agent, up to 51% and 62% savings in energy costs can be achieved on hot and mild summer days, respectively. The significant amount of cost savings over a diverse collection of homes shows the promise of the technology.

Objective 5 on dissemination and commercialization of the technology is discussed in Sections 8.1 and 8.2. On the technology dissemination side, results from this project led to a total of eight publications, including six journal papers and two conference papers, and have been disseminated through community outreach and engagement, and reported through local newspapers and TV channels. On the technology commercialization side, results from this project led to one issued patent, one provisional patent, and the identification of two potential licensees of the technology.

By using the data readily available through smart thermostats, such as AC on/off signals, space air temperature, and humidity and node sensor temperature measurements—as well as cloud-based computing—the validated technologies in this project, including the learning-based thermal model, the optimization agent, and AFDD agent can improve residential AC performance efficiency and cost savings for homeowners with no additional hardware needed.

References

ASTM. 2019. *Standard Test Method for Calibration of Thermocouples by Comparison Techniques*. American Society for Testing and Material. PA, USA. https://www.astm.org/e0220-19.html.

Bell, Ian H., Sylvain Quoilin, Jorrit Wronski, and Vincent Lemort. 2013. “CoolProp: An open-source reference-quality thermophysical property library.” Abstract from ASME ORC 2nd International Seminar on ORC Power Systems, Rotterdam, Netherlands.

Brock, Fred V., Kenneth C. Crawford, Ronald L. Elliott, Gerrit W. Cuperus, Steven J. Stadler, Howard L. Johnson, and Michael D. Eilts. 1995. “The Oklahoma Mesonet: A Technical Overview.” *Journal of Atmospheric and Oceanic Technology.* 12:5–19. https://doi.org/10.1175/1520-0426(1995)012<0005:TOMATO>2.0.CO;2.

Brogan, William L. 1991. *Modern Control Theory*. United Kingdom: Prentice Hall.

Cericola, Rachel. 2015. Emerson ComfortGuard HVAC Monitoring Service Can Predict Problems. Emerson Electric. https://www.electronichouse.com/home-energy-management/emerson-comfortguard-hvac-monitoring-service-can-predict-problems/.

Cetin, Kristen Sara, and Catilyn Kallus. 2016. “Data-Driven Methodology for Energy and Peak Load Reduction of Residential HVAC Systems.” *Procedia Engineering.* 145:852–859. https://doi.org/10.1016/j.proeng.2016.04.205.

Chintala, Rohit, Jon Winkler, and Xin Jin. 2021. “Automated Fault Detection of Residential Air-Conditioning Systems Using Thermostat Drive Cycles.” *Energy and Buildings.* 236:1–11. https://doi.org/10.1016/j.enbuild.2020.110691.

DOE. 2021. “Programmable Thermostats.” https://www.energy.gov/energysaver/programmable-thermostats.

DOE EERE. 2016. “Overview of existing and future residential use cases for connected thermostats.”

Ejenakevwe, Kevwe Andrew, and Li Song. 2021. “Review of Fault Detection and Diagnosis Studies on Residential HVAC Systems.” Presented at ASME 2021 International Mechanical Engineering Congress and Exposition, online. https://asmedigitalcollection.asme.org/IMECE/proceedings/IMECE2021/85642/V08BT08A045/1132971?casa_token=aoZwlqzFQ4YAAAAA:mE0jfeoB3csymszz0mYgCJn90R99k-HeYkPAdx5s8zGPwchuViqePI2AkJIiV7UU95BinlE.

Emerson, E. 2021. Sensi Predict smart HVAC for Contractors. Emerson Electric. https://sensi.copeland.com/documents/sensi-predict-launch-guide-en-us-6158016.pdf.

Emerson, E. 2017. ComfortAlert for Residential Applications. Emerson Electric. https://climate.emerson.com/en-us/shop/1/copeland-comfortalert-for-residential-applications.

Emporia. 2021. “Emporia Vue: Gen 2 Home Energy Monitor.” Emporia. Littleton, CO, USA. https://shop.emporiaenergy.com/products/gen-2-emporia-vue-whole-home-energy-monitor.

Florida Power & Light. 2022. “Residential Time of Use Rate.” https://www.fpl.com/rates/time-of-use.html.

Gharghabi, Shaghayegh, Shima Imani, Anthony Bagnall, Amirali Darvishzadeh, and Eamonn Keogh. 2018. “Matrix Profile XII: MPdist: A Novel Time Series Distance Measure to Allow Data Mining in More Challenging Scenarios.” Presented at: The 2018 IEEE International Conference on Data Mining (ICDM). Singapore. https://doi.org/10.1109/ICDM.2018.00119.

Gowri, Krishnan, David W. Winiarski, and Ronald E. Jarnagin. 2009. *Infiltration Modeling Guidelines for Commercial Building Energy Analysis.* Richland, WA: Pacific Northwest National Laboratory. PNNL-18898. https://doi.org/10.2172/968203.

Grafana Labs. 2021. “Grafana: The Open Observability Platform.” 2021. Accessed January 24, 2023. https://grafana.com.

Grant, Michael, and Stephen Boyd. 2014. *CVX: Matlab Software for Disciplined Convex Programming*, Boston, MA: Springer.

Grigg, O. A., V. T. Farewell, and D. J. Spiegelhalter. 2003. “Use of Risk-Adjusted CUSUM and RSPRT Charts for Monitoring in Medical Contexts.” *Statistical Methods in Medical Research.* 12:147–170. https://doi.org/10.1177/096228020301200205.

Gurobi Optimization, LLC. 2021. *Gurobi Optimizer Reference Manual*. https://www.gurobi.com/wp-content/plugins/hd_documentations/documentation/9.0/refman.pdf.

Imani, Shima, Frank Madrid, Wei Ding, Scott Crouter, and Eamonn Keogh. 2018. “Matrix Profile XIII: Time Series Snippets: A New Primitive for Time Series Data

Mining." Presented at: The 2018 IEEE International Conference on Big Knowledge (ICBK). Singapore. https://doi.org/10.1109/ICBK.2018.00058.

InfluxDB. 2021. *InfluxDB OSS 2.0*. 2021. https://docs.influxdata.com/influxdb/v2.0/get-started.

Jain, Milan, Mridula Gupta, Amarjeet Singh, and Vikas Chandan. 2019. "Beyond Control: Enabling Smart Thermostats for Leakage Detection." *In Proceedings of the ACM Interactive, Mobile, Wearable, and Ubiquitous Technologies*. 3:1–21, https://dx.doi.org/10.1145/3314401.

Johra, Hicham, and Per Heiselberg. 2017. "Influence of Internal Thermal Mass on the Indoor Thermal Dynamics and Integration of Phase Change Materials in Furniture for Building Energy Storage: A Review." *Renewable and Sustainable Energy Reviews.* 69:19–32. https://doi.org/10.1016/j.rser.2016.11.145.

Juniper Technology. 2022. "Juniper Technology." Accessed September 6, 2022. https://www.junipertechnology.co/.

Kim, Jong-Jin, and Jin Woo Moon. 2009. "Impact of Insulation on Building Energy Consumption." Presented at: The Eleventh International IBPSA Conference, Building Simulation. Glasgow, Scotland. http://www.ibpsa.org/proceedings/BS2009/BS09_0674_680.pdf.

Kosny, Jan, T. Petrie, D. Gawin, P. Childs, A. Desjarlais, and J. Christian. 2001. *Thermal Mass-Energy Savings Potential in Residential Buildings*. Oakridge, TN: Pacific Northwest National Laboratory. www.buildingstudies.org/pdf/energy_studies/ORNL_Thermal-Mass_Energy_Savings_Potential_in_Residential_Buildings.pdf.

Kraft, Dieter. 1988. *A Software Package for Sequential Quadratic Programming*. Germany: Wiss. Berichtswesen d. DFVLR.

MathWorks. 2020. *MATLAB Manual.* https://www.mathworks.com/help/matlab/index.html?s_tid=hc_panel.

McPherson, Renee A., Christopher A. Fiebrich, Kenneth C. Crawford, James R. Kilby, David L. Grimsley, Janet E. Martinez, Jeffrey B. Basara, et al. 2007. "Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet." *Journal of Atmospheric and Oceanic Technology.* 24:301–321. https://doi.org/10.1175/JTECH1976.1.

McQuiston, Faye C., Jerald D. Parker, and Jeffrey D. Spitler. 2004. *Heating, Ventilating, and Air Conditioning: Analysis and Design*. United Kingdom: John Wiley & Sons.

Mueen, Abdullah, Sheng Zhing, Yan Zhu, Michael Yeh, Kaveh Kamgar, Krishnamurthy Viswanathan, Chetan Gupta, and Eamonn Keogh. 2022. "The Fastest Similarity Search Algorithm for Time Series Subsequences under Euclidean Distance." http://www.cs.unm.edu/~mueen/FastestSimilaritySearch.html.

Nelder, John A, and Roger Mead. 1965. "A Simplex Method for Function Minimization." *The Computer Journal.* 7:308–313. https://people.duke.edu/~hpgavin/cee201/Nelder+Mead-ComputerJournal-1965.pdf.

OG&E. 2022. "SmartHours – Because it's a SMART Time to Save on Energy." https://www.oge.com/wps/portal/ord/residential/pricing-options/smart-hours.

Ogunsola, Oluwaseyi T., and Li Song. 2015. "Application of a Simplified Thermal Network Model for Real-Time Thermal Load Estimation." *Energy and Buildings.* 96:309–318. https://doi.org/10.1016/j.enbuild.2015.03.044.

Ogunsola, Oluwaseyi T., Li Song, and Gang Wang. 2014. "Development and Validation of a Time-Series Model for Real-Time Thermal Load Estimation." *Energy and Buildings.* 76:440–449. https://doi.org/10.1016/j.enbuild.2014.02.075.

Rogers, A. P., F. Guo, and B. P. Rasmussen. 2019. "A Review of Fault Detection and Diagnosis Methods for Residential Air Conditioning Systems." *Building and Environment.* 161:1–12. https://doi.org/10.1016/j.buildenv.2019.106236.

Rotondo, Julia, Robert Johnson, Nancy Gonzalez, Alexandra Waranowski, Chris Badger, Nick Lange, Ethan Goldman, and Rebecca Foster. 2016. *Overview of Existing and Future Residential Use Cases for Connected Thermostats*. WA: DOE/EE-1508.

Schrijver, Alexander. 1998. *Theory of Linear and Integer Programming*. Germany: John Wiley & Sons.

Oklahoma Electric Cooperative. 2022. "Time of Use Rate." *Oklahoma Electric Cooperative* (blog). Accessed May 9, 2022. https://okcoop.org/time-of-use-rate/.

Waite, Michael B, and Sean M O'Brien. 2010. "Air Leakage: Difficulties in Measurement, Quantification and Energy Simulation." Presented at: Building Enclosure Science and Technology (BEST2) Conference, Portland, OR.

Wang, Junke, Yilin Jiang, Choon Yik Tang, and Li Song. 2022. "Development and Validation of a Second-Order Thermal Network Model for Residential Buildings." *Applied Energy.* 306:118124. https://doi.org/10.1016/j.apenergy.2021.118124.

Wang, Junke, Choon Yik Tang, and Li Song. 2020. "Design and Analysis of Optimal Pre-Cooling in Residential Buildings." *Energy and Buildings.* 216:109951. https://doi.org/10.1016/j.enbuild.2020.109951.

OpenWeather. 2018. "Weather API" Accessed January 27, 2023. https://openweathermap.org/api.

Werling, Eric. 2015. *Building America Research-to-Market Plan*. USA: Office of Energy Efficiency and Renewable Energy (EERE). DOE/EE-1285. https://doi.org/10.2172/1226785.

Yeh, Chin-Chia Michael, Nickolas Kavantzas, and Eamonn Keogh. 2017. "Matrix Profile VI: Meaningful Multidimensional Motif Discovery." Presented at: The 2017 IEEE International Conference on Data Mining (ICDM). New Orleans, LA. https://doi.org/10.1109/ICDM.2017.66.

Appendix

Appendix A. Lists of the Pi and Sensors Installed in the Norman Test Home

Table A.1. Specifications of All Measurements for the Data Acquisition System

Pi No.	Measurement	Channel	Pi No.	Measurement	Channel
Pi 1	T1	0	Pi 10	T25	0
	T2	1		T26	1
	T3	2		Open	2
	T4	3		Open	3
Pi 2	T5	0	Pi 11	T21	0
	T6	1		T22	1
	T7	2		T23	2
	T8	3		T24	3
Pi 3	T9	0	Pi 12	T28	0
	T10	1		T29	1
	T11	2		T30	2
	T12	3		T31	3
Pi 5	T15	0	Pi 13	T32	0
	T16	1		T33	1
	T17	2		T34	2
	T18	3		T35	3
Pi 6	Wind speed	0	Pi 14	T38	0
	Open	1		T39	1
	Open	2		T40	2
	Open	3		T41	3
Pi 7	T19	0	Pi 15	T43	0
	T20	1		T44	1

Pi No.	Measurement	Channel	Pi No.	Measurement	Channel
	T36	2		T45	2
	T37	3		T46	3
Pi 8	T13	0	Pi 16	T47	0
	T14	1		T48	1
	T27	2		T49	2
	T42	3		Open	3
Pi 4	Indoor frequency	0	Pi 9	Solar radiation	0
	Outdoor frequency	1		Airflow from return duct 1	1
	House frequency	2		Airflow from return duct 2	2
	Indoor power	3		Relative humidity from supply air duct	3
	Outdoor power	4		Air temperature from supply air duct	4
	House power	5		Relative humidity from return air duct	5
	Indoor pulse	6		Air temperature from return air duct	6
	Outdoor pulse	7		Total airflow rate	7
	House pulse	8		Open	8

Table A.2. Specifications of Location and Function of Sensors with Connected Pi for the Data Acquisition System

Pi Number	Location	Function
1	Between bedroom 1 and 2	Indoor air temperature measurements
2	Livingroom	Indoor air temperature measurements
3	Bedroom 3	Temperature measurements for indoor air and interior wall surface
4	Dining room	Power measurements
5	Between bedroom 1 and 2	Temperature measurements for the exterior wall surface and leaving and entering air of the outdoor unit
6	Bedroom 3	Wind speed measurement
7	Bedroom 3	Temperature measurements for the exterior wall surface and supply air from diffuser
8	Dining room	Temperature measurements for indoor and outdoor air, exterior wall surface, and supply air from diffuser
9	Bedroom 3	Solar radiation and duct flow rate measurements
10	Attic	Duct temperature measurement
11	Living room	Partition wall surface temperature measurement
12	Bedroom 1	Interior wall surface temperature measurement
13	Bedroom 3	Temperature measurements for interior wall surface and supply air from diffuser
14	Between bedroom 1 and 2	Air temperature measurements from supply diffusers
15	Between bedroom 1 and 2	Air temperature measurements from supply diffusers
16	Living room	Air temperature measurements from return diffusers

Appendix B. Model Parameter Estimation

(1) Identify τ_1 and τ_2 in Equation (3.11)

$$X_1\beta_1 = Y_1 \tag{B. 1}$$

The least-squares solution to Equation (B. 1) is

$$\hat{\beta}_1 = \begin{pmatrix} \hat{\beta}_1(1) \\ \hat{\beta}_1(2) \end{pmatrix} = (X_1^T X_1)^{-1} X_1^T Y_1 \tag{B. 2}$$

Thus,

$$\tau_1 = \Delta t / \hat{\beta}_1(1) \text{ and } \tau_2 = \Delta t / \hat{\beta}_1(2). \tag{B. 3}$$

where X_1 and Y_1 are known matrices; β_1 is the matrix to be identified; and $\hat{\beta}_1$ is the least-squares solution matrix.

$$X_1 = \begin{bmatrix} T_o(2) - T_{ie}(2) & T_{in}(2) - T_{ie}(2) \\ T_o(3) - T_{ie}(3) & T_{in}(3) - T_{ie}(3) \\ \vdots & \vdots \\ T_o(k-1) - T_{ie}(k-1) & T_{in}(k-1) - T_{ie}(k-1) \\ T_o(k) - T_{ie}(k) & T_{in}(k) - T_{ie}(k) \end{bmatrix}, \qquad \beta_1 = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix},$$

$$Y_1 = \begin{bmatrix} T_{ie}(2) - T_{ie}(1) \\ T_{ie}(3) - T_{ie}(2) \\ \vdots \\ T_{ie}(k-1) - T_{ie}(k-2) \\ T_{ie}(k) - T_{ie}(k-1) \end{bmatrix}. \tag{B. 4}$$

(2) Identify τ_3, b_1, b_2, a_1, a_2, a_3, Q_i, and Q_s in Equation (3.12)

$$X_2\beta_2 = Y_2 \tag{B. 5}$$

The least-squares solution to Equation (B. 5) is:

$$\hat{\beta}_2 = \begin{pmatrix} \hat{\beta}_2(1) \\ \hat{\beta}_2(2) \\ \hat{\beta}_2(3) \\ \hat{\beta}_2(4) \\ \hat{\beta}_2(5) \\ \hat{\beta}_2(6) \\ \hat{\beta}_2(7) \\ \hat{\beta}_2(8) \end{pmatrix} = (X_2^T X_2)^{-1} X_2^T Y_2 \tag{B. 6}$$

Thus,

$$\tau_3 = \Delta t/\hat{\beta}_2(1), b_1 = \hat{\beta}_2(2)/\hat{\beta}_2(1), b_2 = \hat{\beta}_2(3)/\hat{\beta}_2(1), a_1 = \hat{\beta}_2(4)/\hat{\beta}_2(1), a_2 = \hat{\beta}_2(5)/\hat{\beta}_2(1), a_3 = \hat{\beta}_2(6)/\hat{\beta}_2(1), Q_i = \hat{\beta}_2(7)/\hat{\beta}_2(1), \text{ and } Q_s = \hat{\beta}_2(8)/\hat{\beta}_2(1). \quad \text{(B. 7)}$$

where X_2 and Y_2 are known matrices; β_2 is the matrix needed to identify; and $\hat{\beta}_2$ is the least-squares solution matrix.

$$X_2 = \left[\begin{matrix} T_{ie}(2) - T_{in}(2) & (T_o(2) - T_{in}(2))W(2) & \big(T_o(2) - T_{in}(2) \\ T_{ie}(3) - T_{in}(3) & \big(T_o(3) - T_{in}(3)\big)W(3) & \big(T_o(3) - T_{in}(3) \\ \vdots & \vdots & \vdots \\ T_{ie}(k-1) - T_{in}(k-1) & \big(T_o(k-1) - T_{in}(k-1)\big)W(k-1) & \big(T_o(k-1) - T_{in}(k - \\ T_{ie}(k) - T_{in}(k) & \big(T_o(k) - T_{in}(k)\big)W(k) & \big(T_o(k) - T_{in}(k) \end{matrix}\right.$$

$$\beta_2 = \begin{bmatrix} \tau_3 \\ b_1 \\ b_2 \\ a_1 \\ a_2 \\ a_3 \\ Q_i \\ Q_s \end{bmatrix}, \quad \text{(B. 8)}$$

$$Y_2 = \begin{bmatrix} T_{in}(2) - T_{in}(1) \\ T_{in}(3) - T_{in}(2) \\ \vdots \\ T_{in}(k-1) - T_{in}(k-2) \\ T_{in}(k) - T_{in}(k-1) \end{bmatrix}.$$

Appendix C. Weather Conditions

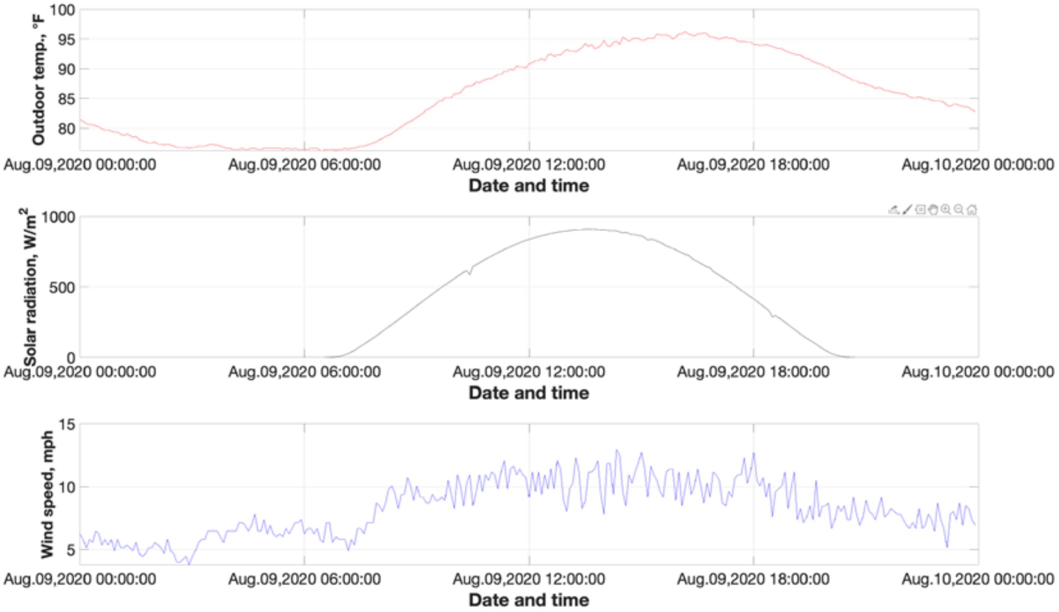


Figure C.1. Weather data on August 9, 2020

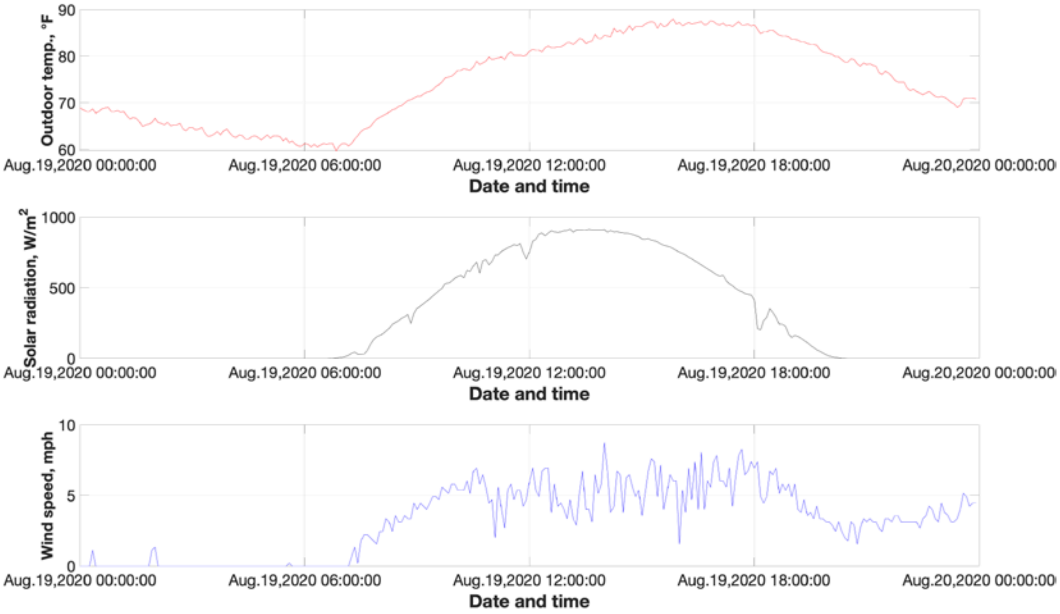


Figure C.2. Weather data on August 19, 2020

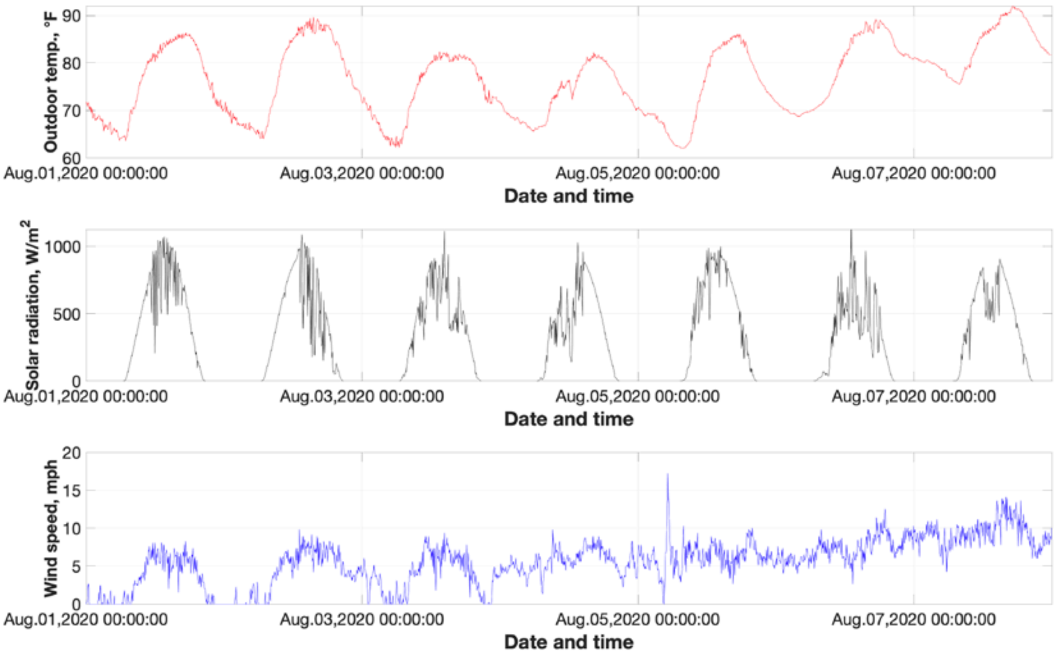
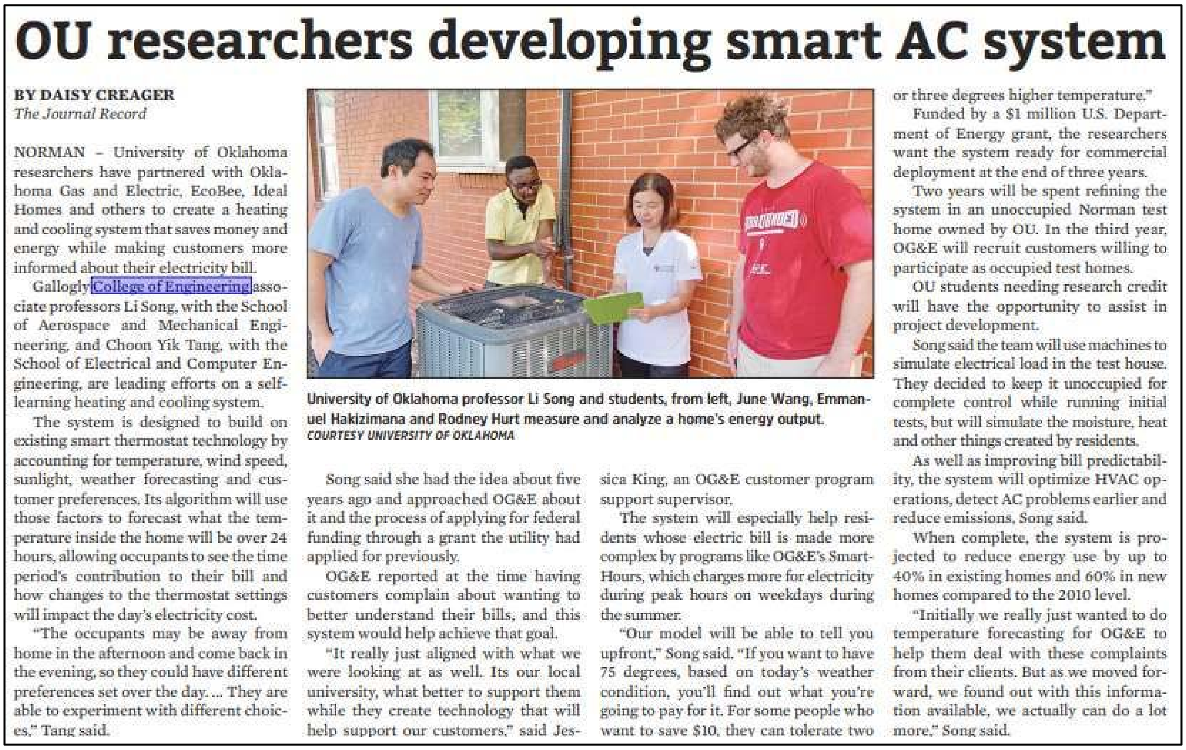


Figure C.3. Weather data from August 1–7, 2020

Appendix D. News Coverage of the Project for Technology Dissemination



OU researchers developing smart AC system

BY DAISY CREAGER
The Journal Record

NORMAN – University of Oklahoma researchers have partnered with Oklahoma Gas and Electric, EcoBee, Ideal Homes and others to create a heating and cooling system that saves money and energy while making customers more informed about their electricity bill.

Gallogly College of Engineering associate professors Li Song, with the School of Aerospace and Mechanical Engineering, and Choon Yik Tang, with the School of Electrical and Computer Engineering, are leading efforts on a self-learning heating and cooling system.

The system is designed to build on existing smart thermostat technology by accounting for temperature, wind speed, sunlight, weather forecasting and customer preferences. Its algorithm will use those factors to forecast what the temperature inside the home will be over 24 hours, allowing occupants to see the time period's contribution to their bill and how changes to the thermostat settings will impact the day's electricity cost.

"The occupants may be away from home in the afternoon and come back in the evening, so they could have different preferences set over the day. ... They are able to experiment with different choices," Tang said.

University of Oklahoma professor Li Song and students, from left, June Wang, Emmanuel Hakizimana and Rodney Hurt measure and analyze a home's energy output.
COURTESY UNIVERSITY OF OKLAHOMA

Song said she had the idea about five years ago and approached OG&E about it and the process of applying for federal funding through a grant the utility had applied for previously.

OG&E reported at the time having customers complain about wanting to better understand their bills, and this system would help achieve that goal.

"It really just aligned with what we were looking at as well. Its our local university, what better to support them while they create technology that will help support our customers," said Jessica King, an OG&E customer program support supervisor.

The system will especially help residents whose electric bill is made more complex by programs like OG&E's SmartHours, which charges more for electricity during peak hours on weekdays during the summer.

"Our model will be able to tell you upfront," Song said. "If you want to have 75 degrees, based on today's weather condition, you'll find out what you're going to pay for it. For some people who want to save $10, they can tolerate two or three degrees higher temperature."

Funded by a $1 million U.S. Department of Energy grant, the researchers want the system ready for commercial deployment at the end of three years.

Two years will be spent refining the system in an unoccupied Norman test home owned by OU. In the third year, OG&E will recruit customers willing to participate as occupied test homes.

OU students needing research credit will have the opportunity to assist in project development.

Song said the team will use machines to simulate electrical load in the test house. They decided to keep it unoccupied for complete control while running initial tests, but will simulate the moisture, heat and other things created by residents.

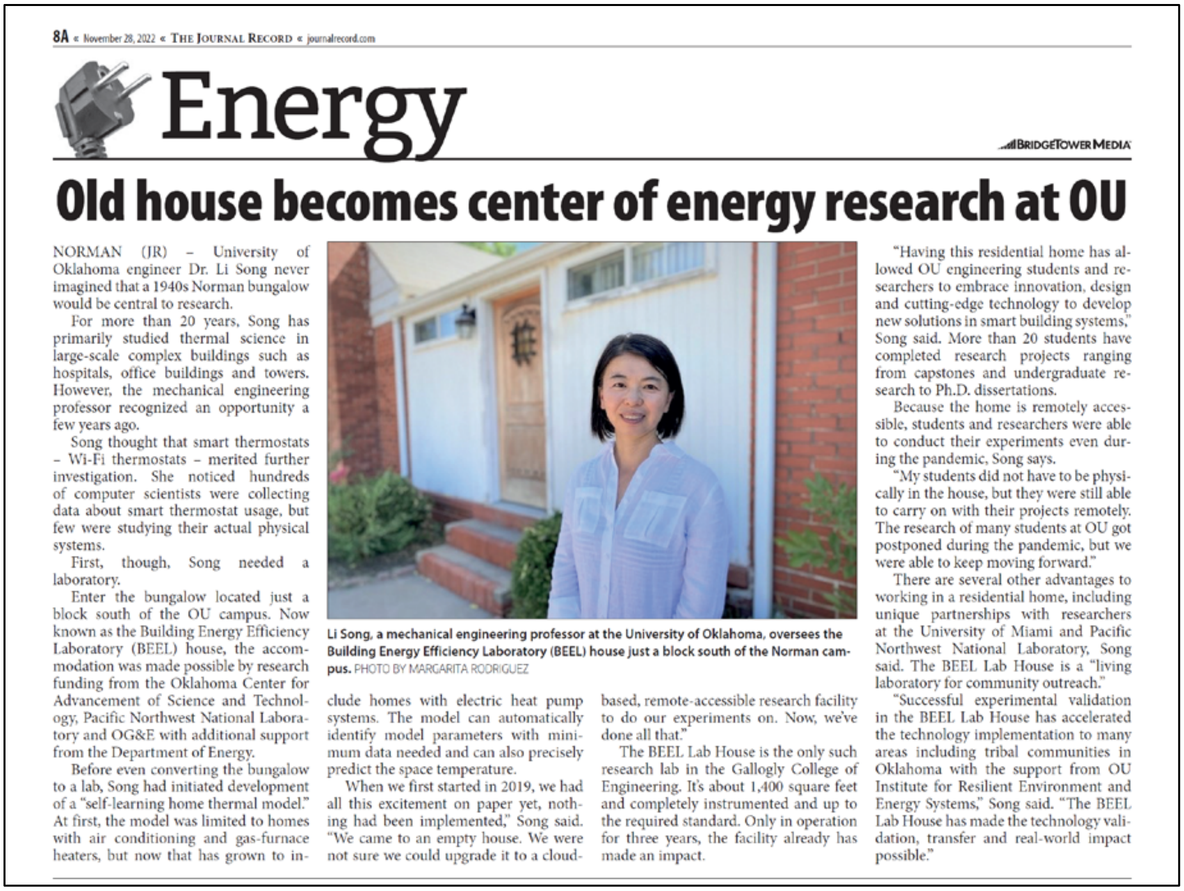
As well as improving bill predictability, the system will optimize HVAC operations, detect AC problems earlier and reduce emissions, Song said.

When complete, the system is projected to reduce energy use by up to 40% in existing homes and 60% in new homes compared to the 2010 level.

"Initially we really just wanted to do temperature forecasting for OG&E to help them deal with these complaints from their clients. But as we moved forward, we found out with this information available, we actually can do a lot more," Song said.

Figure D.1. Newspaper report about the project in August 2019

Article from: https://journalrecord.com/2019/08/ou-researchers-developing-smart-ac-system-that-could-lower-bills-by-40/



8A « November 28, 2022 « THE JOURNAL RECORD « journalrecord.com

Energy

BRIDGETOWER MEDIA

Old house becomes center of energy research at OU

NORMAN (JR) – University of Oklahoma engineer Dr. Li Song never imagined that a 1940s Norman bungalow would be central to research.

For more than 20 years, Song has primarily studied thermal science in large-scale complex buildings such as hospitals, office buildings and towers. However, the mechanical engineering professor recognized an opportunity a few years ago.

Song thought that smart thermostats – Wi-Fi thermostats – merited further investigation. She noticed hundreds of computer scientists were collecting data about smart thermostat usage, but few were studying their actual physical systems.

First, though, Song needed a laboratory.

Enter the bungalow located just a block south of the OU campus. Now known as the Building Energy Efficiency Laboratory (BEEL) house, the accommodation was made possible by research funding from the Oklahoma Center for Advancement of Science and Technology, Pacific Northwest National Laboratory and OG&E with additional support from the Department of Energy.

Before even converting the bungalow to a lab, Song had initiated development of a "self-learning home thermal model." At first, the model was limited to homes with air conditioning and gas-furnace heaters, but now that has grown to include homes with electric heat pump systems. The model can automatically identify model parameters with minimum data needed and can also precisely predict the space temperature.

When we first started in 2019, we had all this excitement on paper yet, nothing had been implemented," Song said. "We came to an empty house. We were not sure we could upgrade it to a cloud-based, remote-accessible research facility to do our experiments on. Now, we've done all that."

The BEEL Lab House is the only such research lab in the Gallogly College of Engineering. It's about 1,400 square feet and completely instrumented and up to the required standard. Only in operation for three years, the facility already has made an impact.

"Having this residential home has allowed OU engineering students and researchers to embrace innovation, design and cutting-edge technology to develop new solutions in smart building systems," Song said. More than 20 students have completed research projects ranging from capstones and undergraduate research to Ph.D. dissertations.

Because the home is remotely accessible, students and researchers were able to conduct their experiments even during the pandemic, Song says.

"My students did not have to be physically in the house, but they were still able to carry on with their projects remotely. The research of many students at OU got postponed during the pandemic, but we were able to keep moving forward."

There are several other advantages to working in a residential home, including unique partnerships with researchers at the University of Miami and Pacific Northwest National Laboratory, Song said. The BEEL Lab House is a "living laboratory for community outreach."

"Successful experimental validation in the BEEL Lab House has accelerated the technology implementation to many areas including tribal communities in Oklahoma with the support from OU Institute for Resilient Environment and Energy Systems," Song said. "The BEEL Lab House has made the technology validation, transfer and real-world impact possible."

Li Song, a mechanical engineering professor at the University of Oklahoma, oversees the Building Energy Efficiency Laboratory (BEEL) house just a block south of the Norman campus. PHOTO BY MARGARITA RODRIGUEZ

Figure D.2. Newspaper report about the project in December 2022

Available from: https://journalrecord.com/2022/11/old-house-becomes-center-of-energy-research-at-ou/

U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

For more information, visit: buildingamerica.gov

DOE/GO-102024-5884 • May 2024