Residential Heat Pump Efficiency Rating Representativeness Project PHASE 2

OCTOBER 2024

Contents

List of Tables

List of Figures

List of Figures, continued

Acknowledgements

This report reflects the invaluable contributions of multiple individuals and organizations. This project is the result of a collaboration between a diverse range of heat pump market actors including regional energy efficiency organizations, energy efficiency program administrators, energy efficiency advocates and manufacturers. Northeast Energy Efficiency Partnerships (NEEP) enabled the partnership by serving as a facilitator, contributor, and fiscal agent.

Direct project funding contributions were provided by Air-Conditioning, Heating and Refrigeration Institute (AHRI), BC Hydro, CLASP, ComEd, New York State Energy Research and Development Authority (NYSERDA), Natural Resources Canada (NRCan), Northwest Energy Efficiency Alliance (NEEA), Pacific Gas & Electric (PG&E), Southern Cal Edison, and Xcel Energy.

With co-funding from this diverse range of organizations, NEEP facilitated the hiring of DNV as lead project researcher, Underwriters Laboratory (UL) to conduct lab testing, and Bruce Harley Energy to provide technical assistance.

NEEP would like to recognize the report's lead author from DNV, Jennifer McWilliams and DNV scope project manager Vivek Jaiswal. Lab testing was led by Mark Baines and Titus Mowry from UL Solutions. In-field monitoring and data analysis was led by David Yuill, Jim Butler, and Yuxuan Chen from University of Nebraska Lincoln (UNL), and Chris Williams from DNV. Bruce Harley Energy provided technical support and analysis.

Representatives from the funding organizations served as advisors throughout the project. NEEP would also like to recognize and thank members of the advisory committee for their participation in reviewing this report and providing input into the creation of this document.

Several leading heat pump manufacturers supported the project directly by assisting the lab testing, without which the project would not have been possible

Dave Lis, NEEP's Director of Technology Market Transformation, served as project manager. Formatting and edits were provided by Lisa Cascio, Director of Communications and External Relations, Krysia Wazny McClain of Owl Eye Edits, and Marianne Michalakis of designMind.

About NEEP

NEEP was founded in 1996 as a non-profit whose mission is to serve the Northeast and Mid-Atlantic to accelerate regional collaboration to promote advanced energy efficiency and related solutions in homes, buildings, industry, and communities. Our vision is that the region's homes, buildings, and communities are transformed into efficient, affordable, low-carbon resilient places to live, work, and play.

Disclaimer: NEEP verified the data used for this report to the best of our ability. This paper reflects the opinion and judgments of the NEEP staff and does not necessarily reflect those of NEEP Board members, NEEP Sponsors, or project participants and funders.

Section 3: INTRODUCTION

This report details the second phase of the residential heat pump efficiency rating representativeness study conducted by DNV and UNL (the research team) for Northeast Energy Efficiency Partnerships (NEEP). The study is part of an effort to modernize lab-based test procedures and energy efficiency ratings used to predict the in-field efficiency of heat pumps by ensuring that ratings are representative of in-field efficiency. Lab-based performance ratings provide critical information to the market for the development, sales and selection of heat pump systems, so it is essential that they are representative of real-world performance. With the emergence of variable speed heat pumps that rely on proprietary controls to manage the systems operation and efficiency, the representativeness of existing test procedures that determine performance under fixed speed conditions have come under increased scrutiny. Energy efficiency programs also have a keen interest in representative performance ratings to ensure incentive investments are achieving expected savings. Representative test procedures and ratings enable the adoption of high efficiency heat pumps across North America, commonly identified as an essential step toward decarbonizing homes and businesses.

3.1: Study objectives

This study has three primary aims:

1. Build a robust set of rigorous and well-controlled in-the-field measurement data to enable in-depth comparisons of the field data to the ratings produced by the two major laboratory test procedures specified above (SPE-07 and M1) for a set of ducted and ductless heat pumps. The Department of Energy (DOE) "Appendix M1" is the governing document that the study follows in the lab. As noted above, it is harmonized generally with AHRI 210/240 with field test conditions the same or very close to conditions set by Appendix M1.
2. Use the data to inform policy on the value of load-based testing[1] (SPE-07) relative to static testing[2] (M1).
3. Determine the shortcomings or differences that diminish the relevance, or representativeness, of the lab test results compared with measured field performance.

In this second phase, the research team carefully packed and sent the six heat pumps used in the first phase field testing to Underwriters Laboratory (UL) in Plano, Texas for laboratory testing according to the two procedures:

1. Department of Energy (DOE) Code of Federal Regulations Title 10, Chapter II, Subchapter D, Part 430, Subpart B, Appendix M1 to Subpart B of Part 430[3]: Uniform Test Method for Measuring the Energy

[1] Heat pump performance testing that imposes heating/cooling loads on heat pump systems and allows native controls to determine response. Compressor speeds are not fixed or locked.

[2] Heat pump performance testing that utilizes fixed compressor speeds

[3] Appendix M1 is harmonized generally with AHRI 210/240, but M1 is the governing document that is followed in the lab.

Consumption of Central Air Conditioners and Heat Pumps (January 2017) (hereafter M1). This standard is harmonized generally with Air-conditioning, Heating and Refrigeration Institute (AHRI) 210/240-2023, used for regulation in Canada and the USA.

2. Canadian Standards Association (CSA) SPE-07:23[4] (hereafter SPE-07), which is being considered for regulatory adoption in Canada.

The results of the analysis comparing the laboratory data to the field data are detailed in this report.

3.2: Field testing overview

The field-testing work consisted of testing six heat pumps in three nearly identical and well calibrated mobile homes in Lincoln, Nebraska. The instrumentation was of lab grade quality and resolution including energy, outdoor conditions climate, indoor controls, humidity, refrigerant flow etc. Indoor conditions and loads were created to reflect an occupied house. Each house had one ducted and one ductless heat pump which were used on alternate weeks to maximize data collection with only three homes. The data was collected from August 2022 through February 2023 to provide ample variation across a range of outdoor conditions to provide confidence in the applicability of the results for both cooling and heating conditions. Greater detail is available in the Residential Heat Pump Efficiency Rating Representativeness Project Phase 1 report[5].

Because of set-up and commissioning delays, the cooling test period was shorter than planned. The large number of hotter-than-normal days allowed us to extend the end of the cooling test period past mid-October. We are confident that the test period included ample variation across heating and cooling seasons with the exception of one unit (D) which faced commissioning challenges both in the field and in the lab. Because these challenges led to highly questionable test results, system D has been removed from the analysis. A summary of the testing period dates and number of testing days are shown in Table 3-1. Detailed descriptions are further given in Sections 2.5.1 and 2.5.5 of the Phase 1 report. Internal loads were simulated as described in Section 2.4.5 of the Phase 1 report.

Table 3-1. ***Summary of testing periods***

Heating or Cooling	Start	End	# of days
Cooling	August 19, 2022	October 24, 2022	67
Heating	October 25, 2022	February 28, 2023	127

[4] The first edition of *Load-based and climate-specific testing and rating procedures for heat pumps and air conditioners*. It supersedes the document CSA EXP-07:19 of the same name.

[5] Add weblink once created

The research objectives of the field-testing phase were never to characterize the heat pump performance for a particular year in a particular location; rather, it was to:

1. Develop performance profiles of the heat pump units over wide temperature ranges and realistic loads and then apply the performance profiles to normalized weather bins discussed in section 5;
2. Compare the heat pump systems' normalized seasonal performance profiles to those derived in laboratory testing based on the CSA SPE-07 and DOE-Appendix M1 lab performance test and rating protocols. This is discussed in section 7.3.
3. Establish recommendations for lab grade field testing of heat pump performance that could be duplicated by future studies.

3.3: Lab testing overview

After field testing, the five heat pump units were carefully decommissioned, packaged and shipped to the UL testing lab in Plano, Texas where the SPE-07 and M1 tests were performed. In preparation to ship the units, the refrigerant was evacuated and weighed so the same refrigerant mass could be added when the units were assembled in the lab testing facility. To maintain the same refrigerant line length and volume, the mass flow sensors remained in the refrigerant lines and were shipped along with the heat pumps to the testing labs. Further description of the heat pump decommissioning and packing process is found in Section 2.6 of the Phase 1 report.

Alignment of the test homes to the thermal conductance and thermal capacitance assumed in the SPE-07 testing algorithm was sought, but it was not possible to entirely match them. Details of these efforts are described in the Phase 1 report. The load lines[6] used in the SPE-07 testing procedure were modified to align with the field conditions described in Section 2.2.2 of the Phase 1 report for each heat pump to ensure comparability between the actual houses where field tests were performed and the house response assumed as part of the SPE-07 test procedure. Despite these efforts, the loads experienced by the field homes differed from those in the lab. We corrected for this difference as described in the field data analysis Section 5.2.

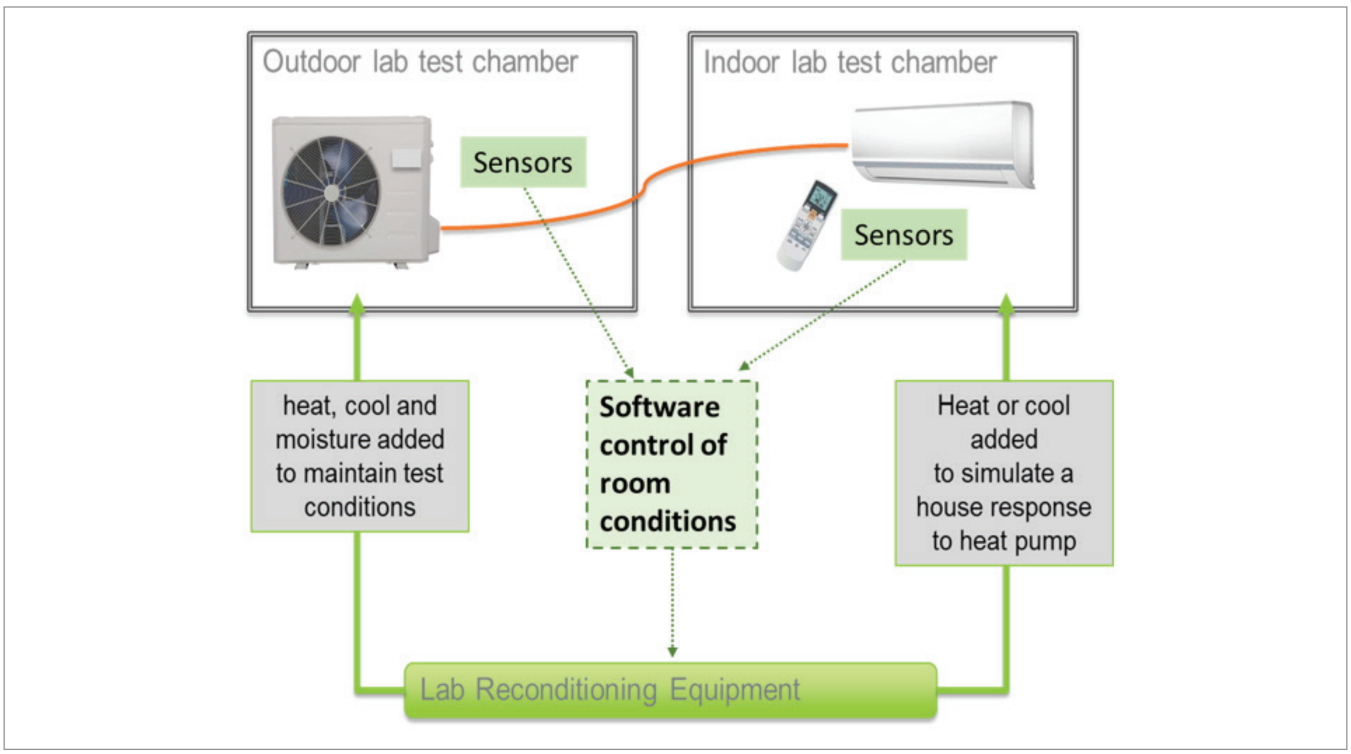
Because there wasn't enough data to get coherent capacitance values from the field data, we used the capacitance assumptions from SPE-07 without modification. These values should be reasonable based on the measured mobile home capacitance values described in Section 2.2.3 of the Phase 1 report which were close to the SPE-07 assumptions.

[6] A load line is a linear relationship between the building load and the outdoor temperature. The SPE-07 test procedure has an assumed load line that is used as part of the algorithm to adjust the indoor test chamber temperature based on the outdoor temperature of the test and the capacity produced by the equipment under test.

Section 4: LAB TESTING ACTIVITIES

At the laboratory, SPE-07 and M1 tests were performed on each of the five heat pumps. The lab facility consists of an indoor chamber and an outdoor chamber, with one simulating the indoor condition and another the outdoor condition. In the M1 test, the conditions in each test chamber are static over the duration of each testing condition. The SPE-07 standard consists of a load-based test where the indoor lab test chamber temperature responds to the capacity of the system under test rather than being set at a static temperature. Figure 4-1 shows a diagram of the laboratory testing chambers for the SPE-07 test. The tests are run until the coefficient of performance (COP) converges or an elapsed time limit is reached.

Figure 4-1. ***Diagram of CSA SPE-07 testing apparatus***



Graphic by Parveen Dhillon, Purdue University

4.1: Lab testing descriptions

All the systems were tested in the lab after the completion of the field test. The SPE-07 test included nine cooling conditions (five dry and four humid) and seven heating conditions (six continental and one marine) for a total of 16 testing conditions. The indoor and outdoor conditions for the SPE-07 laboratory cooling and heating tests are shown in Table 4-1 and Table 4-2 respectively. All tests were performed except for one on system E (HL_C was not conducted due to a lab equipment issue when the outdoor chamber reached -21 °F and the reconditioning equipment faulted). All tests were performed three times on systems C and F to evaluate the repeatability of the SPE-07 test procedure following the changes from its earlier version (EXP-07) that were made in 2022 to improve repeatability.

Table 4-1. ***SPE-07 test room conditions for cooling (SCOPC) test series***

Test	Humid Test Conditions[ii]			Dry Test Conditions		
	Outdoor dry-bulb temperature, °F	Indoor dry-bulb temperature,[iii] °F	Indoor humidity ratio[iv]	Outdoor dry-bulb temperature, °F	Indoor dry-bulb temperature,[iii] °F	Indoor humidity ratio[iv]
CA[i]	N/A	74	0.010	113	79	0.0087
CB	104			104		
CC	95			95		
CD	86			86		
CE	77			77		

i.) Temperature "CA" conditions are required only for a "Hot/Dry" climate rating.
ii.) Outdoor humidity conditions during cooling mode tests where the system rejects condensate to the outdoor coil shall be selected to maintain an outdoor relative humidity of 40%. The values for humidity ratio are: 0.025 at 113 °F DB; 0.019 at 104 °F DB; 0.015 at 95 °F DB; 0.011 at 86 °F DB; and 0.0082 at 77 °F DB. For single-package systems where all or part of the indoor section is located in the outdoor test room, the average humidity ratio of the air entering the outdoor coil during wet coil tests must be within 0.0011 of the average humidity ratio of the air entering the indoor coil, over the convergence or measurement period used to calculate capacity and power input.
iii.) Indoor room conditions at start of testing, target for equipment to meet during dynamic test intervals, and indoor room test temperature for full-load test intervals.
iv.) Indoor room conditions at start of testing, and indoor room test condition for full-load test conditions.

Table 4-2. ***SPE-07 test room conditions for heating (SCOPH) rating test series***

Test	Continental outdoor conditions		Marine outdoor conditions		Indoor conditions	
	Dry-bulb temperature, °F	Humidity ratio	Dry-bulb temperature, °F	Humidity ratio	Dry-bulb temperature,[iii] °F	Humidity ratio[iv]
HB[i]	5	0.00080	N/A	N/A	70	0.0092 (max)
HC	17	0.0013	N/A	N/A		
HD	34	0.0031	34	0.0035		
HE	47	0.0042	N/A	N/A		
HF	54	0.0045	N/A	N/A		
HL[i]	LCT	ii	N/A	N/A		

i.) Condition HL (LCT) is an optional test at the lowest catalogued temperature
ii.) The humidity ratio for test HL, if conducted, shall be considered to be a maximum, and shall be calculated as W = 0.000000543 * TDB2 + 0.0000357 * TDB + 0.000554, for the range of LCT −24 °F ≤ TDB ≤ 5 °F. For TDB below −24 °F, W = 0.0000001.
iii.) Indoor conditions at start of testing, target for equipment to meet during virtual-load test intervals, and indoor room test condition for full-load test intervals.
iv.) Indoor room conditions at start of testing, and indoor room test condition for full-load test conditions.

The M1 test included five cooling conditions and nine heating conditions. All five systems received cooling tests A2, B2, Ev, B1, and F1 except system B (a 2-speed unit) that received additional tests C1 and D1. It also received B1 and F1 twice using external static pressure of 0.4 and 0.6 inches of water column. System B did not receive the Ev test.

All systems (except system D) received four basic heating tests: H01, H11, H32, and H42. All but system B also received H1N and H2v tests. Instead, system B received H01 and H11 at two external static pressure conditions (0.4 and 0.6 inches of water column) and received four extra tests: H21, H22 and H1C1 (twice.) Table 4-3 describes the M1 cooling and heating testing conditions.

Table 4-3. ***Detailed testing conditions required in Appendix M1***

COOLING MODE TESTING CONDITIONS						
Test description	**Air entering indoor unit temperature (°F)**		**Air entering outdoor unit temperature (°F)**		**Compressor speed**	**Cooling air volume rate**
	Dry bulb	**Wet bulb**	**Dry bulb**	**Wet bulb**		
A_2 (steady, wet coil, cooling)	80	67	95	75	Cooling Full	Cooling Full-Load
B_2 (steady, wet coil, cooling)	80	67	82	65	Cooling Full	Cooling Full-Load
E_V (steady, wet coil, cooling)	80	67	87	69	Cooling Intermediate	Cooling Intermediate
B_1 (steady, wet coil, cooling)	80	67	82	65	Cooling Minimum	Cooling Minimum
F_1 (steady, wet coil, cooling)	80	67	67	53.5	Cooling Minimum	Cooling Minimum
C_1 (steady, dry coil)	80	57	82	N/A	Cooling Minimum	Cooling Minimum
D_1 (cyclic, dry coil)	80	57	82	N/A	Cooling Minimum	Cooling Minimum

HEATING MODE TESTING CONDITIONS						
Test description	**Air entering indoor unit temperature (°F)**		**Air entering outdoor unit temperature (°F)**		**Compressor speed**	**Heating air volume rate**
	Dry bulb	**Wet bulb**	**Dry bulb**	**Wet bulb**		
$H0_1$ (required, steady)	70	60(max)	62	56.5	Heating Minimum	Heating Minimum
$H1_1$ (required, steady)	70	60(max)	47	43	Heating Minimum	Heating Minimum
$H1_2$ (optional, steady)	70	60(max)	47	43	Heating Full	Heating Full-Load
$H1_N$ (required, steady)	70	60(max)	47	43	Heating Full	Heating Nominal
$H2_2$ (optional, steady)	70	60(max)	35	33	Heating Full	Heating Full-Load
$H2_V$ (required, frost)	70	60(max)	35	33	Heating Intermediate	Heating Intermediate
$H3_2$ (required, steady)	70	60(max)	17	15	Heating Full	Heating Full-Load
$H4_2$ (optional, steady)	70	60(max)	5	4(max)	Heating Full	Heating Full-Load
H_1C_1 (optional, steady)	70	60(max)	47	43	Low	N/A

In heating mode, the indoor temperatures are 70 °F for all tests and we also set the field test heat pump thermostats at 70 °F.[7] For cooling, the SPE-07 and M1 indoor test conditions differ, so field test conditions were set to match SPE-07 using the heat pump thermostats. However, the actual room conditions the systems were controlling were never calibrated to these temperatures, and the temperature of the return air entering the heat pump varied even with a constant setpoint, so there was some variation which was accounted for in the analysis phase. Because none of the tested heat pumps had electric resistance backup heat installed, the factory-installed mobile home electric furnace was used as backup to prevent the house temperature from dropping so much in cold weather that the heat pump performance for those hours would be seriously affected. When the backup furnace was used, field data were not used. A summary of the indoor field test conditions (i.e., the thermostat set points) is shown in Table 4-4.

Table 4-4. ***Heat pump field test conditions***

Heating or Cooling	Heat pump thermostat set point °F	Backup electric furnace heating set point °F
Cooling	74	62-64
Heating	70	62-64

As noted in the Phase 1 report, the indoor humidity ratio in the field test homes was not controlled to the specifications outlined in Table 4-1 to Table 4-3 but was monitored and controlled to the specifications developed for simulating internal loads. In the SPE-07 tests, the indoor humidity ratio is not a controlled indoor room condition past the initial set-up of the psychometric chamber. After set-up, the load-based test uses a virtual latent load model following an assumed sensible heat ratio of the total load to introduce moisture as if it were a real house. The unit under test is allowed to respond under its native controls. The humidity ratio is not maintained by the room conditioning equipment in a steady state as it is in M1, and the test results include reporting of humidity maintained during the test conditions.

[7] There is no thermostat setback used in either the field or the lab testing.

Section 5: FIELD DATA ANALYSIS

The field data analysis section provides detailed information on the field data collection process for the five residential heat pump systems run and tested in three mobile homes. The section outlines the field data used in this study and presents the detailed results.

Table 5-1 lists the specifications of each system, including the system type, energy efficiency, capacity, and other relevant information. To keep the report simple, each unit will be referred to by its unique system label. As the table shows, four of the five units have a cooling capacity of 1.5 tons, and the other two units have a cooling capacity of two tons. The rated heating capacity varies between 18 kBtu/hr to 24 kBtu/h, but it does not always correspond to the cooling capacity. The Seasonal Energy Efficiency Ratio (SEER2) values vary between 15 to 22, and the Heating Season Performance Factor (HSPF2) values vary between nine to 11 according to manufacturers' data.8 Most systems had an Energy Efficiency Ratio at 95 °F of 12.5, though systems A and B were higher at 14.5 and 13.0 respectively.

Table 5-1. ***Specifications of the systems tested in this study***

System label	Type	Cooling capacity (tons)	Rated heating capacity (kBtu/hr)	SEER2	HSPF2	EER
A	Ducted, variable speed	2	24	18.1	8.9	14.5
B	Ducted, two speed	2	22	15.2	7.7	13.0
C	Ducted, variable speed	1.5	22	18.7	9.2	12.5
D	Ductless, variable speed	1.5	22	19.8	8.5	12.5
E	Ductless, variable speed	1.5	18	21.5	11.3	12.5
F	Ductless, variable speed	1.5	19	21.0	10.3	12.5

Field data was collected in one-minute intervals and most instrumentation used one-second sampling. Most sensors remained stable with only one known outright failure. The failed sensor was replaced, and its erroneous data was flagged in the dataset.

Most sensors appear to have measured accurate and confident readings. Confidence in the measurements was afforded by comparing like sensors measuring the same event. For example, nine different sensors measure the temperature of the ductless heat pump discharge/outlet air. When the system is off, all nine temperature

[8] Systems A and B use the AHRI crosswalk to determine the SEER2 and HSPF2 values from the older SEER and HSPF values. The others are from AHRI directory.

sensors converge to the same temperature, within the reported accuracy range of the sensors. Positioning and placement of the sensors had some influence on their measurements, but these were distinct from sensor failure or drift.

DNV supplied UNL data analysts with the complete field data set and supplemental data collection notes marking critical events and spot measurements like:

- Internal gain schedules and power draws
- Known observations of sensor malfunctions or periods of low testing confidence (e.g., a sensor was inadvertently disconnected, both heat pumps ran simultaneously)
- Daily date flags marking heat pump schedules and high-level confidence in field data

UNL used these data to flag and filter field data depending on the type of analysis being performed. These data cleaning steps are further detailed in the next section.

5.1: Data cleaning

Five data sets with one-minute trend data corresponded to the five systems tested in three mobile homes. Each system was switched on or off throughout the cooling and heating testing periods to ensure that only one system operated in each house to meet the indoor load9. The time span of the cooling and heating periods can be found in Table 3-1 in Section 3. In the field study, each unit was operated solely to satisfy the load requirements of each house. However, the building load of each house varied depending on ambient conditions and affected the operation of each system separately.

A collection plan was made for each system in each house to collect operating data across a range of temperatures for each system in heating and cooling modes. Because the ducted and ductless systems in each house are switched on in turns, the operation data are collected when only one system is on and the other system is off, in effect the short overlapping period when both systems were operating was discarded. Table 52 summarizes the number of days of data collection for each system. Those days were tagged and excluded for possible data quality issues when the data should not be used. As described above system D was excluded from the analysis partly because it only had ten days of cooling data due to an install mismatch between the indoor and outdoor units that was not resolved until late in the summer. For all systems, the number of days in the heating testing period was longer than the number of days in the cooling period.

---

[9] During the cooling period, the systems were switched on and off manually. During the heating period, the system thermostats were programmed to turn on and off on fixed schedules

Table 5-2. ***Number of running days for each system***

System label	House	Cooling period (days)	Heating period (days)
A	1	31	73
B	2	28	70
C	3	21	68
D[10]	1	10	54
E	2	25	54
F	3	26	52

To enable the analysis, UNL developed criteria to identify when each system is on or off, when it is in cooling or heating mode, and when it is operating in defrost mode. Each system's operating status was determined using the mass flow rate of refrigerant in combination with the power input of the indoor unit. Table 53 shows the criteria for deciding the "on" status of each system's operation in heating or cooling mode. The standby power varied significantly across the units, necessitating different thresholds for each unit in each operating mode. The mass flow rate is positive for cooling, and negative for heating.

Table 5-3. ***Criteria and thresholds that are used for on/off status identification in each system***

System label	Thresholds of indoor power draw (W)	Thresholds of mass flow rate (g/s)	Criteria used for ON status identification
A	Cooling: 50	Cooling: 10	Cooling: Indoor power > 50 and mass flow rate > 10
	Heating: 25	Heating: -10	Heating: Indoor power > 25 and mass flow rate < -10
B	Cooling: 50	Cooling: 20	Cooling: Indoor power > 50 and mass flow rate > 20
	Heating: 50	Heating: -10	Heating: Indoor power > 50 and mass flow rate < -10
C	Cooling: 20	Cooling: 10	Cooling: Indoor power > 20 and mass flow rate > 10
	Heating: 20	Heating: -10	Heating: Indoor power > 20 and mass flow rate < -10
E	Cooling: 6	Cooling: 10	Cooling: Indoor power > 6 and mass flow rate > 10
	Heating: 8	Heating: -10	Heating: Indoor power > 8 and mass flow rate < -10
F	Cooling: 6	Cooling: 5	Cooling: Indoor power > 6 and mass flow rate > 5
	Heating: 4	Heating: -10	Heating: Indoor power > 4 and mass flow rate < -10

When the ambient temperature falls below a particular threshold during heating, frost development on the outdoor coil is unavoidable. All systems tested in this study employed a demand defrost strategy. In defrost, the

[10] Unit D was not used in the analysis.

mass flow changes direction compared to regular heating mode operation resulting in a sign change for the mass flow sensor reading. This sign change, together with a threshold for a minimum value (to avoid measurement noise), indicates defrost operation. The defrost mass flow thresholds are given in Table 5-4.

Table 5-4. ***Criteria and thresholds that are used for defrost identification in each system***

System label	Thresholds of mass flow rate (g/s)
A	10
B	20
C	10
E	10
F	0

5.2: Capacity calculations

Air thermodynamic properties are indispensable when calculating the air-side capacity. To obtain air-side capacity, enthalpies are calculated first based on dry bulb temperature and relative humidity during the monitoring period. In the field, in order to improve the accuracy of the measured values, multiple temperature and humidity sensors were installed in the airflow upstream and downstream of the indoor fan. Some sensors have periods of bad readings due to the position of the sensor or sometimes a communication problem. Because there were redundant sensors, those sensors prone to inaccurate readings when calculating air-side enthalpies could be excluded.

Because air-side capacity measurements are prone to error when airflow or air condition is not evenly distributed in the cross section of the duct, difficulties measuring the mass of water condensed on the coil, and time delays in measuring dynamically-changing humidity conditions, refrigerant mass flow meters and temperature sensors were installed to calculate the refrigeration-side capacity. This was discussed at length with the technical advisory committee because the refrigerant mass flow meters could affect the performance of the heat pump, as they are installed in the refrigerant loop, adding additional volume and pressure drop to the loop. To compensate for the additional volume, a very small amount of refrigerant (0.1 oz.) was added to the required refrigerant for each HP line11. The flow meter vendor provided a calculation summary of operating pressure drops using R-410A over a range of flow rates.12 The pressure drop range is from 0.02 psig to 0.22 psig. This pressure drop was considered at the time of selection to be very low and within the committee's tolerance. The refrigerant flow rates observed in the field fell within the operating ranges specified by the meter manufacturer, so any effect from the meters should be small. But more importantly, the mass flow meters were kept in place

[11] We removed all of the refrigerant, then added refrigerant back according to the manufacturer's refrigerant mass specification, with an adjustment for line length, and the 0.1 oz. to account for the MF sensor volume.

[12] At 50 degrees Celsius

during the laboratory testing, so that the results of field and laboratory tests would be comparable, even if the performance varied from the manufacturers' official rating performance.

Although there were periods of time when the refrigerant-side measurements were unreliable due to two-phase flow through the meter or a lack of superheat or subcooling in the suction or liquid lines (which prevents determination of enthalpy at those points), in general, findings were that the refrigerant side capacity better represented the system performance compared to air-side capacity. In the analysis, refrigerant side capacity was used wherever possible. A linear regression of air-side to refrigerant-side capacity was developed, and that regression was used to fill in time periods when the refrigerant-side capacity was unreliable.

It is important to note that the primary heating or cooling capacity (and corresponding power input) that was ultimately needed for the analysis is not the instantaneous capacity or even the average capacity "while running". Because the seasonal efficiency bin models used in both rating systems is based on the average building load at each outdoor temperature bin, the capacity ultimately used in the analysis is the net capacity on an hourly basis: that is, the total delivered heating or cooling energy per period of time. This ensures that the capacity and power input reported for each hour (corresponding to the hours represented by each temperature bin for a given climate) is the net, inclusive of any start/stop operations, compressor modulation, multiple cycles at low loads, and defrost operations that send heat in the opposite direction for several minutes.

5.3: COP calculations

The calculation of field COP requires accurate power measurements, so the total power inputs were inspected for each system. Abnormalities in the current transducer (CT) sensor outputs during the system "on" periods were checked for by comparing the measured values to a redundant alternate calculation of total power input of each system using the following equation:

CALCULATED POWER INPUT =
Panel power – House furnace power – the other heat pump system total power input – base load power

The panel power, house furnace power, and other system's input power are all measured using CT sensors, so measurement values can be directly applied in the above equations. The base load power varies for different systems and it dynamically depends on the status of the electric heaters operating on a schedule, which simulated typical occupied internal gains. The other heat pump (i.e. the one not currently under test) was powered, so any standby power drawn from that heat pump is also subtracted from the panel power. Table 55 gives the values for base load power with respect to each system. The table also provides the times in a day when the electric heater is scheduled.

Table 5-5. *Base load power for each system when electric heater is on or off*

System	Base load power (W)	Times in a day when the heater is scheduled on
A or **D**	1350 (heater on); 420 (heater off)	1 am – 2 am; 4 am – 5 am; 7 am – 8 am; 9 am – 10 am; 11 am – 12 pm; 1 pm – 2 pm; 3 pm – 4 pm; 5 pm – 6 pm; 7 pm – 8 pm; 9 pm – 10 pm; 11 pm – 12 am
B or **E**	1550 (heater on); 630 (heater off)	
C or **F**	1460 (heater on); 560 (heater off)	

The indoor temperatures (and humidity in cooling) in the homes were different from the SPE07 lab tests. COP is quite dependent on those, so where possible we used manufacturers' data to adjust the COP for the field data, to provide the COP the heat pumps would have had if they had the same return air conditions as the laboratory SPE07 tests. For some outdoor temperatures, there are no regression data available or no equivalent lab test. In those cases, no adjustment was made.

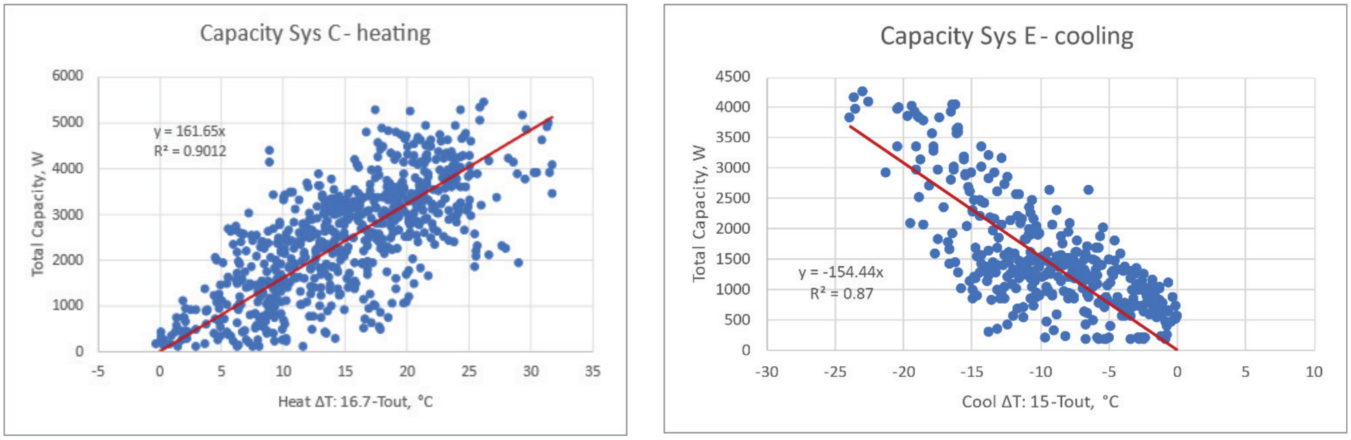
5.4: Field load line calculation

The SPE07 load-based test procedure includes several default parameters that define the virtual load model: the heating and cooling load lines, the load lines' relationship to the equipment size and to each other, the assumed thermal mass time constant to be used during testing, and the sensible heat ratio of the equipment. These parameters must be standardized for a test and rating procedure, but they will never match a particular real house precisely. After much early discussion, the advisory committee agreed that instead of using SPE07 "as-is", these parameters would be measured in the field to the extent possible and duplicated in the lab testing of SPE07. The research objective was not to test the validity of these default parameters to the field sites, but to "test the test procedure" to assess its representativeness.

The intention was that the houses had consistent heating and cooling loads among them and were reasonably close to the SPE07 and 210/240 heating load lines for the selected variable-speed units. In the end, the building heating loads lines, as estimated using the field measurements, averaged about 90 percent of the target (ranging from 70-120 percent). The cooling loads averaged 68 percent of the 210/240 load line but averaged 81 percent of the SPE07 target for the humid climate, which is a bit lower than the 210/240 cooling load line. (Note that 210/240 only uses the load lines in the bin model to calculate the seasonal performance, but not in the test procedure itself.)

It was also necessary to estimate the heating and cooling load lines for each tested unit from the field data, before commencing the laboratory testing. For each of the 10 cases, the load line was estimated by a linear regression of equipment capacity versus outdoor temperature. Figure 5-1 shows examples of the regressions for one heating and one cooling system. It is noteworthy that the heating and cooling load lines were calculated for each system, not for each house; because the ductless systems were not effectively conditioning as much of the house as the ducted systems, each one's effective load was noticeably smaller than that of the ducted system in the same house, particularly during cooling operation. The average sensible heat ratio (SHR) during the cooling season was also estimated for each unit by dividing the total delivered sensible capacity across all valid operating hours by the delivered total capacity for the same hours.

Figure 5-1. ***Typical heating and cooling load line regressions from field data***

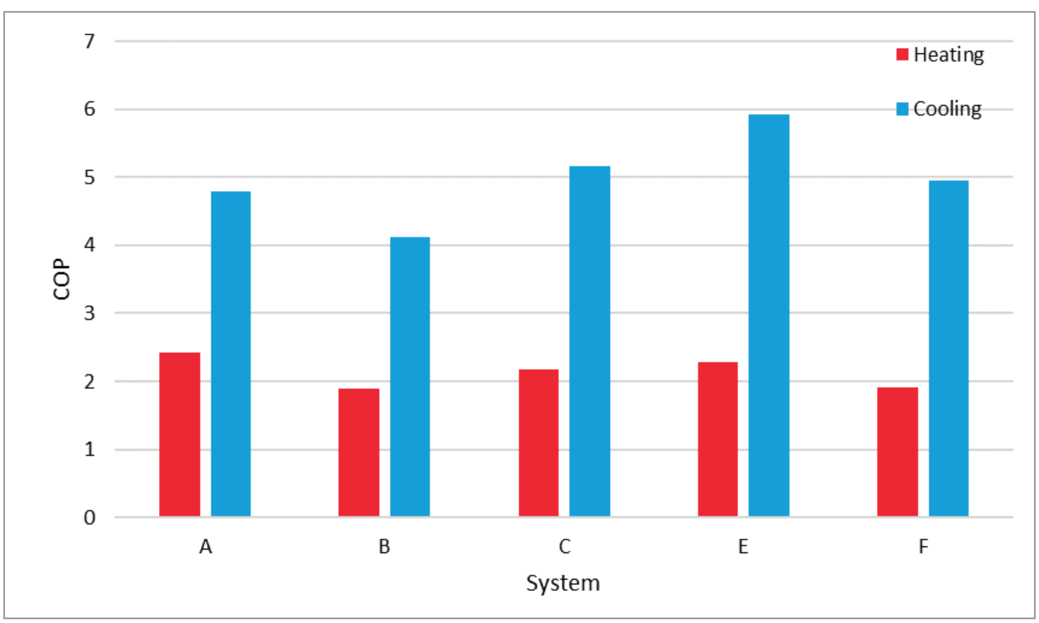


Although it was intended to estimate the shallow mass capacitance of the homes, because of the variable-speed operation of all but one of the five units, adequate data on cycle times was not available to estimate the capacitance.

5.5: Results

The following figure shows the results of the field data analysis. The results in Figure 5-2 are a simple mean of the delivered COP over all the hours that each system was under test. The systems were not running during the same hours, so their operating conditions varied, hence the results are not comparable from one system to the next.

Figure 5-2. ***Unweighted mean cooling and heating COP from field testing over full test duration***



Section 6: LAB DATA ANALYSIS

The lab data analysis section aims to comprehensively examine the data obtained from experimental procedures conducted in the laboratory setting. This section outlines the composition of testing data obtained from lab tests and presents the results obtained. The lab testing activities, including the test descriptions, were provided in Section 4.

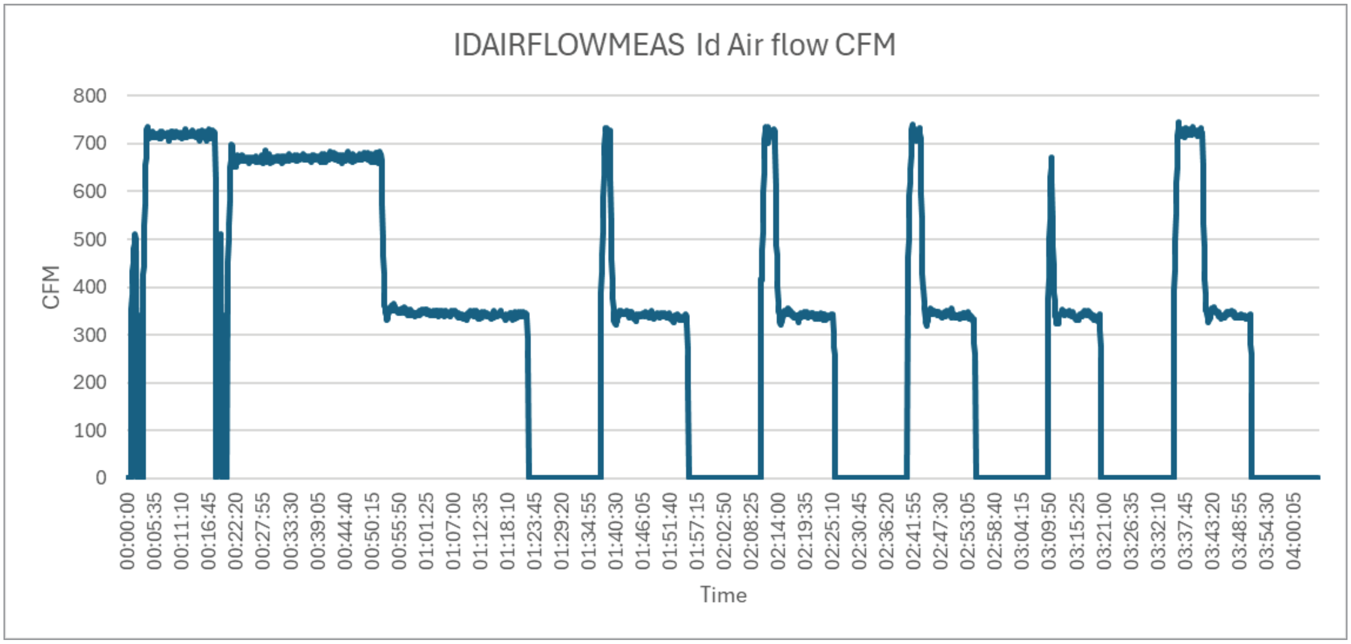
6.1: Data composition

The laboratory data set contains 205 columns in total, with each column representing a variable that is monitored by the instrumentation in the psychrometric chamber. The sampling interval for lab tests is five seconds, and the testing period ranges from one to four hours, depending on the system and the testing conditions. In addition to the existing sensors used by the chamber, the sensors that were installed during field testing are also connected to the data collection system, in part to avoid changing the geometry of any system, and to enable comparison with laboratory sensors. In the laboratory measurement results, the air-side capacity values were used, which is consistent with typical laboratory practice.

For the SPE-07 procedure, calculating the SCOPs for different climate zones requires multiple variables obtained in laboratory results. For cooling and heating SCOPs the calculation requires measured capacity, power input and outdoor dry bulb temperature. As a result, SCOPs for eight heating climate zones and seven cooling climate zones were calculated according to the procedures in SPE07.

One observation from the SPE07 laboratory data was the modulation of the indoor fan when the system was being tested. During the test, the system was acting under its own control to match the building load simulated in the laboratory. At low loads this can lead to cycling of the system, and several monitored variables can be affected. One typical variable is the indoor air flow rate. Figure 61 indicates the on-off status of the indoor conditioning unit of System A using air flow rate. As seen in the figure, indoor airflow rate modulated as the test continued, a result of the equipment controller. Despite the variation of indoor air flow rate, when calculating SEER2 and HSPF2 for each system, we used the mean airflow value.

Figure 6-1. *Indoor air flow rate of System A observed at zero during some periods when the system was under test*



In M1 lab tests, cooling and heating tests follow the conditions outlined in the standard. Five cooling tests and six heating tests are considered in calculating the SEER2 and HSPF2 metrics. The input variables for SEER2 and HSPF2 are capacity and power input under each condition. During M1 lab tests, indoor and outdoor air temperature and wet bulb temperature are controlled by the psychrometric chamber controls for both heating and cooling conditions. Compressor speed and indoor blower air flow rate are manually adjusted to satisfy the requirement of each testing condition.

There were a total of three M1 test conditions among the five units that were missing from the laboratory data. These were the H31 test for system B (heating), the B1 test for system C (cooling), and the H01 test for system F (heating). In all of these cases we estimated the value by interpolating the results from other tests. For systems B and C the HSPF2 and SEER2 results (respectively) were not very sensitive to the estimated test condition. For system F the HSPF2 value was more sensitive to the estimated value. We pushed that one estimated test value to an extreme, to make the lab HSPF2 as low as it conceivably could be, and still ended up with an HSPF that was higher than the manufacturer-rated value. The lab SEER2 value for unit F was higher than the rated value to a similar extent, so this appears to be a reasonable approach.

AHRI has provided an online calculator to assist in calculating SEER2 and HSPF2 for various types of heat pumps. This calculator has been validated and adopted widely to calculate such values, so it was used to calculate SEER2 and HSPF2 values for all tested systems in this study, as well as using off-line calculations to confirm the values. Inputs into the calculator were: capacity, power input, and standard air flow rate in CFM from the lab testing results.

6.2: Results

After obtaining all the values of the input variables for each system, we calculated the results of SCOPs, SEERs, and HSPFs for the laboratory tests. For all the required inputs to SCOP, SEER2, and HSPF2 calculations, the lab measurements of capacity and power were determined according to the requirements of the two test procedures. Figure 6-2 and Figure 6-3 show the calculated SCOP values using SPE-07 in cooling and heating respectively. In each case the SCOP values are shown for the SPE-07 climate zones most similar to the M1 climate zones.

Figure 6-2. ***Laboratory SPE-07 cooling results (SCOP) in the mixed climate vs. Laboratory M1 cooling results (SCOP)***

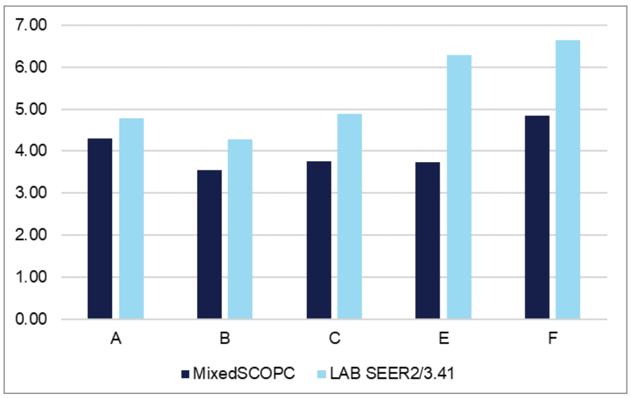
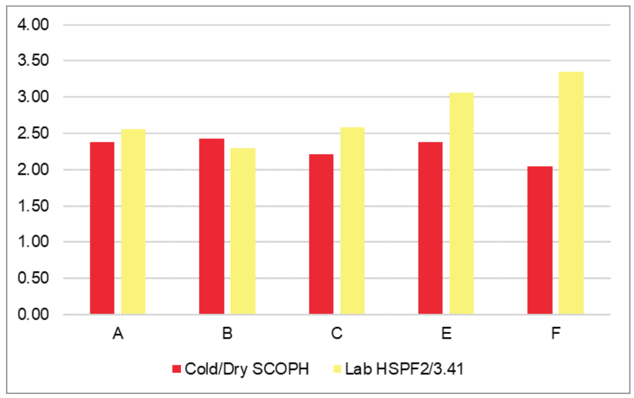


Figure 6-3. ***Laboratory SPE-07 heating results (SCOP) in the cold/dry climate vs. Laboratory M1 heating results (SCOP)***



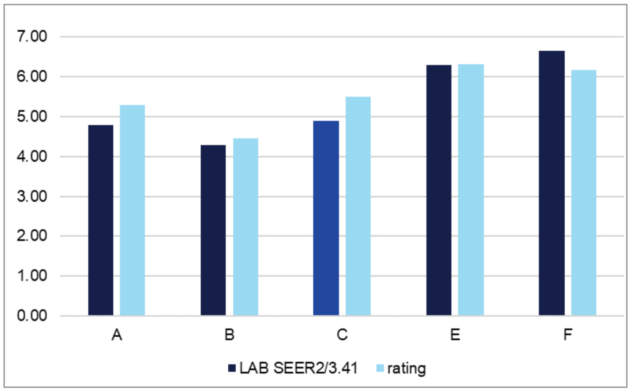
6.3: M1 ratings comparison

In Figure 6-4 and Figure 6-5, the efficiency ratings from the original equipment manufacturers (OEMs) rated values as published in the AHRI directory[13] are compared to those generated for the specific machines in the UL laboratory tests. In both graphs, the lab SEER2 and HSPF2 values that contain estimated test results (as discussed in Section 6.1) are shown in a different color than the other lab values. Note that the pairs of values in Figures 6-4 and 6-5 are not derived from identical sources. First, the lab "ratings" report the results of the test procedure on individual devices, whereas published ratings are typically generated based on lab tests of multiple devices (and sometimes also using modeling) that are further qualified by what the manufacturer chooses to report. Published ratings may tend to be conservative (that is, on the low side) in order to manage risk and ensure that products predominantly meet or exceed their published ratings. Second, about half of the lab results are slightly lower than the published ratings. It is not unexpected that many of these units would have slightly reduced results in the lab, because of the installed refrigerant mass-flow sensors which would slightly reduce capacity (and thus efficiency).[14]

[13] Systems A and B use the AHRI/DOE crosswalk to determine the SEER2 and HSPF2 values from the older SEER and HSPF values. The others are from AHRI directory.

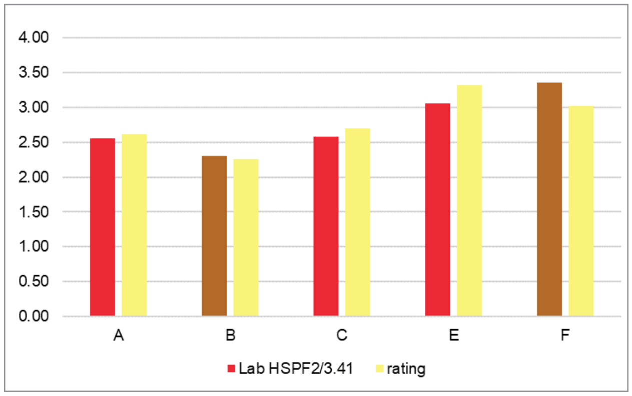
[14] Because the mass-flow sensors were present in both settings, they should not bias comparisons between lab and field performance in the study.

Figure 6-4. ***Comparison of manufacturer's published cooling efficiency rating with values measured in the laboratory (SCOP)***



System C lab testing contained an estimated test result (as discussed in Section 6.1)

Figure 6-5. ***Comparison of manufacturer's published heating efficiency rating with values measured in the laboratory (SCOP)***



Systems B and F lab testing contained an estimated test result (as discussed in Section 6.1)

Section 7: FIELD AND LAB COMPARISONS

To determine how well the two major laboratory tests represent field performance, the team compared the SPE-07 SCOPC and SCOPH, and M1 SEER2 and HSPF2 (which are specific to these devices' lab tests, and thus varied from the "rated" values as shown above) to the field-measured equivalents of the same metrics as outlined in the test condition tables shown in Appendix A. Field-measured data were used to generate seasonal Field SCOP values based on the bin-hour models for the cold/dry heating and the mixed cooling climate zones. (These climates are closest to the DOE Region IV heating and the national cooling bin-hours provided by M1). The lab-measured SPE07 testing data were used to generate lab SCOP values for the same climate zones. The lab-measured M1 test results were also adjusted to reflect the same cold/dry and mixed climates, so that all three sets of data were normalized to the same climates for heating and cooling, respectively. Finally, the M1 results were used to generate "lab HSPF2" and "lab SEER2" values (as differentiated from the product ratings), which were then divided by 3.412 to make them dimensionally equivalent to SCOP values for comparison.

The Appendix M1, SPE-07:23, and field-tested performance SCOP values were compared with a particular focus on relative rankings and performance. The field performance-calculated metrics were thus compared on a climate-normalized basis to the test laboratory performance (e.g., percent deviation of HSPF2 and SCOPH from the field SCOP). Each data set (lab and field) was processed to enable accurate "apples to apples" comparisons. The repeatability of the tests was also investigated and reported in Section 4.1. Daily COP vs. daily outdoor air temperature values from the field-measured tests are plotted and analyzed in Section 7.3. Section 7.4 addresses aspects of each test procedure required to generate ratings that reasonably reflect annual performance under different climate conditions.

We used the climate model of SPE-07 to compare the field results with the laboratory results by calculating a seasonal COP (SCOP) for each data set using the same climate model for all three. SPE-07 uses an assigned number of operation hours over a set of temperature bins to define each climate for which results are reported. The field test climate (in Nebraska) was most similar to the SPE-07 cold/humid climate, but past comparisons of SPE07 and M1 have been made using the cold/dry climate for heating and the mixed climate for cooling, because they were the closest to the climate bins used in reporting HSPF(2) and SEER(2). Therefore, those climates were used for the comparisons in Section 7.1. However, when the systems were operating in the field, the outdoor temperatures did not cover the entire range of SPE-07 temperature bins, so several of the extreme bins use extrapolated values (or in the case of the coldest heating temperatures, values from the SPE07 tests themselves). These substitutions were expected to have little impact on the results, because they only affect the extreme temperatures and thus have very few annual hours.

The field SCOP result was calculated based on the capacity and heat pump input power in each temperature bin. The number of hours in the field for each bin was based on the weather experienced in the home during the hours each unit was under test (about half the time for each heat pump), and this could differ significantly from the number of hours assumed in the SPE-07 lab tests. The variations in weather during the study period could significantly affect the SCOP result. To make the field and lab tests more comparable, we normalized the field results by taking the entire capacity and input power for all the hours at each temperature range, and applying them to the bin hours from the SPE-07 climate model. Ultimately, this is saying "the house had a range of different actual loads at that temperature than the lab had, but if it had the same load and the same climate, this is how much heat it would have extracted (or delivered) and how much power it would have used."

The same normalization process was used for M1 test results. The M1 bin model has a far larger deviation from the field load line than SPE-07. In particular, the heating load line for M1 is based on the measured cooling capacity at 95 (the Afull test), and in the field houses the actual cooling load (which could not be measured or calibrated in advance) was much smaller relative to the design heating load than is assumed by M1. Because the M1 calculation interpolates between high, intermediate, and low speeds based on its own load line assumptions, the research team decided that the fairest way to normalize them to compare with SPE-07 and the field results was to take the COP calculated using M1 for each temperature bin and apply it directly to the climate bin model to match the other two. In the case of cooling, this was straightforward because the M1-calculated COP defined the input power based on the cooling load line of each bin. For heating, the normalized bin models included electric resistance heat. The resistance heat was applied consistently to the bin model across all three normalized data sets (field, SPE-07, and M1) to ensure a fair and comparable normalization process for heating.

7.1: Normalized field and lab ratings comparison

Figure 7-1 presents three metrics to compare cooling COPs across field and lab tests. The leftmost, normalized field SCOPc, is the cooling SCOP value calculated from the field data, using the calculation procedure from SPE-07 with bin-hours based on the mixed climate. As described above, the indoor temperatures have been adjusted to match the indoor temperatures of the SPE-07 test and the field data has been adjusted to account for the load line being different in the field than in the SPE-07 lab tests. The second value is the SCOP value calculated from the SPE-07 laboratory test with bin-hours based on the same climate. The third value is the M1 test results from the lab, normalized to use the same load line and climate as the SPE-07 test. The rightmost value is the SEER2 value from the laboratory for climate Region IV (not normalized), divided by 3.412 Btu/h·W-1, to make it dimensionless and comparable with the SCOP values. The Region IV climate, however, is not the same as the mixed climate, so the "Lab SEER2" value is *not* directly comparable to the other three; it is shown for reference only, so the reader can see the impact of the climate normalization on the M1 results.

The normalized M1 result has the highest value in all cases except System A, where the field value is slightly higher. Overall, SPE-07 appears to have moderately underpredicted efficiency in cooling, whereas M1 moderately overpredicted for two ducted units (B, and C) and overpredicted for ductless units (E and F).

Figure 7-1. *Comparison of cooling SCOP values calculated from field and lab data*

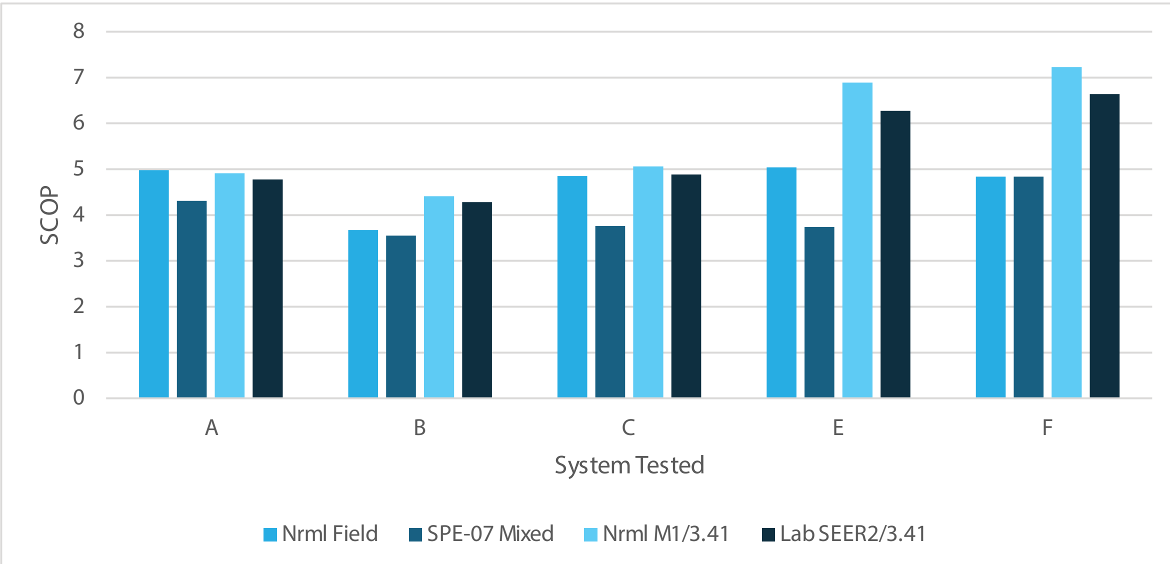
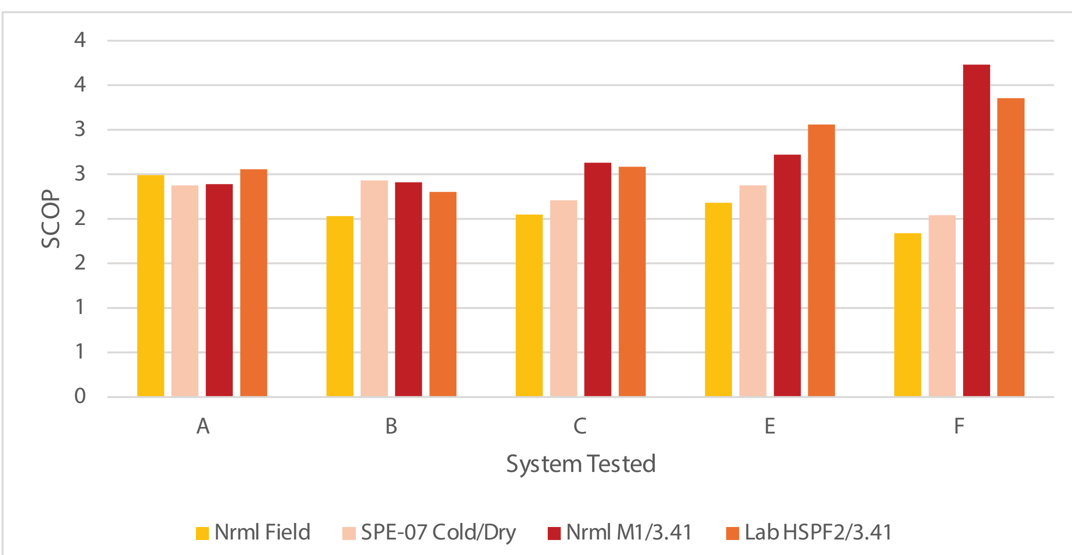


Figure 7-2 is analogous to Figure 7-1, but showing heating results. Interestingly, some of the same patterns emerge that were apparent in the cooling results. The four values are more similar for Systems A, B, and C, which were the ducted systems, than they are for systems E and F, the ductless systems. In most cases, field SCOP is higher than SPE-07 lab SCOP in cooling, while it is lower in heating. Comparing field SCOP and M1 results across both heating and cooling, the field results are lower in all cases except System A. These similar patterns suggest that random error is not the source of the discrepancies between metrics and that there could be some systematic differences.

Figure 7-2. *Comparison of heating SCOP values calculated from field and lab data*



If the field SCOP metric is defined as the correct value, then a holistic comparison can be made by calculating the root mean square error (RMSE) (shown in Table 7-1) and the mean absolute percent error (MAPE) (shown in Table 7-2) for each of the other two metrics. Using both of these methods, negative and positive errors do not cancel, and in the RMSE method large errors have more effect on the result than smaller errors. Minimizing the RMSE is commonly used for curve fitting.

As above, the cooling metrics are calculated for the mixed climate and the heating metrics are calculated for the cold/dry climate. In cooling, the combined errors for SPE-07 are about half those for M1, while in heating they are about one third. One might expect SPE-07 to outperform M1 due to the rigor of that testing procedure and its load-based approach. The distinction between the two tests is clearer for the ductless heat pumps, where SPE-07 performs better compared to field tests than M1, though it is not as clear for the ducted heat pumps.

Table 7-1. ***The root mean squared errors (RMSE) for SPE-07 and M1 metrics, using field SCOP as a reference*.**

	Cooling		Heating	
	SPE-07	M1	SPE-07	M1
Ducted	0.74	0.45	0.26	0.40
Ductless	0.92	2.14	0.20	1.39
Combined	0.82	1.40	0.24	0.93

Table 7-2. ***The mean absolute percent errors (MAPE) for SPE-07 and M1 metrics, using field SCOP as a reference.***

	Cooling		Heating	
	SPE-07	M1	SPE-07	M1
Ducted	13%	9%	11%	17%
Ductless	13%	43%	10%	64%
Combined	13%	22%	10%	36%

Another way to look at the results visually is by showing the rank order. The visual layout in Figure 7-5 and Figure 7-6 show comparisons between SPE07 (and EXP07) lab tests and the published ratings of the same models. In these cases, the lab tests were both conducted on the same unit and were both normalized to the same climate, along with the insertion of the normalized field seasonal efficiencies in the center. For cooling, there is not a lot of variation in the efficiencies in the field. While both rating systems show more variation, M1 shows a much wider spread than SPE07. The somewhat "low" bias of SPE07 compared to the field is also evident visually, even as M1 appears to be biased somewhat "high." For heating, SPE07 seems to retain the actual field efficiency ranking somewhat better than M1, and has a more similar spread of efficiencies relative to the field measurements, compared with M1.

Figure 7-3. *Rank order of cooling SCOP values from field and lab data – normalized to Mixed climate*

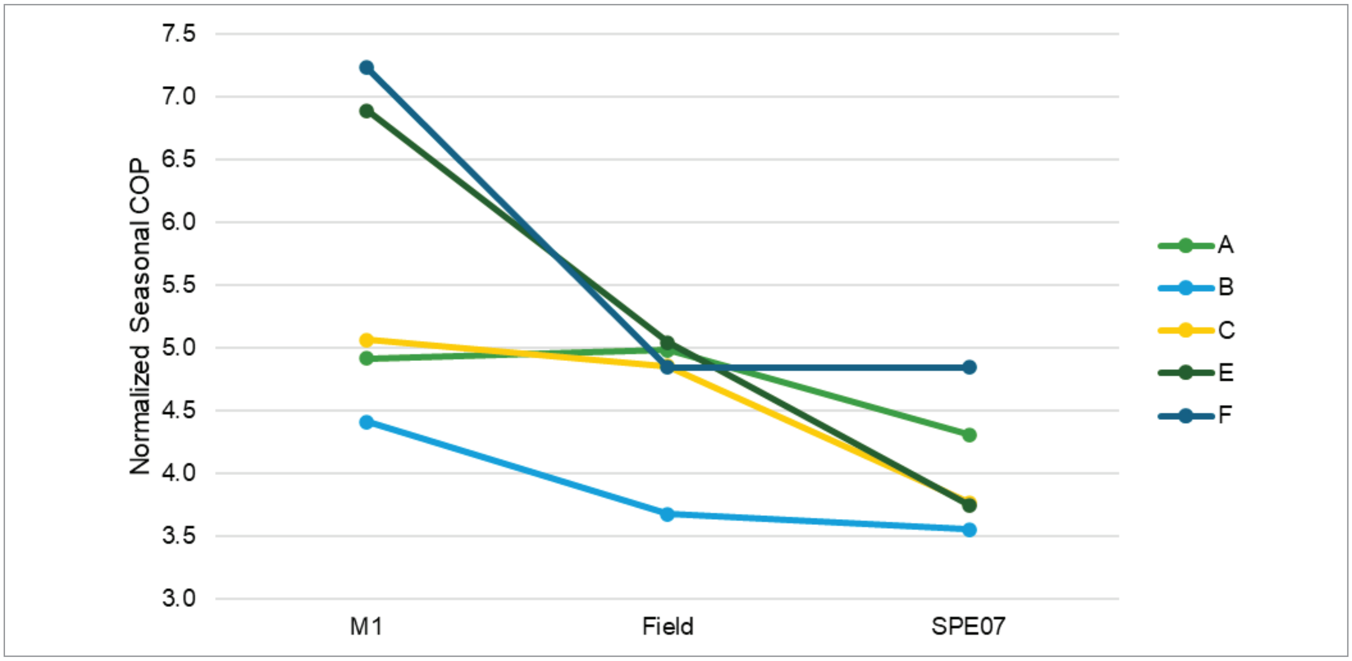
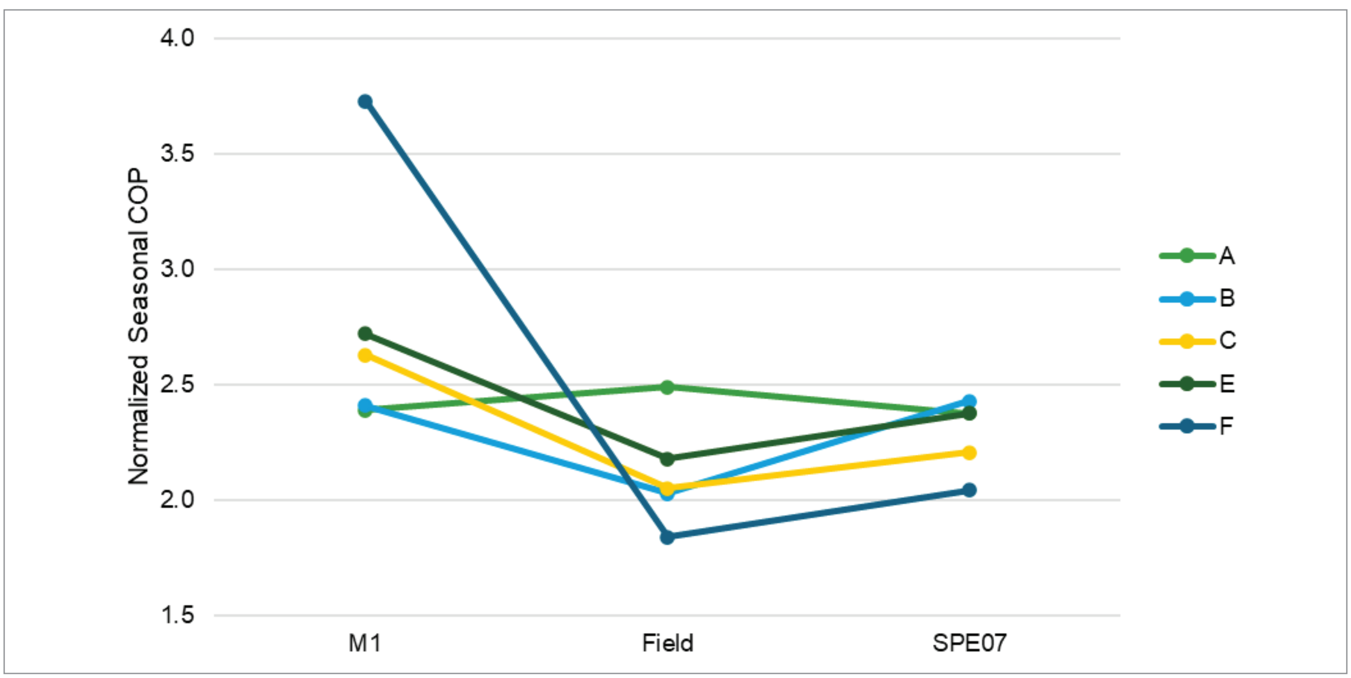


Figure 7-4. *Rank order of heating SCOP values from field and lab data – normalized to Cold/Dry climate*

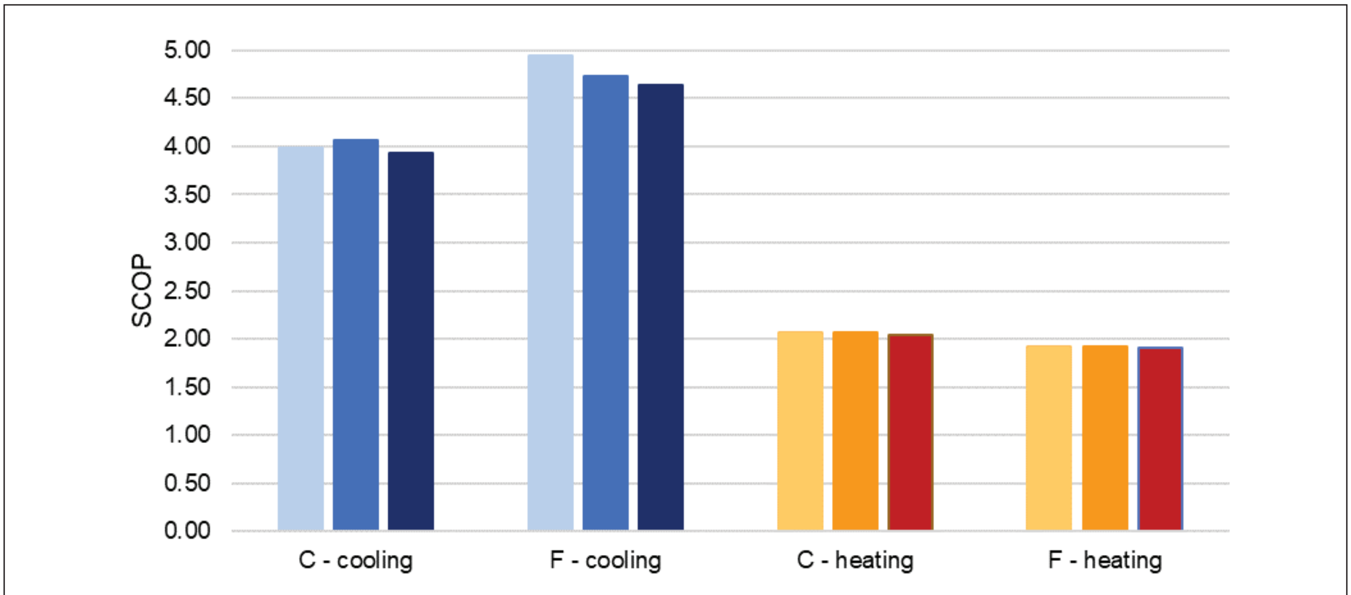


7.2: Repeatability of laboratory tests

SPE-07 tests require systems to run under native controls, which means that the unit under test may modulate to different speeds and capacities in ways that are not controlled by the laboratory testing facility. This raises the question of whether minor variations in test conditions and procedures, or nonlinear controller behavior, may result in variations in the rating. To give an indication of the repeatability of the SPE-07 test, two systems – C and F – were each tested three times in the laboratory, in both heating and cooling modes. This section describes a comparison of the repeated tests.

Figure 7-5 shows the SCOP results of the repeated tests for cold humid climates. The SPE-07 values are used throughout the repeatability analysis.

Figure 7-5. ***SCOP results from three repetitions of SPE-07 test for cold/humid climate on systems C and F***



In Table 7-3 the numerical results of the SPE-07 values are given, with some descriptive statistics, including the mean, standard deviation, and confidence interval for a 95 percent confidence level. The confidence interval is shown in both SCOP and percentage. As an example of interpreting the confidence interval, the results show that for unit C in cooling, for 95 percent of the applications of the rating tests, the results will be within the range of ±1.6 percent of the mean value. The cooling values show much more variation than heating values, potentially because of the complexity of controlling humidity repeatably in a laboratory.

Table 7-3. *Descriptive statistics for the repeatability of SCOP test results, based on three repeated tests of Units C and F for cold/humid climate*

	C - cooling	F - cooling	C - heating	F - heating
Sample 1	3.99	4.95	2.07	1.92
Sample 2	4.07	4.74	2.07	1.92
Sample 3	3.93	4.64	2.04	1.91
Std Dev	0.06	0.13	0.01	0.01
Mean	3.99	4.78	2.06	1.92
Conf. Interval ±	0.0634	0.1473	0.0164	0.0092
Conf. Interval ±	1.6%	3.1%	0.8%	0.5%

If these repeatability results are combined to consider the question of how much variation is likely in multiple tests of systems in a given laboratory, including the heating and cooling results, the 95 percent confidence interval is within 3.1 percent, based on this small sample of four sets of three SCOP results.

7.3: Performance maps

The following sections show the COP and capacity performance maps in cooling and heating modes from each data set (field, SPE-07, and M1) for each heat pump system tested. The cooling M1 tests included A2 and B2 (max load) with B1 and F1 (min load), and the heating tests included H01 and H11 (min load) with H1N, H2V, H32, and H42 (max load). Note that both the field and the SPE-07 data are building-load driven and expected to align, whereas the M1 tests have fixed indoor temperatures and humidities for each test point. System D is omitted as discussed earlier. The M1 tests are shown with lines connecting them only for graphical clarity.

7.3.1: Capacity

The following ten figures show the cooling capacity performance maps from field data, and both lab testing procedures for each of the five systems first in cooling mode and then in heating mode. The field capacity and the SPE-07 capacity align well for most systems except for system B and E in heating mode where the SPE-07 loads are higher. These systems were both tested in house #2 indicating there may have been a problem with the load line for that house.

Figure 7-6. ***System A cooling capacity by outdoor dry bulb temperature***

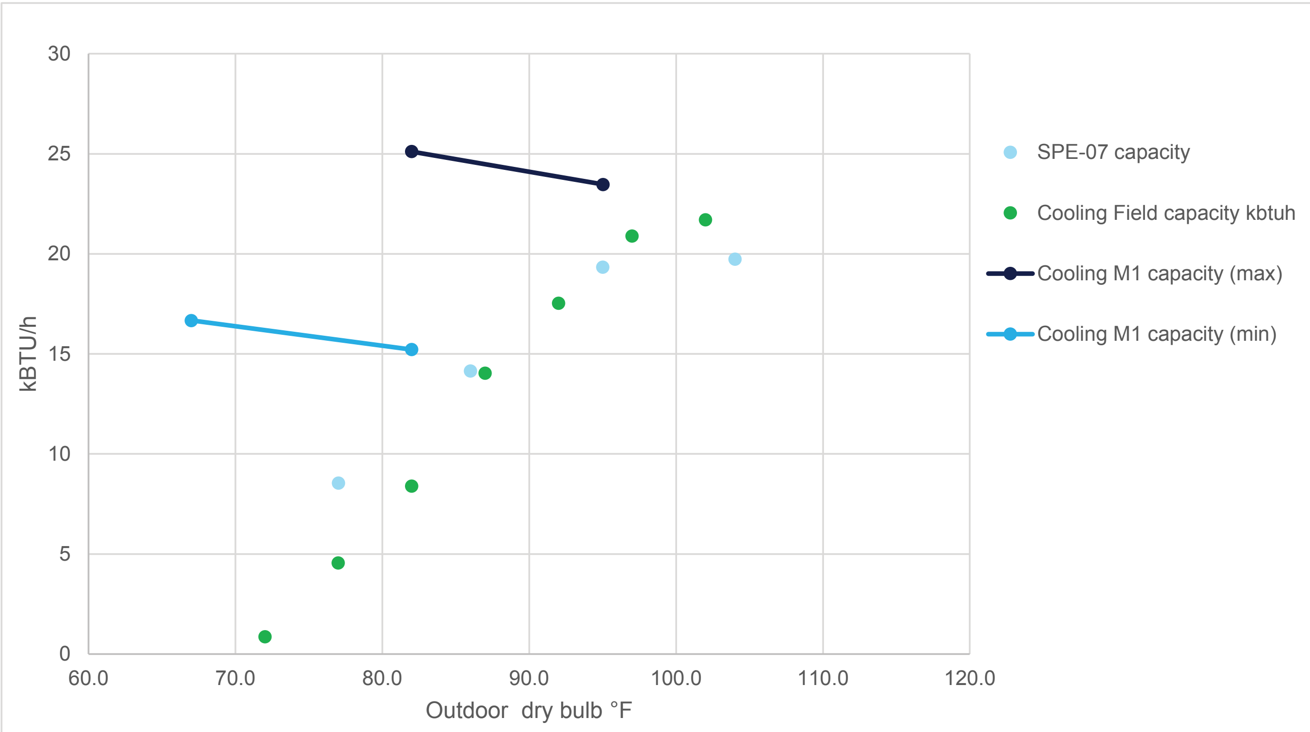


Figure 7-7. ***System B cooling capacity by outdoor dry bulb temperature***

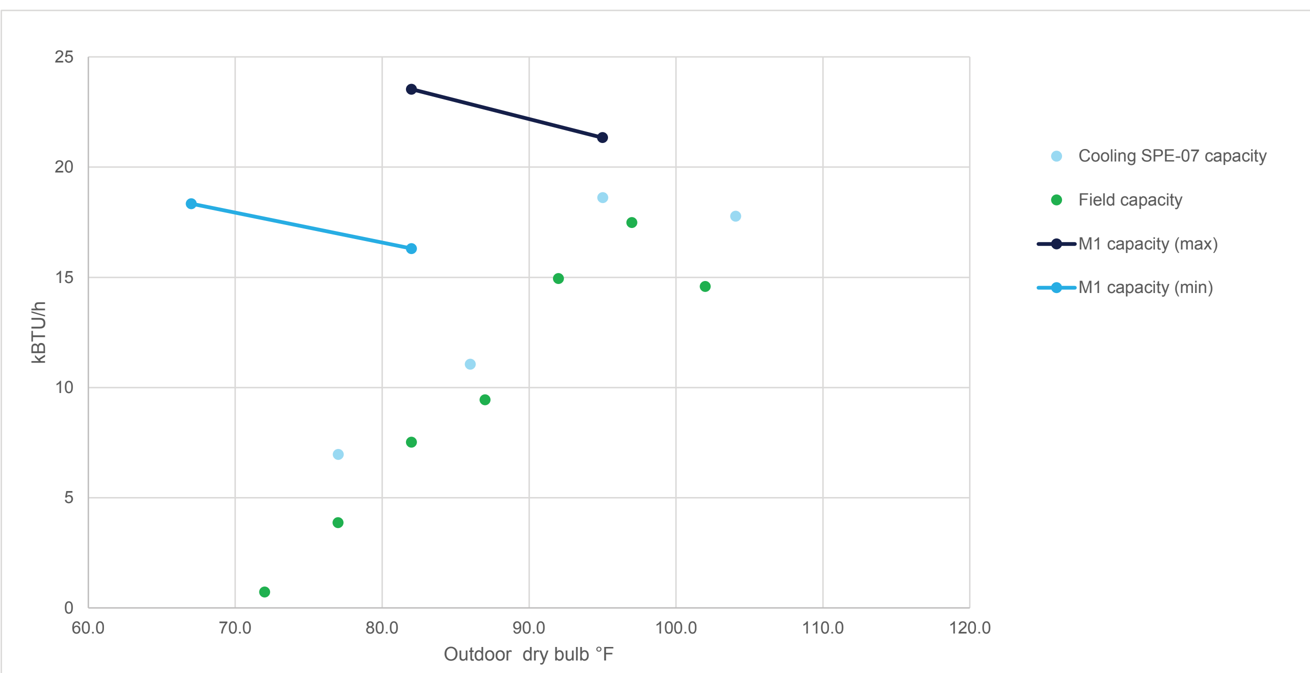


Figure 7-8. ***System C cooling capacity by outdoor dry bulb temperature***

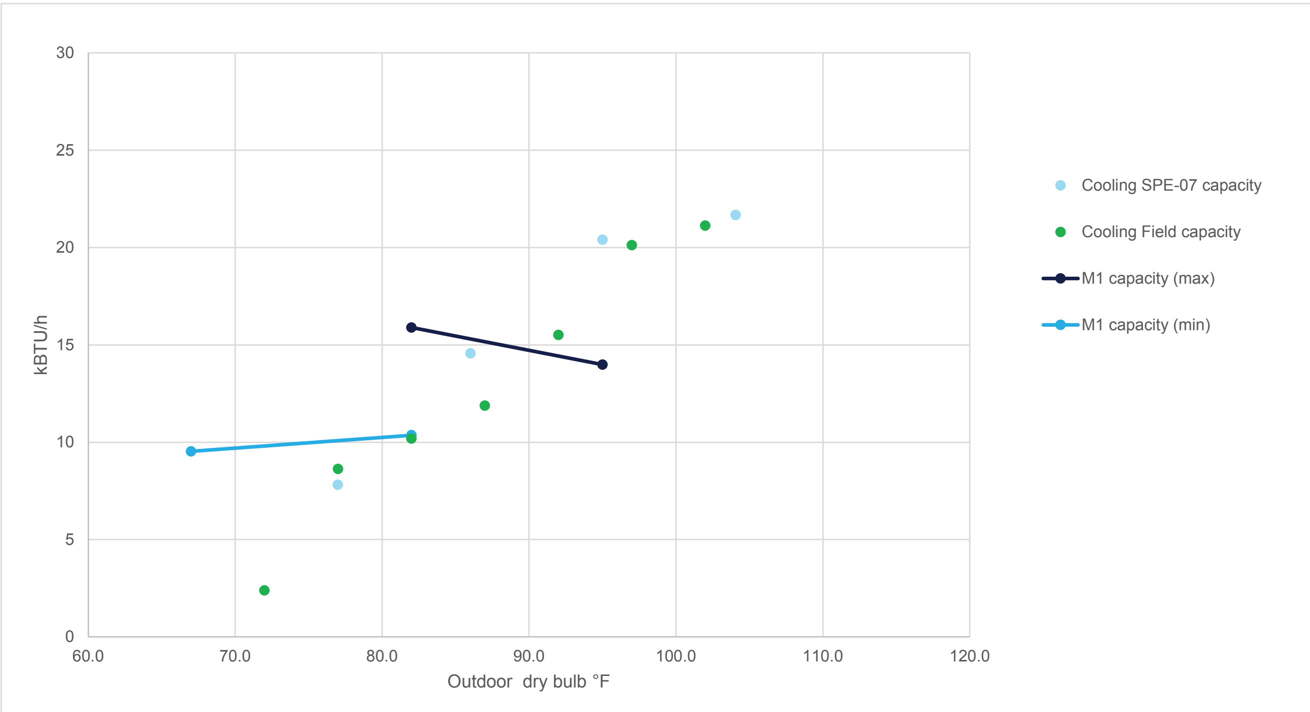


Figure 7-9. ***System E cooling capacity by outdoor dry bulb temperature***

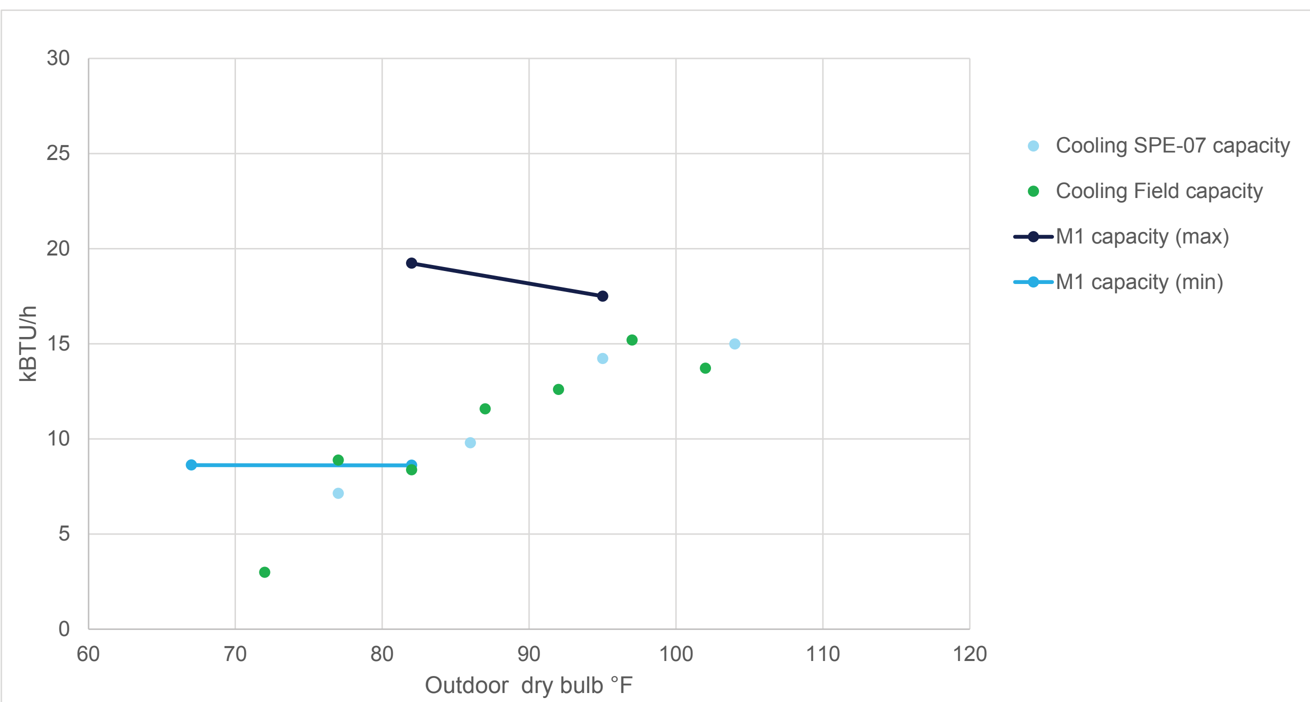
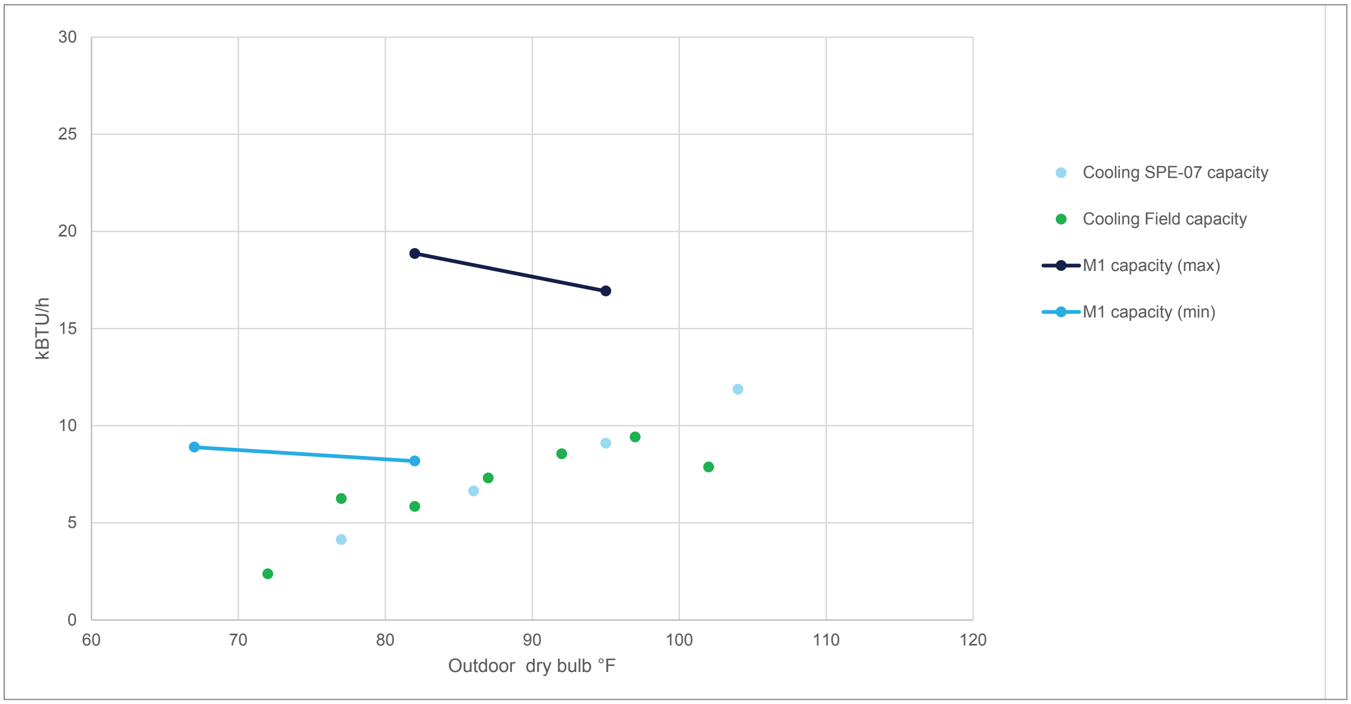


Figure 7-10. ***System F cooling capacity by outdoor dry bulb temperature***



The following figures show heating capacity performance maps from field data, and both lab testing procedures for each of the five systems.

Figure 7-11. ***System A heating capacity by outdoor dry bulb temperature***

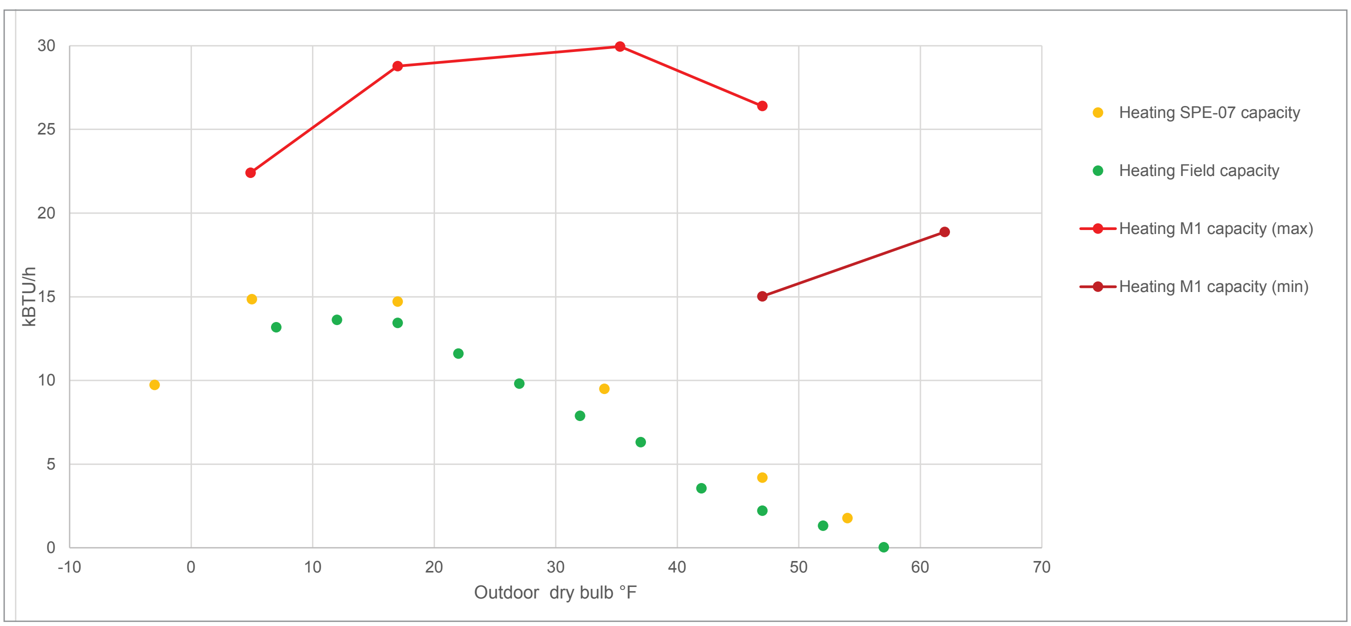


Figure 7-12. *System B heating capacity by outdoor dry bulb temperature*

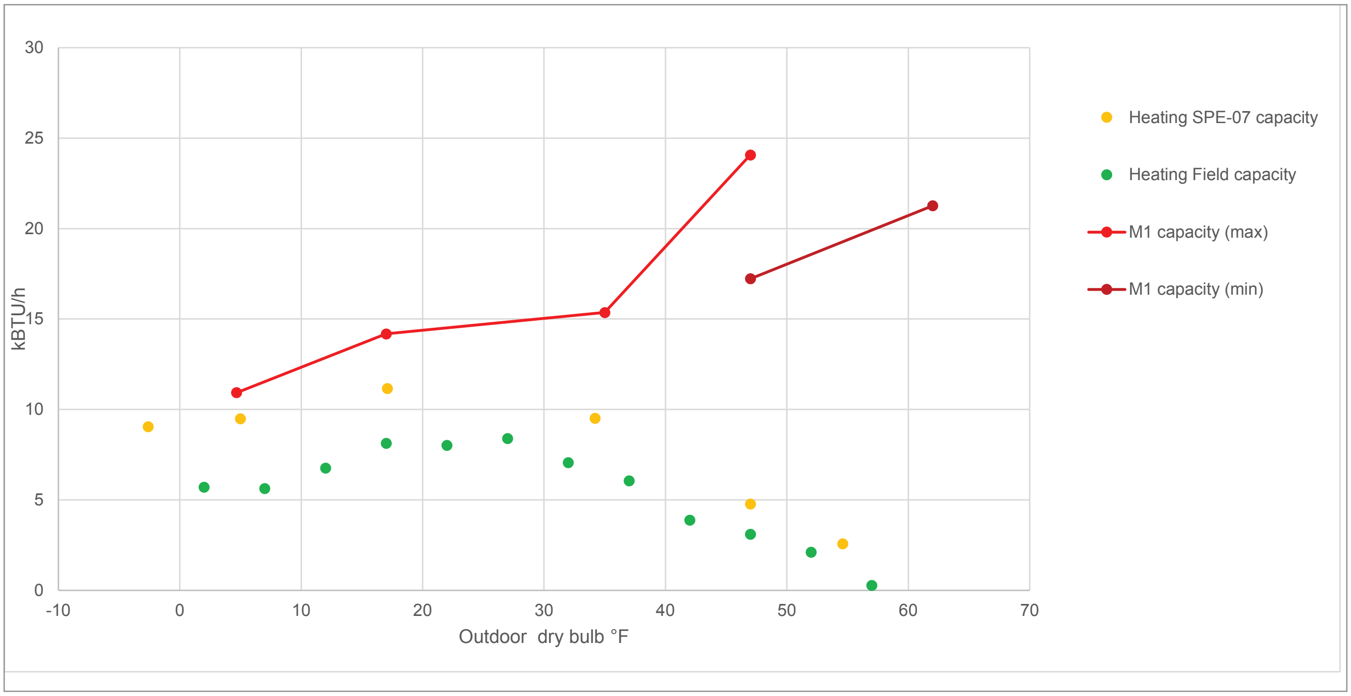


Figure 7-13. *System C heating capacity by outdoor dry bulb temperature*

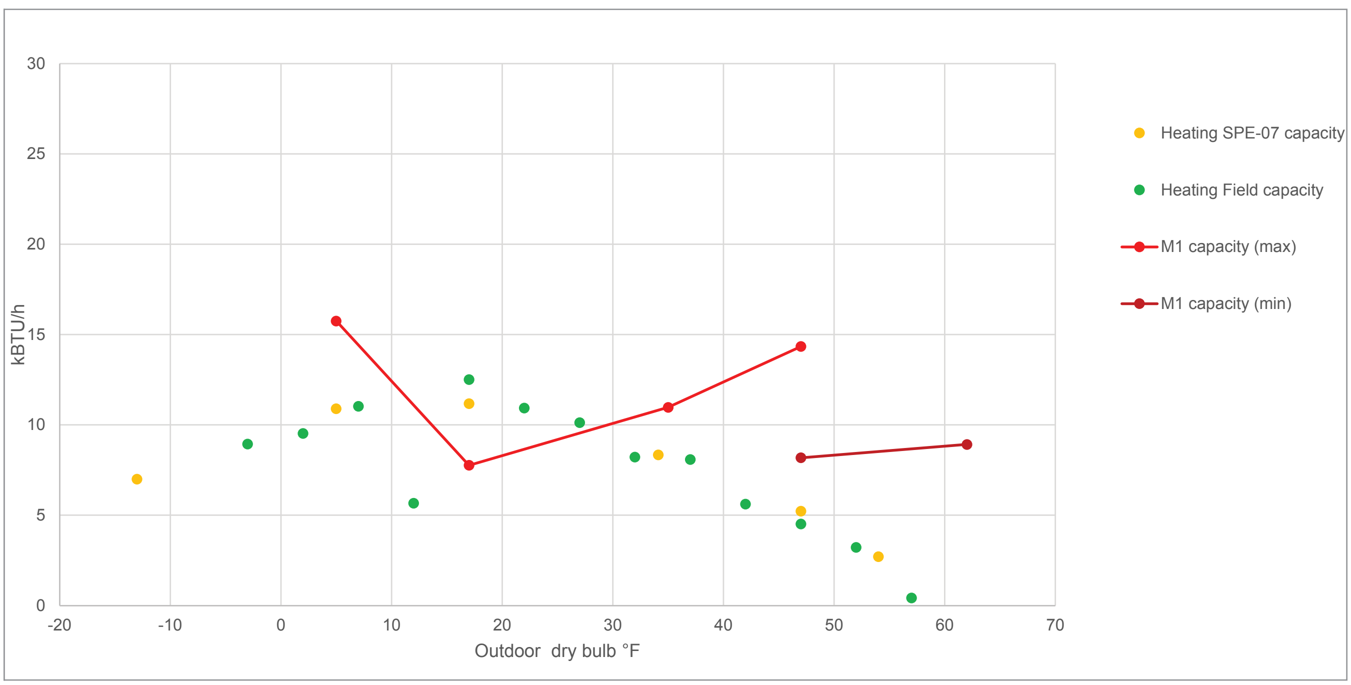


Figure 7-14. ***System E heating capacity by outdoor dry bulb temperature***

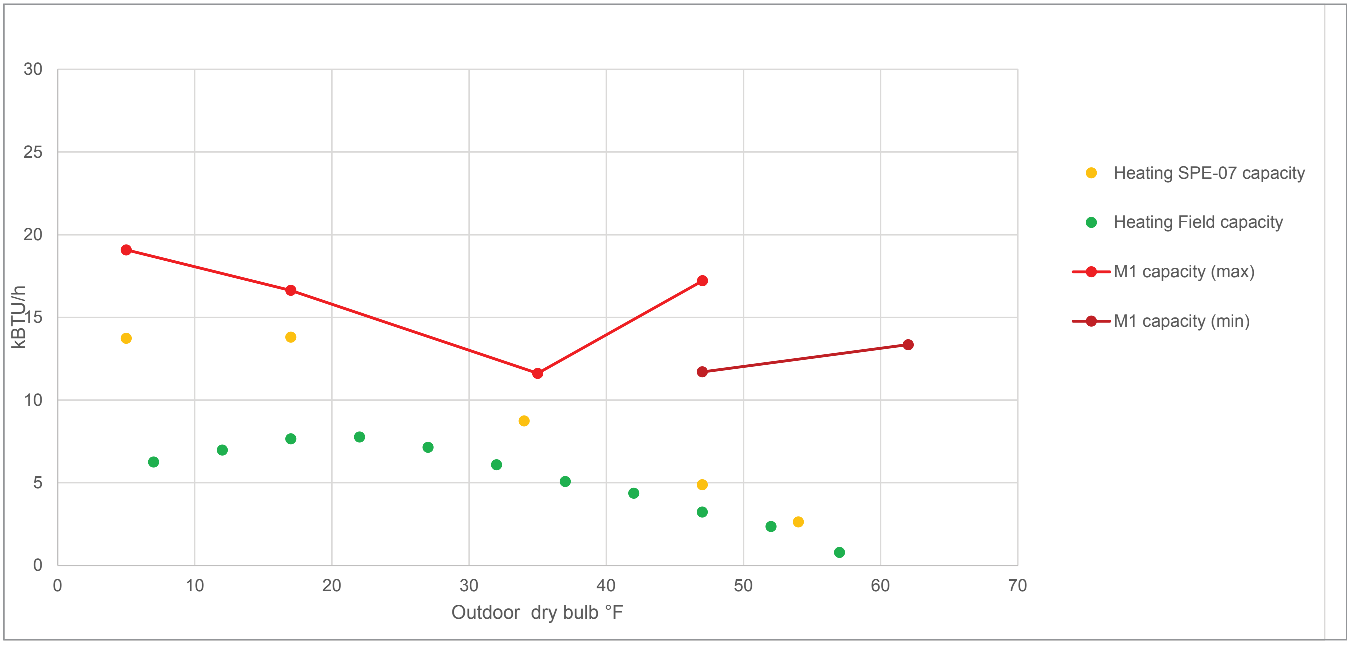
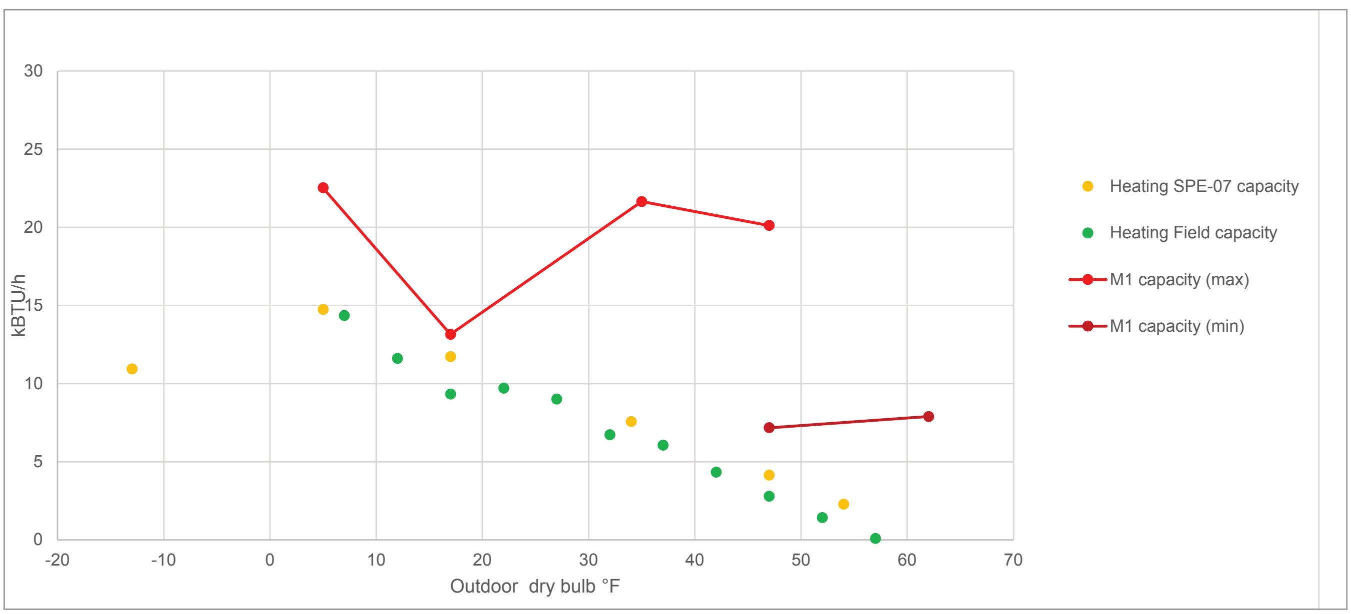


Figure 7-15. ***System F heating capacity by outdoor dry bulb temperature***



7.3.2: Coefficient of performance

The following ten figures show the COP performance maps from field data, and both lab testing procedures for each of the five systems first in cooling mode and then in heating mode. In general, the COPs agree across all tests better than the capacity in the previous plots. The M1 minimum load tests have the best COPs as expected. At those milder temperatures the systems are most likely cycling in the field and SPE-07 tests. The heating field data is very different from M1 and SPE-07 lab tests for systems B and E indicating there may have been a bias error in those tests.

To understand if SPE-07 does a better job at measuring low load efficiency we compared the heating COP values at 47°F and the cooling COP values at 82°F across the field and lab tests (using minimum test for M1) and calculated the mean absolute percent error (MAPE) and root mean squared error (RMSE) for each of the lab tests using the field data defined as the correct performance values. The raw heating COP values and the results are shown in Table 7-4 and the cooling COP values are shown in Table 75. For all the heating tests, SPE-07 provides a lower COP than M1, and for all systems except System A, it matches the field performance better than M1. Although cooling results vary more between the systems, SPE-07 consistently under-predicts performance at low cooling loads while M1 over-predicts most of the time. When looking at all five systems together the cooling tests show that SPE-07 tests match field performance slightly better than M1 lab tests.

Table 7-4. ***The mean absolute percent error and root mean squared error for SPE-07 and M1 COP at 47°F (minimum heating load), using field COP as a reference.***

	COP at 47F min heating			Percent error		Squared error	
System	SPE-07	Field	M1	SPE-07	M1	SPE-07	M1
A	2.8	3.92	3.49	-28%	-11%	1.25	0.18
B	3.23	2.38	3.39	36%	43%	0.73	1.03
C	3.11	2.41	4.59	29%	91%	0.50	4.77
E	2.63	2.81	3.59	-6%	28%	0.03	0.61
F	2.26	1.91	6.02	18%	215%	0.12	16.87
				MAPE		RMSE	
				24%	77%	0.72	2.17

Table 7-5. ***The mean absolute percent error and root mean squared error for SPE-07 and M1 COP at 82°F (minimum cooling load), using field COP as a reference.***

	COP at 82F Min cooling			Percent error		Squared error	
System	SPE-07*	Field	M1	SPE-07	M1	SPE-07	M1
A	4.60	5.65	6.94	-19%	23%	1.10	1.67
B	3.65	4.34	4.65	-16%	7%	0.47	0.10
C	4.51	6.17	5.40	-27%	-13%	2.77	0.60
E	4.14	6.06	7.13	-32%	18%	3.69	1.14
F	5.21	5.34	7.69	-2%	44%	0.02	5.51
				MAPE		RMSE	
				19%	**21%**	**1.27**	**1.34**

* Average of SPE-07 COP at 77°F and 86°F

Figure 7-16. ***System A cooling COP by outdoor dry bulb temperature***

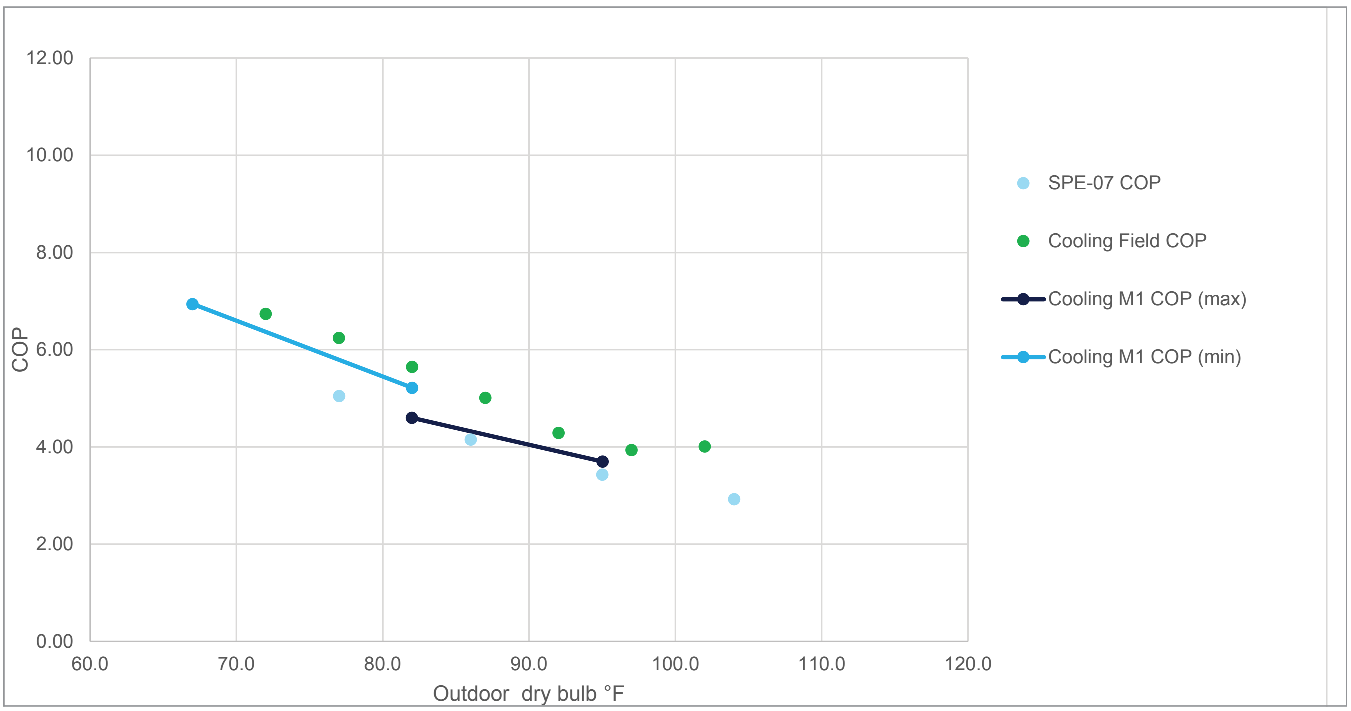


Figure 7-17. *System B cooling COP by outdoor dry bulb temperature*

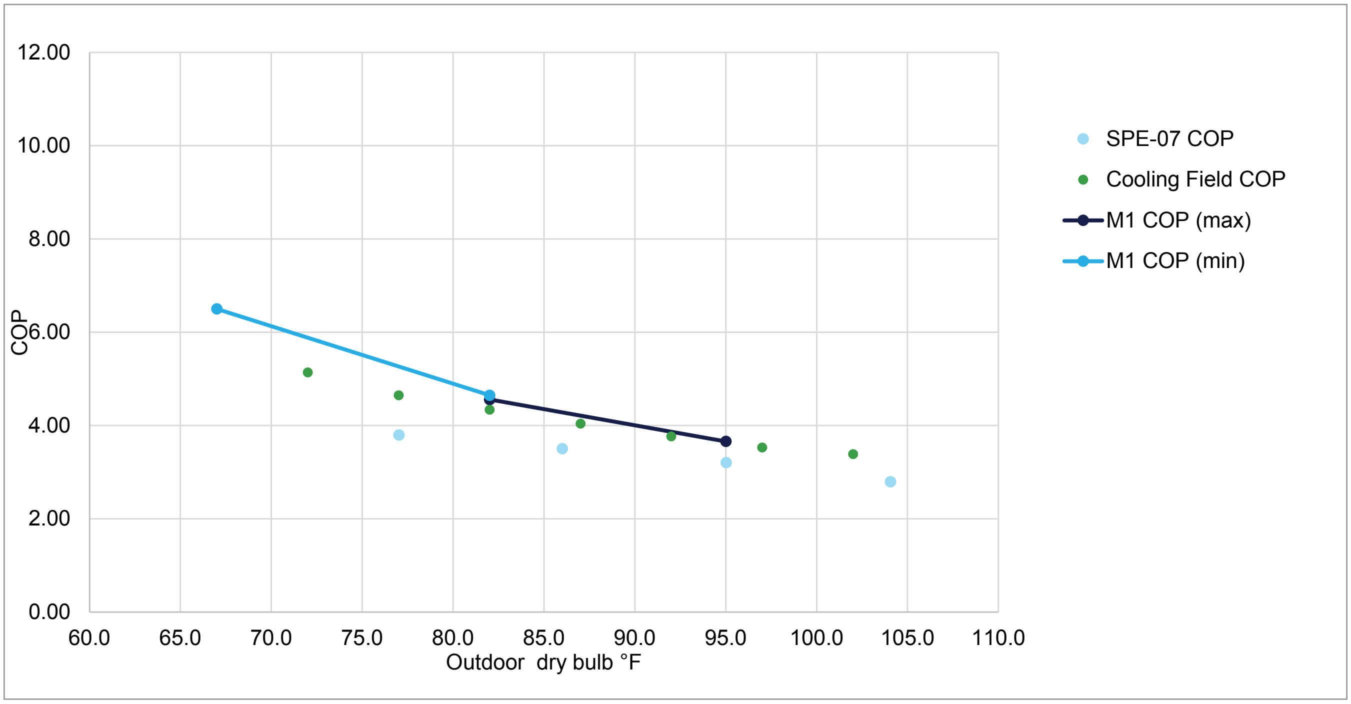


Figure 7-18. *System C cooling COP by outdoor dry bulb temperature*

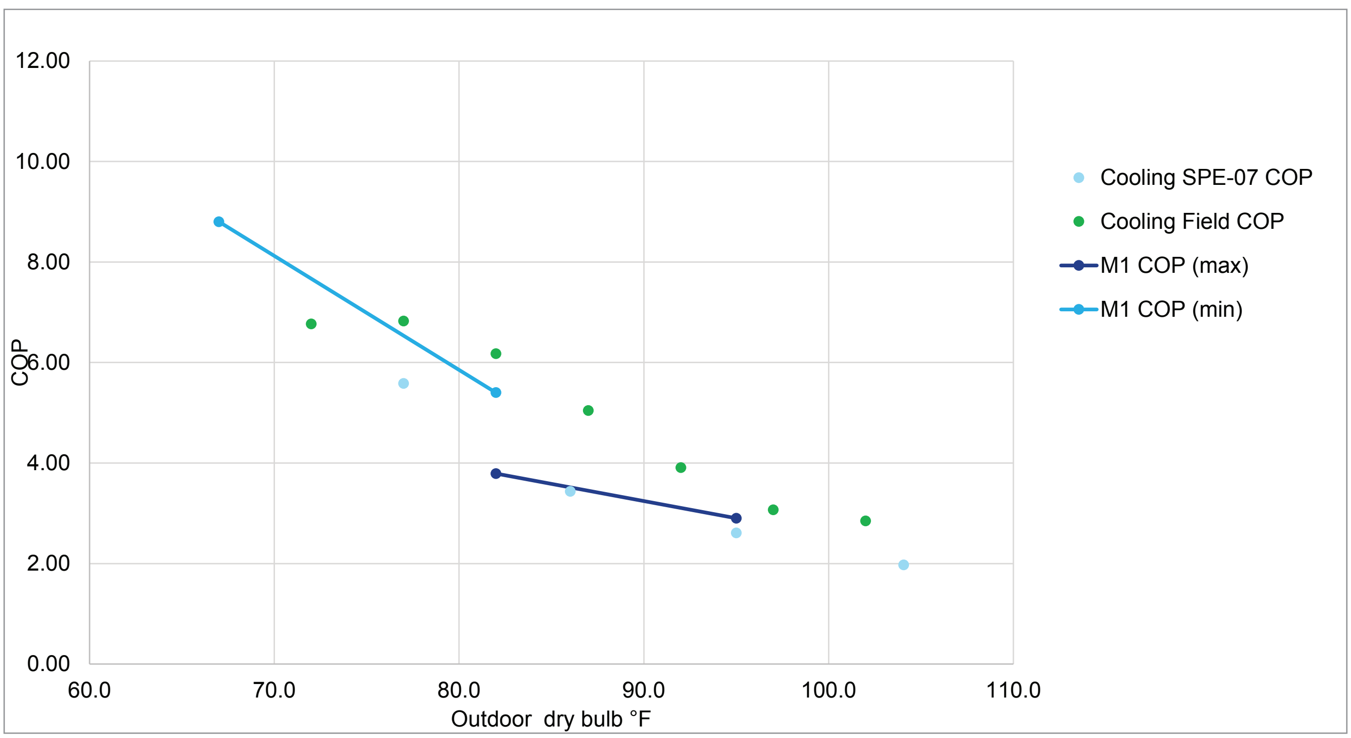


Figure 7-19. *System E cooling COP by outdoor dry bulb temperature*

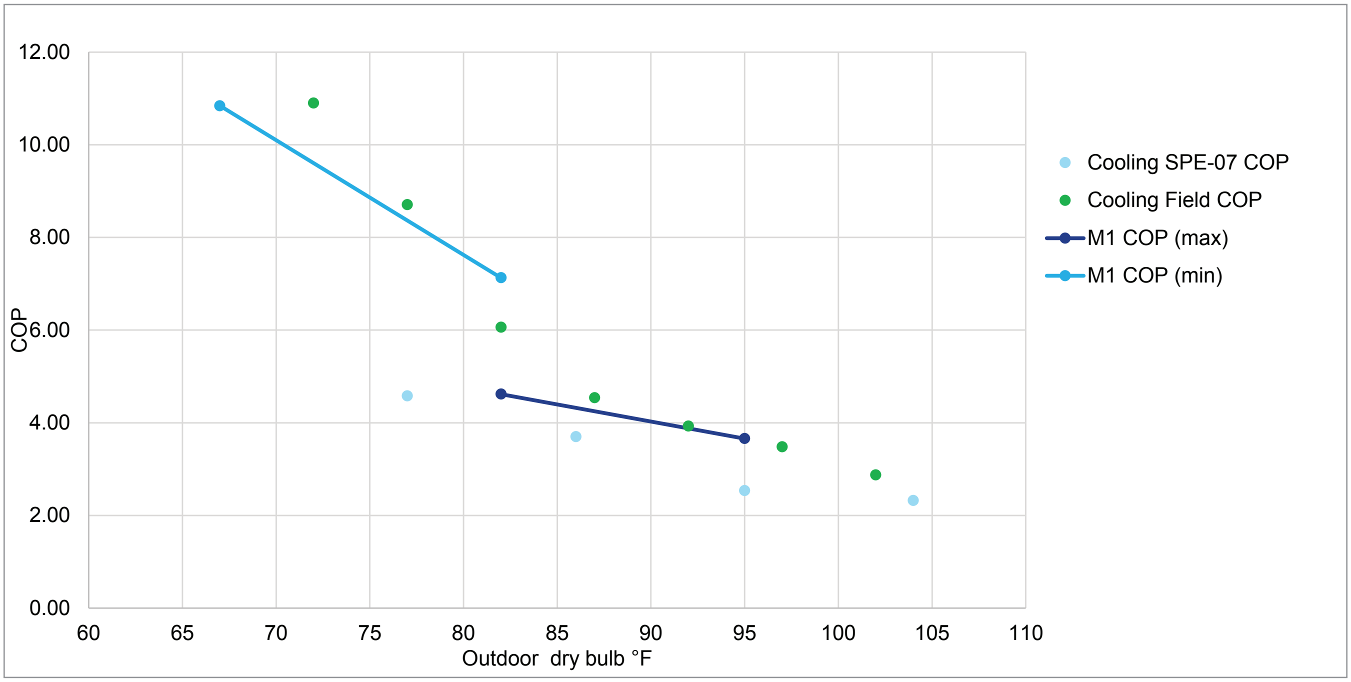
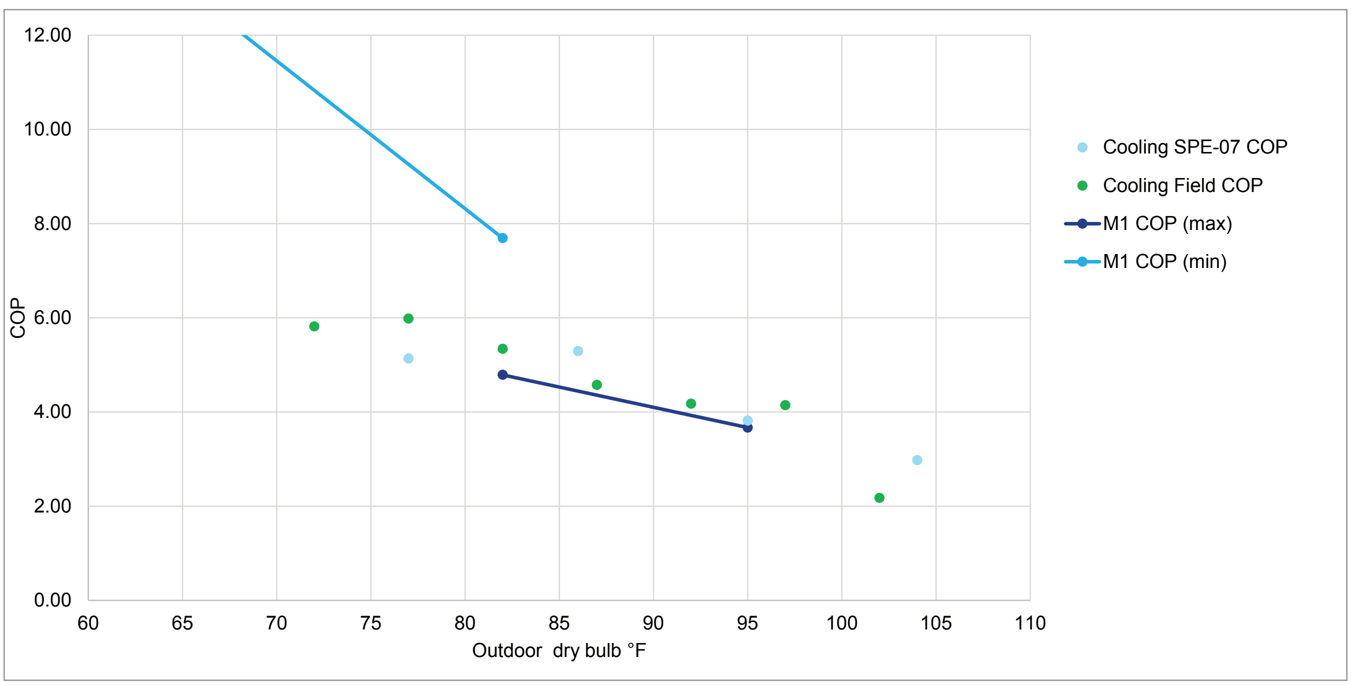


Figure 7-20. *System F cooling COP by outdoor dry bulb temperature*[15]



[15] The M1 minimum COP of 12.4 at 67 °F is off the chart.

Figure 7-21. *System A heating COP by outdoor dry bulb temperature*

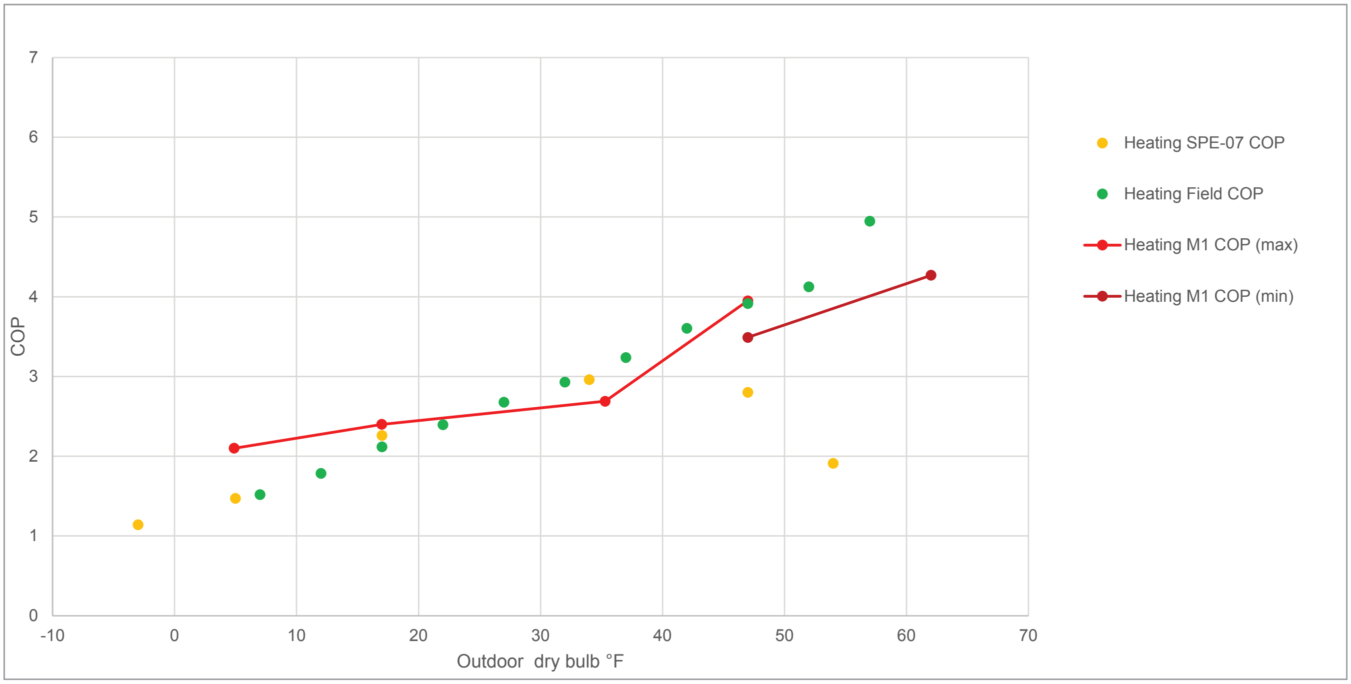


Figure 7-22. *System B heating COP by outdoor dry bulb temperature*

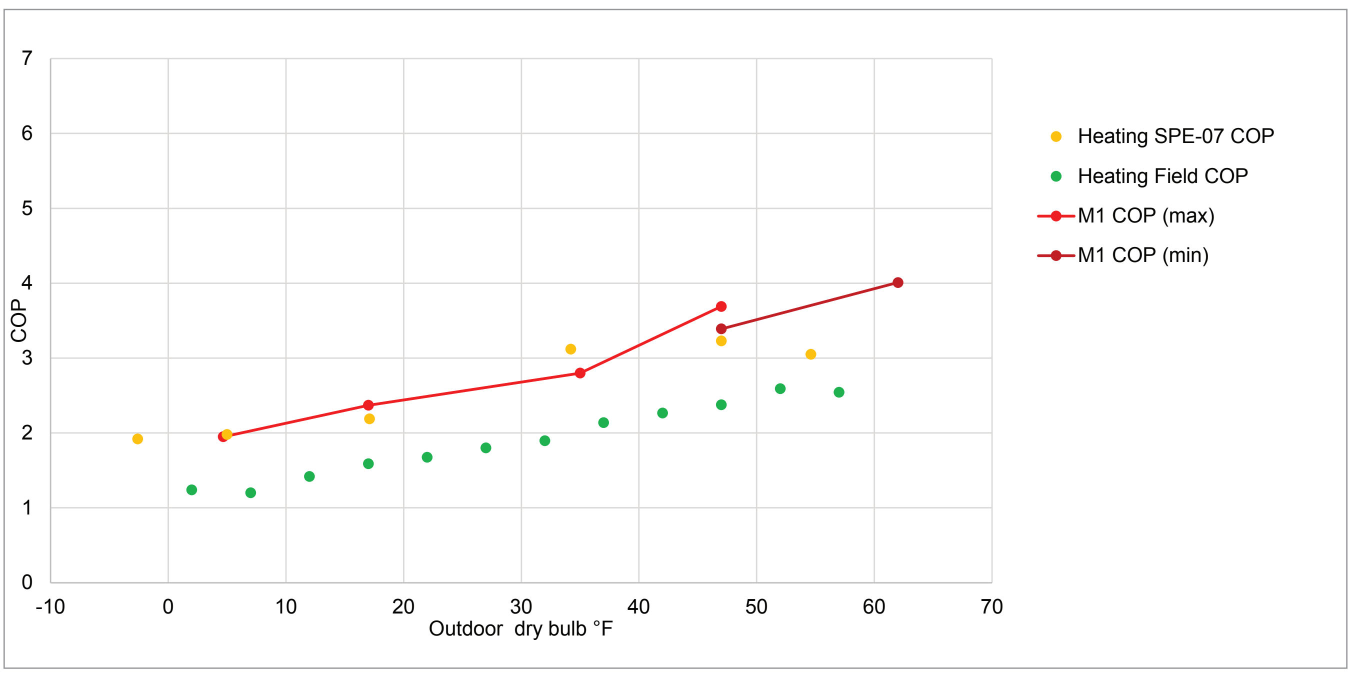


Figure 7-23. ***System C heating COP by outdoor dry bulb temperature***

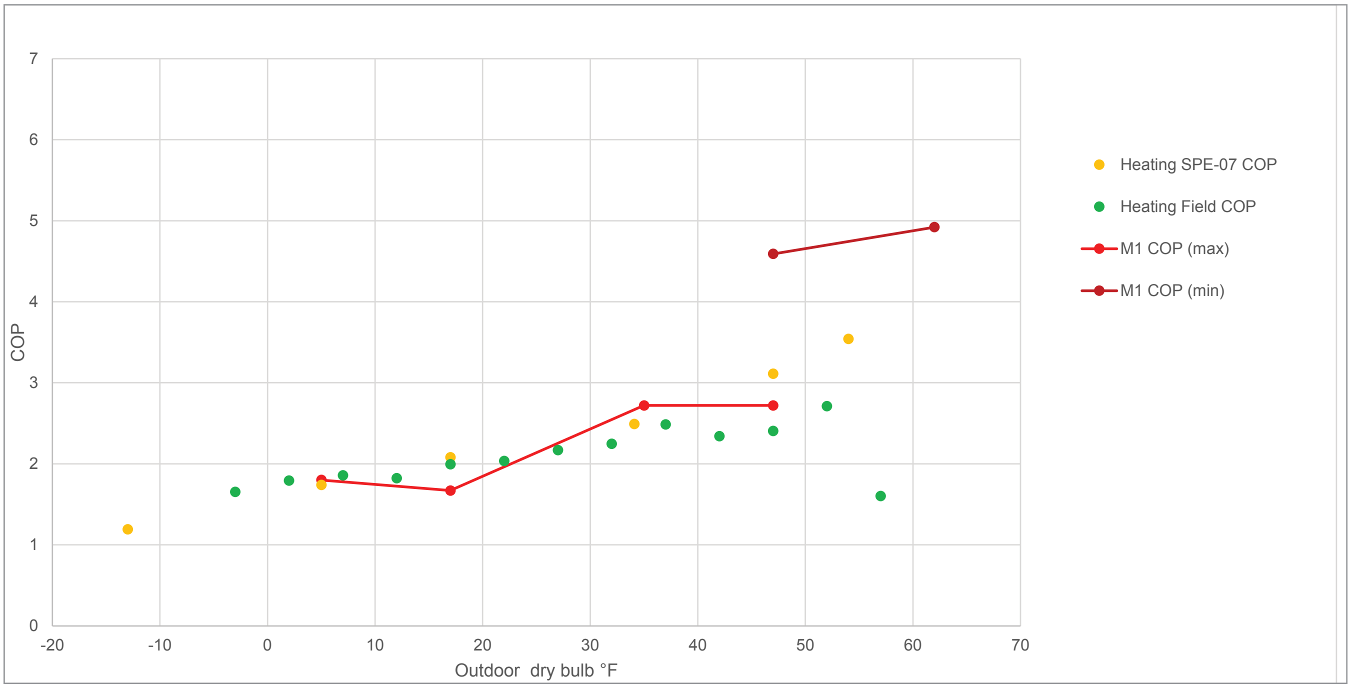


Figure 7-24. ***System E heating COP by outdoor dry bulb temperature***

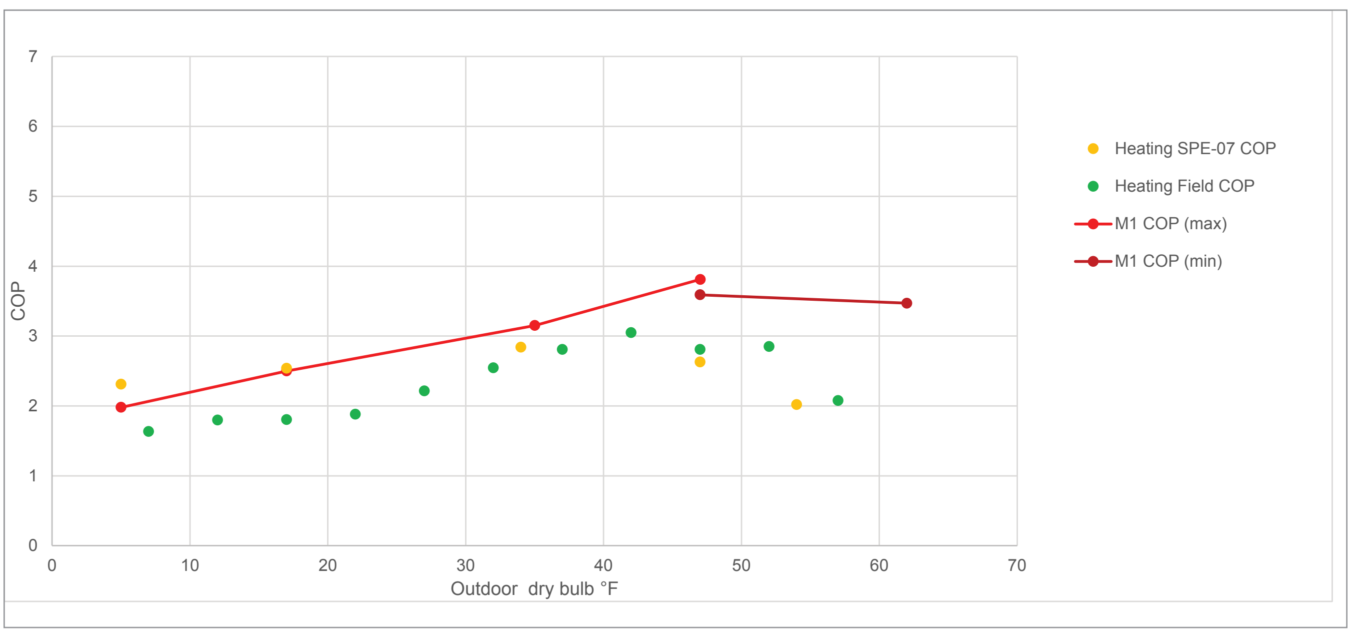
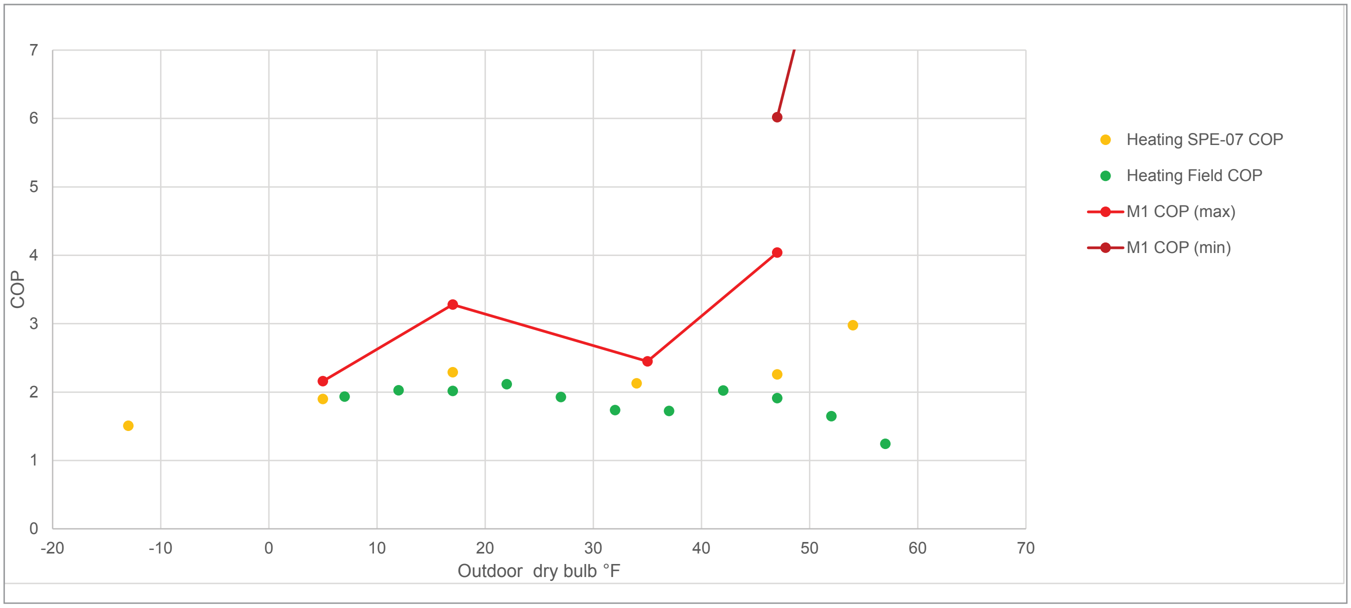


Figure 7-25. *System F heating COP by outdoor dry bulb temperature*



7.4: Additional characteristics investigation

To determine the characteristics that might be factored into heat pump efficiency metrics or enhance test procedures, the team investigated suspected characteristics or conditions that might affect field operation relative to laboratory operation. Many of the investigations were in response to specific requests from the oversight committee. These include instances where there is significant cycling, turndown ratio, and defrost energy.

7.4.1: Cycling

To investigate instances where there is significant cycling in field and in lab data, "frequent cycling" must first be defined. For this use, it is defined as the occurrence of power below the "off" threshold twice or more in a two-hour period. Using this definition, the number of cycles per two-hour period, the calculated average capacity, and the average total power input in each two-hour period when the system was cycling were obtained. Figure 726 to Figure 729 illustrate the average capacity and COP in each temperature bin when the system was cycling. System B and D are not plotted because B did not have any cycling periods, and D has insufficient field data. The COPs seemed to be most affected by cycling at the lower temperatures when cycling was occurring more frequently.

Figure 7-26. *Average capacity and COP during cycling by outdoor temperature bins for system A in cooling modewith field COP for all hours in each bin shown as labels.*

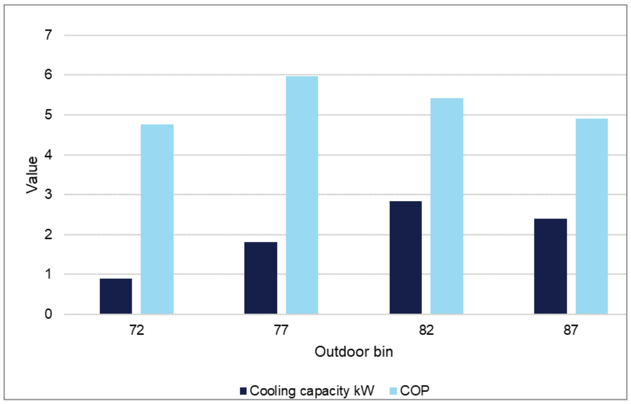


Figure 7-27. *Average capacity and COP during cycling by outdoor temperature bins for system C in cooling mode with field COP for all hours in each bin shown as labels.*

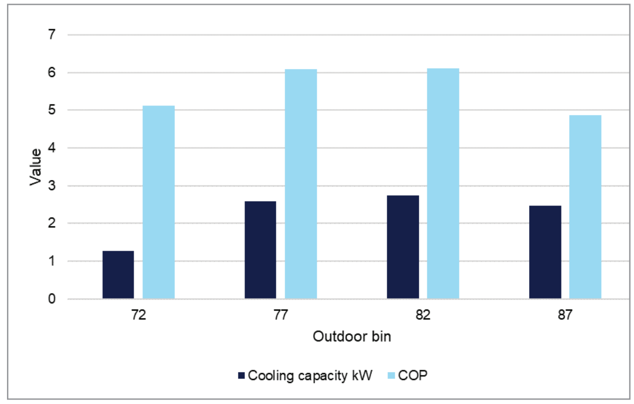


Figure 7-28. *Average capacity and COP during cycling in 72 °F temperature bin for system E in cooling mode with field COP for all hours in each bin shown as labels.*

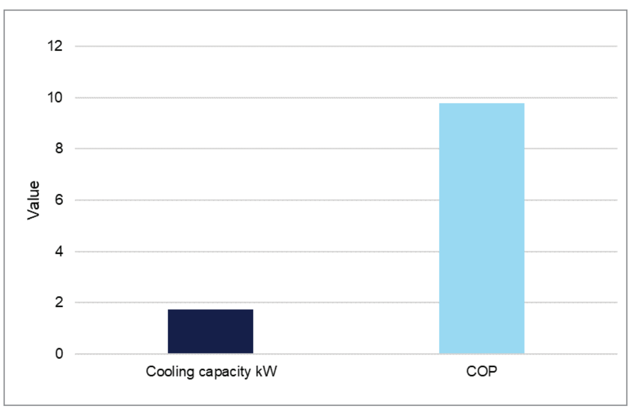
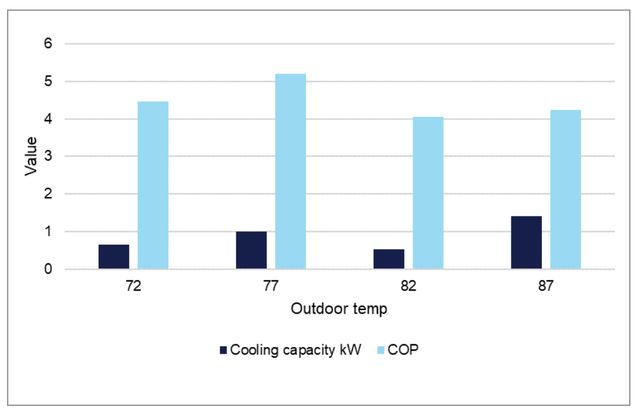


Figure 7-29. *Average capacity and COP during cycling by outdoor temperature bins for system F in cooling mode with field COP for all hours in each bin shown as labels*.



7.4.2: Turndown Ratio

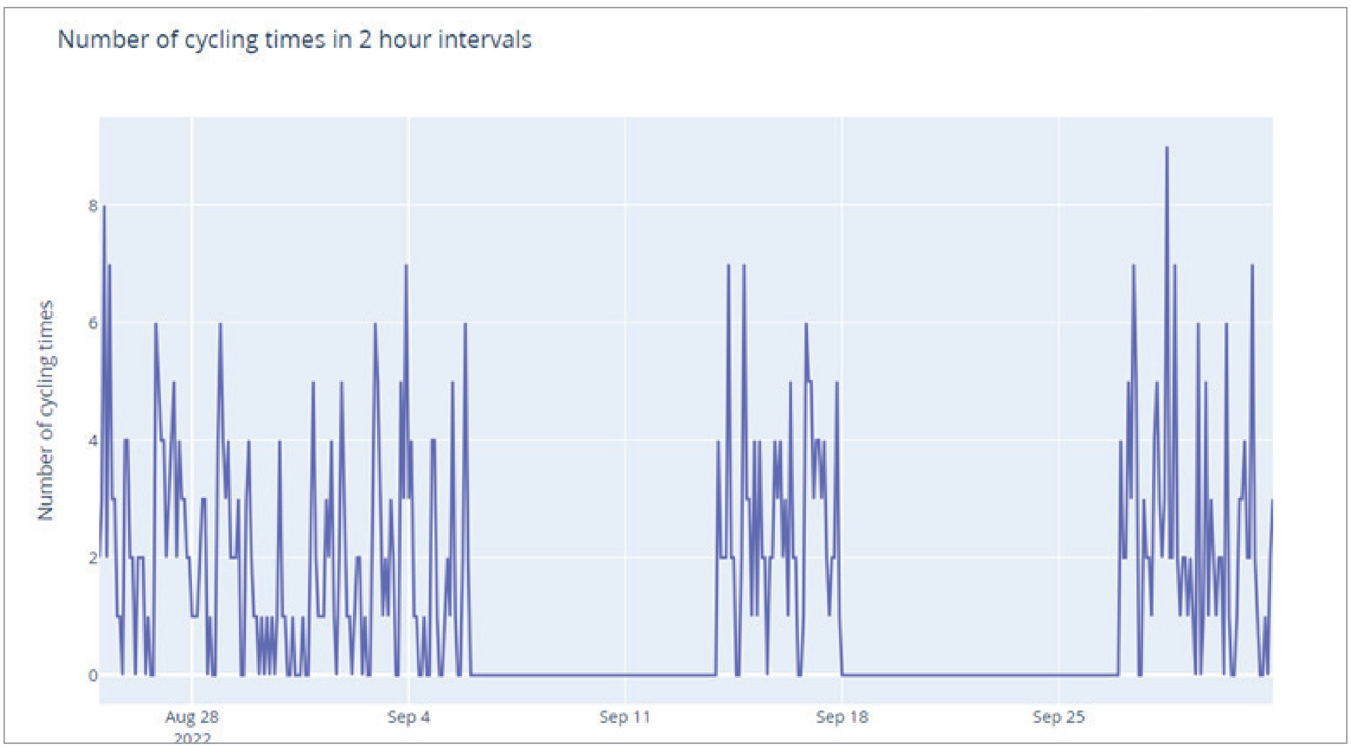
By inspecting the values of turn down ratio (TDR) over time, the highest TDR was obtained, as shown in Table 7-6. The turn down ratio was calculated as the lowest stable output compared to the highest stable output and is a descriptor of the system's ability to modulate capacity under low loads, to avoid cycling inefficiencies. There was a wide variation in turn down ratio: from only three to one for System C to a whopping 10 to one turn down ratio for System F. The two ductless systems (E and F) had higher turn down ratios than the ducted systems.

Table 7-6. ***The turndown ratios obtained from field running data for all five systems***

System	Turndown ratio
System A	6.0:1
System B	3.4:1
System C	2.9:1
System E	5.6:1
System F	10.0:1

To further understand typical low-load behavior, the number of cycles during each two-hour period was counted. Figure 7-30 shows the number of cycles per two-hour period for System A over a span of roughly a month in cooling mode. The largest number of cycles is nine for this system, but most of the cycling numbers are between one and six. There are two periods in the plot in which System A was not the unit being tested, so it had no cycling.

Figure 7-30. ***Number of cycles during each two-hour period for System A in cooling mode***



7.4.3: Defrost Energy

Figure 7-31 illustrates the duration of each defrost cycle, in minutes, when System A was operating in heating mode. The duration of the defrost cycle ranges from three to 11 minutes, depending on the operating conditions, and the longer periods and shorter periods of defrost each tend to cluster, consistent with weather conditions that require more or less defrosting. This indicates that the system has a demand defrost approach (not a timed defrost), and it shows the range of typical defrost periods.

Figure 7-31. ***Duration of each defrost cycle in heating mode of System A***

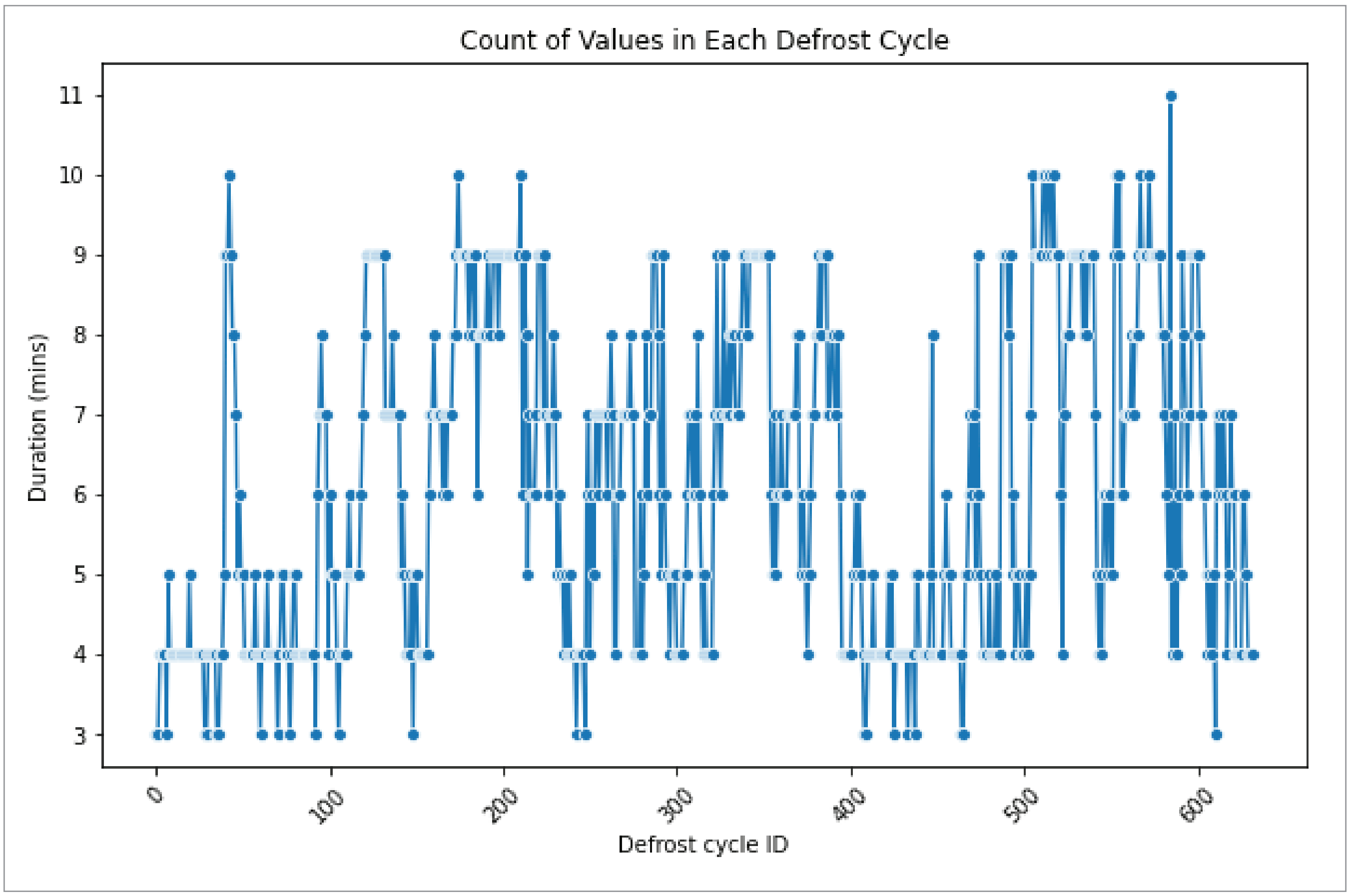


Figure 7-32 depicts the average duration and power input during defrost mode for system A. Although the data is noisy, the trend is for shorter defrost during warmer weather (which makes sense) and higher power draw during warmer weather (less intuitive). Possibly, the fact that there is more compressor lift during warmer weather (because we're essentially in cooling mode) and it's probably at a fixed speed, causing the power draw to be higher during the warmer temperature. Figure 7-33 shows System A's energy usage during a single cycle of defrost and non-defrost modes organized by temperature bins. A single defrost cycle's energy consumption ranges from 150 to 450 Btu but is not closely correlated with temperature. The non-defrost cycles are short when the outdoor temperatures are cold, limiting the energy consumption of those cycles. Non-defrost energy usage increases significantly as the outdoor temperature rises over 32 °F because the operating cycles can be much longer when the system does not enter defrost mode.

Figure 7-32. ***Defrost mode average duration and power input for system A for each temperature bin***

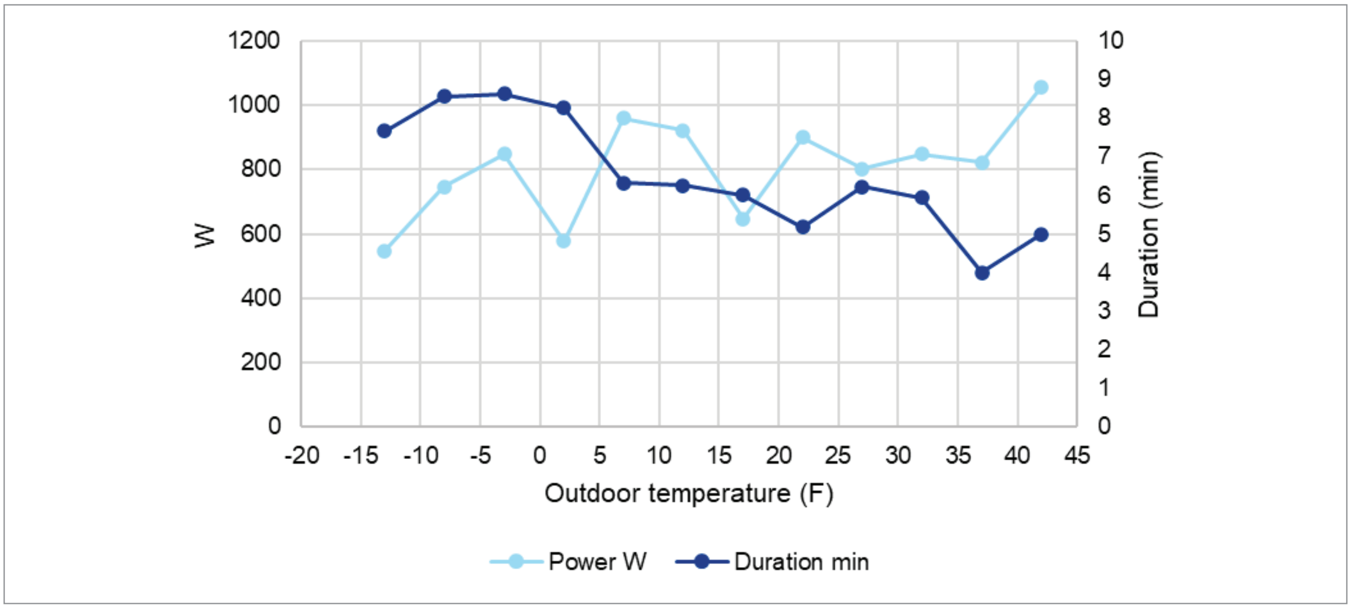
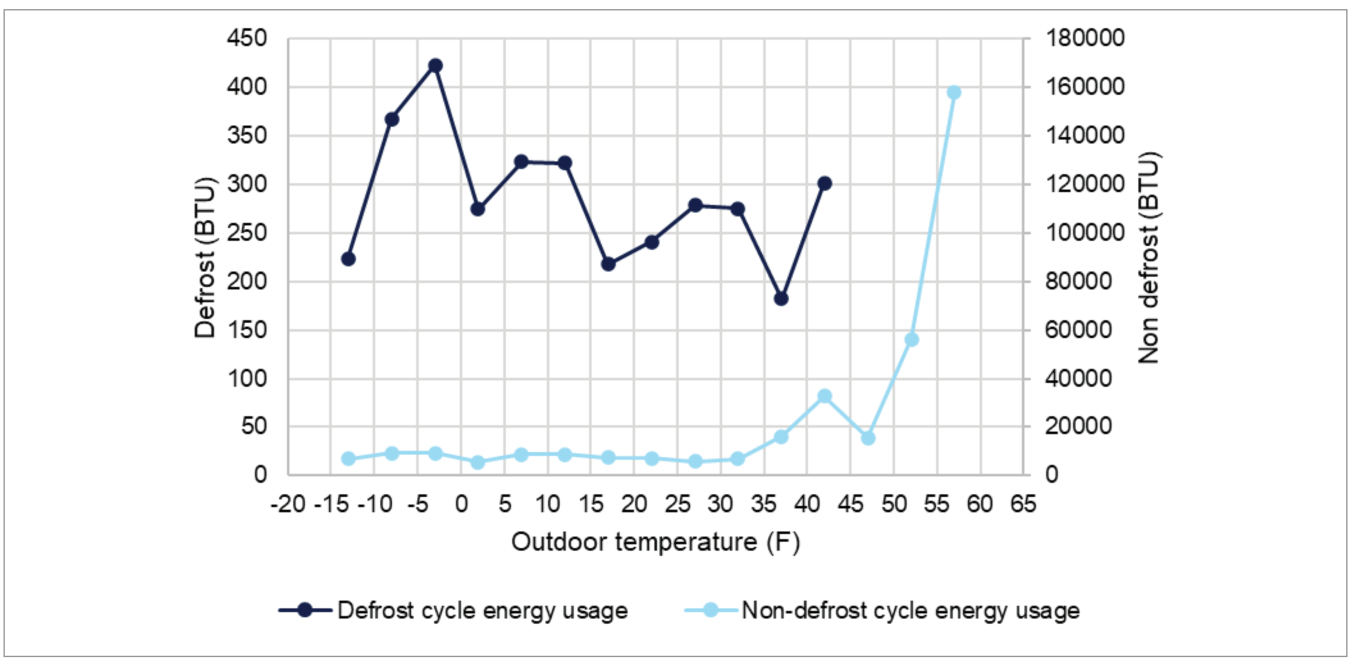


Figure 7-33. ***Defrost and non-defrost energy usage for a single cycle of system A for each outdoor temperature bin***



Section 8: CONCLUSIONS

Ultimately, this project achieved its objectives to build a robust set of rigorous and well-controlled field measurement data and to enable detailed comparisons of the field data to the ratings produced by SPE-07 and M1. At the outset of this project, the project partners and lead researchers expected the SPE-07 lab test procedure to outperform the DOE M1 test procedure when compared to field performance because the systems operate under their own controls in the SPE-07 test.

Although the sample size is small, because of the low variance in the repeatability tests[16], there is high confidence in the rating comparison to field measurements described in this report indicating that the SPE-07 test is more representative of field operation than M1 testing (normalized), using both the root mean squared error and the mean absolute error metrics.

Based on the mean absolute error percentage analysis, normalized M1 efficiency ratings for heat pumps diverge from actual performance by 22% for seasonal cooling efficiency and by 36% for seasonal heating efficiency. Based on the same analysis, SPE-07 ratings diverged by 13% on seasonal cooling ratings and 10% on seasonal heating ratings.

Based on the root mean squared error (RSME) measurement of difference, normalized M1 efficiency ratings for heat pumps produced a deviation of 1.40 for seasonal cooling efficiency and .93 for seasonal heating efficiency. Based on the same analysis, SPE-07 ratings produced a deviation of .82 for seasonal cooling ratings and a deviation of .24 on seasonal heating ratings.

The larger errors in M1 tests of ductless heat pumps suggest that the M1 testing procedure is less representative for ductless units than for ducted, although the sample size is too small to generalize the differences between ducted and ductless with much confidence. Even for the ducted systems, M1 more often overpredicts the performance of the units in both heating and cooling. In contrast, SPE-07 under-predicts efficiency in cooling mode fairly consistently, and in most cases slightly over-predicts in heating mode.

The results using both the root mean squared error and the mean absolute error metrics also show that the SPE-07 lab test is more representative of field operation at low heating loads than the M1 lab test. Particularly, System F (in both heating and cooling) and System C (in heating) had large differences between field and M1 performance for COPs tested at low loads.

Preliminary data from this study has already influenced the planned changes to the next M1 test procedure update. In the summer of 2023 AHRI convened members and efficiency stakeholders to develop updates to the AHRI 210/240 test procedure that DOE uses by reference. Specifically, the load equation and climate-based bin hours definition from the SPE-07 standard were adopted as well as a controls verification test that can be used to ensure the part load capabilities under a system's own controls.

---

[16] The SPE-07 repeatability tests performed on units C and F showed low variance across three repeated tests with average 95 percent confidence interval of ± 2.35 percent in cooling and ± 0.65 in heating.

Section 9: FURTHER RESEARCH

More exploration of SPE-07 or other load-based testing is warranted to confirm the results seen in this study. In addition, more exploration of the ability of laboratory tests to capture low load operation is recommended. Further research is planned to laboratory-test the six machines at part load to understand whether it is possible to use the H11 test (heating, minimum at 47°F) as a good indicator of heat pump performance at part load conditions.

For budgetary reasons, the control verification procedure (CVP) tests were not performed in the laboratory as part of this study, so this project does not include a comparison between CVP results and field operating results. We recommend that this work be performed at some point in the future. The CVP testing investigation task would include working with the lab to determine periods during each heat pump testing where CVP tests can be run without compromising the timeline and budget for the primary lab tests. The technical advisory committee will also need to determine how to prioritize which CVP tests (e.g., low load 47 °F, low load 17 °F, max speed 5 °F, etc.) will be performed on which heat pumps, time permitting. Comparison with field operating results requires that the heat pumps experience similar operating conditions as the CVP tests. CVP tests have been adopted by EPA and DOE for heat pump programs (ENERGY STAR and Cold Climate Challenge) and regulatory tests (AHRI 1230). Obtaining CVP results for specific units whose behavior and performance have also been observed/measured in the field will be useful in evaluating the CVP concept for possible broader adoption for residential single-zone systems.

Appendix A: DATA ANALYSIS DETAILS

The SPE-07 bin factors are provided in Table A-1 and A-2 below. Note: These test temperatures do not align exactly with the bin centers that are reported in the tables above. The test conditions for SPE07 (and M1) were never meant to correspond exactly with the bin centers - there are so many more bins than test temperatures. This is true for heating as well, where the 17 and 47 test points line up with the bin centers, but 5 and 34 do not.

Table A-1. ***Bin factors provided in SPE-07 for cooling SCOP calculations***

COOLING										
	Outdoor DB bin (°F)	**DB bin range (°F)**	**Subarctic**	**Very cold**	**Cold/dry**	**Cold/ humid**	**Marine**	**Mixed**	**Hot/ humid**	**Hot/ dry**
1	72	<74.5	N/A	0.336	0.289	0.316	0.335	0.284	0.19	0.213
2	77	74.5–79.5	N/A	0.192	0.154	0.21	0.137	0.232	0.305	0.143
3	82	79.5–84.5	N/A	0.202	0.157	0.209	0.137	0.199	0.255	0.154
4	87	84.5–89.5	N/A	0.162	0.138	0.147	0.104	0.15	0.146	0.131
5	92	89.5–94.5	N/A	0.089	0.172	0.095	0.154	0.1	0.081	0.163
6	97	94.5–99.5	N/A	0.016	0.076	0.019	0.094	0.029	0.019	0.109
7	102	99.5–104.5	N/A	0.002	0.013	0.005	0.028	0.006	0.003	0.058
8	107	104.5–109.5	N/A	—	0.002	—	0.007	0.001	—	0.025
9	112	>109.5	N/A	—	—	—	0.002	—	—	0.004

Table A-2. Bin factors provided in SPE-07 for heating SCOP calculations

HEATING										
	OD DB bin (°F)		Subarctic	Very cold	Cold/ dry	Cold/ humid	Marine	Mixed	Hot/ humid	Hot/ dry
1	−23	< −20.5	0.043	0.004	—	—	—	—	—	—
2	−18	−20.5 to −15.5	0.024	0.006	0.001	0.002	—	—	—	—
3	−13	−15.5 to −10.5	0.028	0.009	0.002	0.003	—	—	—	—
4	−8	−10.5 to −5.5	0.036	0.015	0.004	0.007	—	—	—	—
5	−3	−5.5 to 0.5	0.038	0.019	0.005	0.01	—	0.001	—	—
6	2	0.5–4.5	0.05	0.028	0.01	0.02	—	0.003	—	0.001
7	7	4.5–9.5	0.05	0.035	0.014	0.029	—	0.007	—	0.002
8	12	9.5–14.5	0.051	0.045	0.025	0.041	—	0.015	0.001	0.004
9	17	14.5–19.5	0.062	0.061	0.047	0.062	0.002	0.033	0.005	0.011
10	22	19.5–24.5	0.057	0.067	0.064	0.071	0.005	0.046	0.015	0.022
11	27	24.5–29.5	0.089	0.102	0.12	0.116	0.019	0.094	0.049	0.054
12	32	29.5–34.5	0.123	0.136	0.162	0.15	0.061	0.137	0.098	0.093
13	37	34.5–39.5	0.114	0.145	0.171	0.151	0.147	0.171	0.165	0.145
14	42	39.5–44.5	0.076	0.108	0.125	0.104	0.199	0.147	0.173	0.161
15	47	44.5–49.5	0.082	0.108	0.123	0.114	0.282	0.161	0.217	0.218
16	52	49.5–54.5	0.055	0.077	0.079	0.078	0.205	0.118	0.171	0.175
17	57	> 54.5	0.022	0.034	0.047	0.041	0.079	0.065	0.105	0.114

Appendix B: LAB AND FIELD COMPARISON CHARTS

The following charts show the comparisons between the field and lab tests across all of the SPE-07 climate zones. The orange "actual" values are the unadjusted COPs measured in the field. These are always shown in the cold/humid climate because that was the climate in which the measurements were made. The red "field" values were normalized as described in section 7. The blue "SPE07" values were from the lab tests which used the adjusted field load line in the testing procedure. The purple "M1 SEER2/3.41" are the M1 lab results, not normalized according to the procedure in section 7, but divided by 3.41 to make it unitless and comparable to a COP value. Finally, the blue "rated/3.41" values use the AHRI rated SEER2 or HSPF2 values for each system again divided by 3.41 to make it unitless and comparable to a COP value. The M1 and the Rating values are shown in the mixed climate for cooling and in the cold/dry climate for heating because those are the SPE-07 climates most similar to the M1 Region IV used for those lab tests. The cooling charts are shown first, followed by the heating charts.

Figure B-1. ***Cooling seasonal cops compared – Unit A***

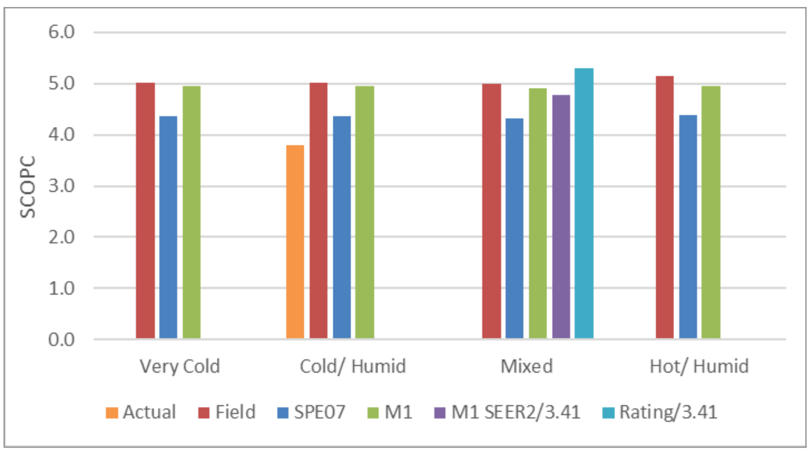


Figure B-2. ***Cooling seasonal COPs compared – Unit B***

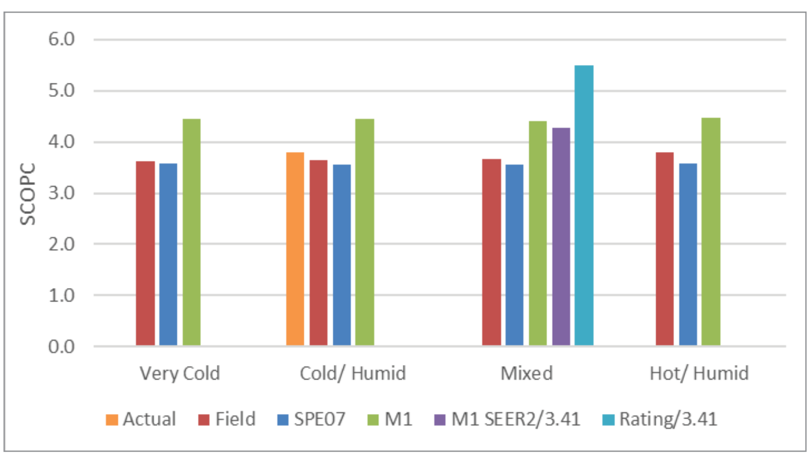


Figure B-3. ***Cooling seasonal COPs compared – Unit C***

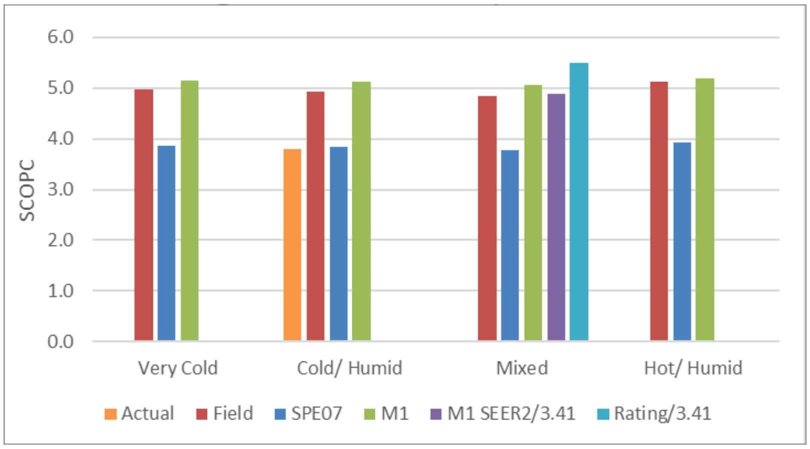


Figure B-4. ***Cooling seasonal COPs compared – Unit E***

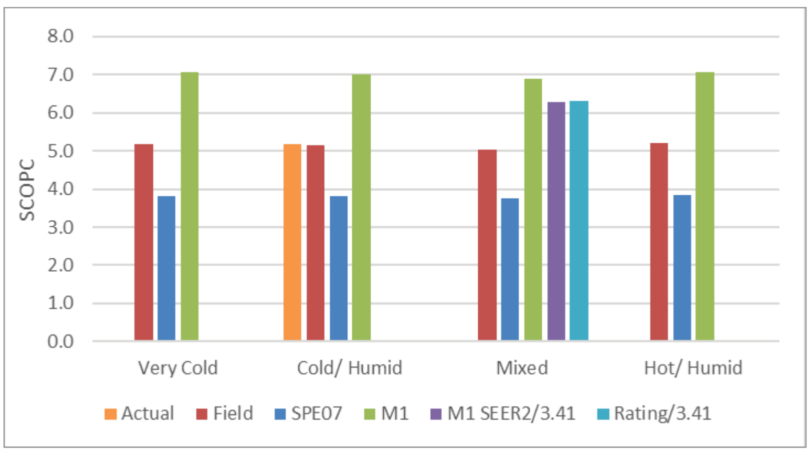


Figure B-5. ***Cooling seasonal COPs compared – Unit F***

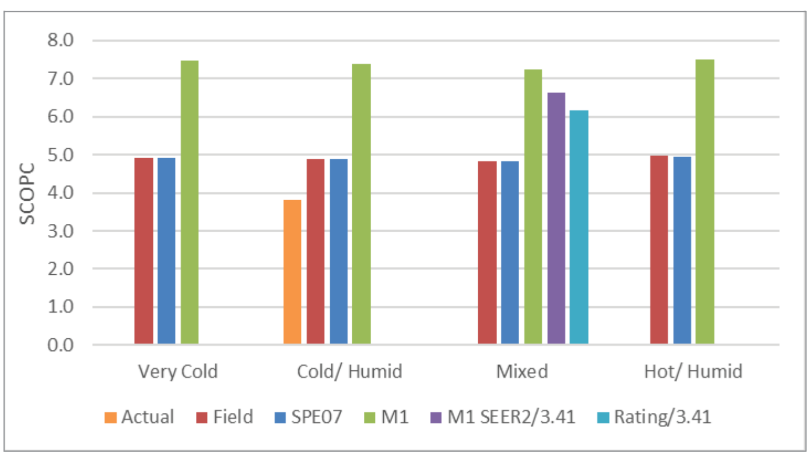


Figure B-6. *Heating seasonal COPs compared – Unit A*

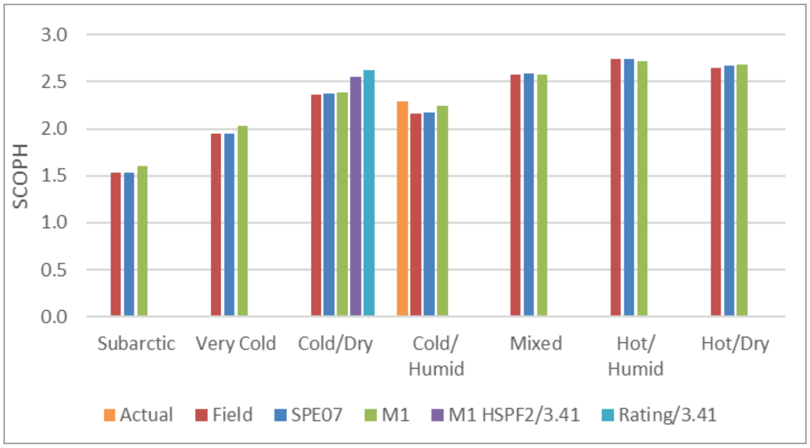


Figure B-7. *Heating seasonal COPs compared – Unit B*

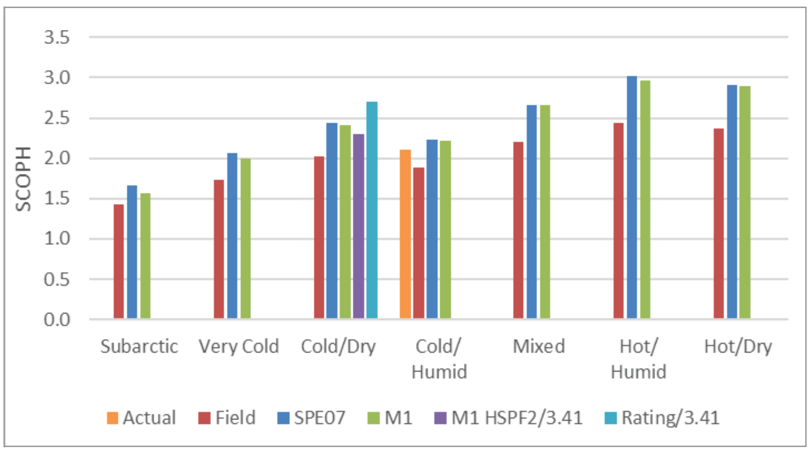


Figure B-8. *Heating seasonal COPs compared – Unit C*

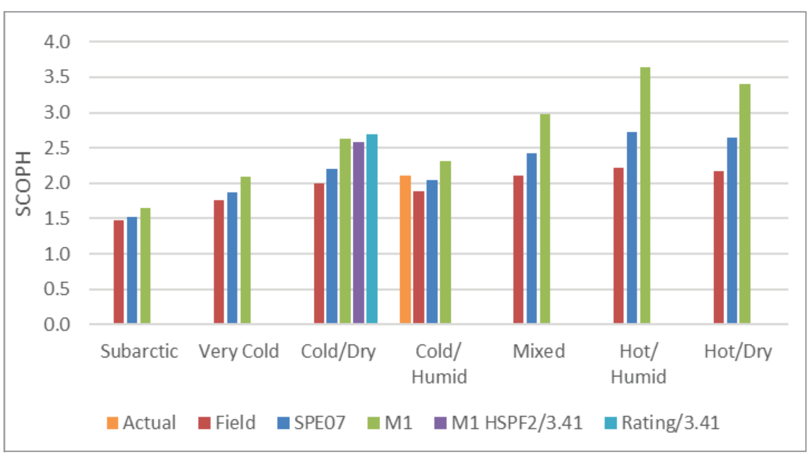


Figure B-9. ***Heating seasonal COPs compared – Unit E***

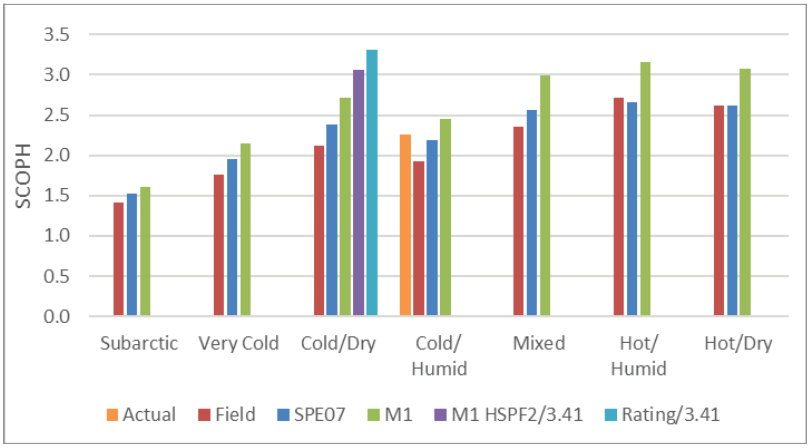


Figure B-10. ***Heating seasonal COPs compared – Unit F***