Natural Resources Canada / Ressources naturelles Canada

CanmetENERGY

Evaluating the Reproducibility of CSA SPE-07:23 Using Two Centrally Ducted Air-Source Cold-Climate Heat Pumps

Prepared by:

Jean-Philippe B. Poirier, P.Eng., Research Engineer, CanmetENERGY-Ottawa

Jeremy Sager, M.Eng., Senior HVAC and Renewables Research Engineer, CanmetENERGY-Ottawa

Bruce Harley, Principal, Bruce Harley Energy Consulting LLC

Prepared for:

Natural Resources Canada - Office of Energy Efficiency – Equipment Division

Northwest Energy Efficiency Alliance

30 April 2025

"CanmetENERGY- Ottawa leads the development of energy science and technology solutions for the environmental and economic benefit of Canadians."

Canada

Disclaimer

Neither Natural Resources Canada nor any of its employees makes any warranty expressed or implied or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of this report's contents. Reference in the report to any specific commercial product, process, service or organization does not necessarily constitute or imply endorsement, recommendation or favouring by Natural Resources Canada. The views and opinions of the authors expressed in this report do not necessarily state or reflect those of Natural Resources Canada.

Funding for this work was provided by Natural Resources Canada through the Program of Energy Research and Development (PERD), as well as the Office of Energy Efficiency (OEE). Additional funding was provided by the Northwest Energy Efficiency Alliance (NEEA). The heat pumps tested were provided by the Northeast Energy Efficiency Partnerships (NEEP).

Cat. No. M154-199/2025E-PDF

ISBN: 978-0-660-79839-4

Acknowledgements

The authors would like to acknowledge Liam Jones and Cory Balderson from Natural Resources Canada / CanmetENERGY-Ottawa for their help setting up the test facility, installing the heat pumps, and conducting the tests at CanmetENERGY in Ottawa, ON.

The authors would also like to acknowledge the help received from Mark Baines and Titus Mowry of UL Solutions who provided invaluable insights into testing conducted according to CSA SPE-07:23.

Executive Summary

In this report, CanmetENERGY-Ottawa evaluated the reproducibility of the CSA SPE-07:23 (SPE-07) test procedure using two centrally ducted cold-climate air-source heat pumps (CC-ASHP). Both CC-ASHPs were 2-ton variable speed systems with communicating thermostats that had previously been field and lab tested for evaluations of representativeness and repeatability. The objective was to produce heating and cooling seasonal COPs ($SCOP_C$ and $SCOP_H$) from a complete series of SPE-07 rating tests using the same building load line and equipment installation as that used during previous testing conducted in another laboratory.

The first CC-ASHP tested (Unit A) achieved an overall reproducibility of “fair” with 8.6% mean difference in $SCOP_C$, and 8.8% mean difference in $SCOP_H$ between facilities. Unit A was consistent in reproducing individual test part-load and full-load behaviour under matching test conditions, which reflects favourably on the procedure’s ability to consistently induce an equipment response between testing facilities.

The second CC-ASHP tested (Unit C), achieved an overall reproducibility of “poor” with 15.8% mean difference in $SCOP_C$, and 9.8% mean difference in $SCOP_H$ between facilities. Unit C testing at Natural Resources Canada (NRCan) experienced issues with thermostat temperature sensor drift that negatively impacted equipment behaviour. Furthermore, Unit C experienced cut-offs at the lowest catalogued temperature heating test interval (-25°C), resulting in no useable data for this test condition. For these reasons, it is difficult to draw conclusion from Unit C testing on the reproducibility of the SPE-07 procedure.

There were several factors affecting reproducibility identified during testing, including: (a) instrumentation lag differences between testing facilities; (b) potential indoor chamber response lag in Lab A; (c) issues in properly measuring unit under test (UUT) capacity during defrost operations in Lab A; (d) installation differences between facilities; (e) equipment and thermostat dynamic response differences between labs; (f) an unverifiable thermostat setup in Lab A (for Unit A); and (g) thermostat temperature sensor drift in the NRCan lab (for Unit C). Several recommendations were made to the development committee of SPE-07 to help improve reproducibility as a result of this project.

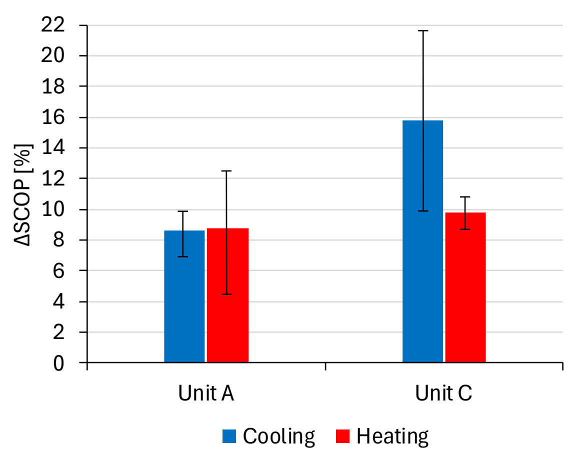


Figure E.1. Mean difference and standard deviation in $SCOP_C$ and $SCOP_H$ for Unit A and Unit C

Table of Contents

List of Figures

List of Tables

Glossary and Abbreviations

AHRI	American Heating and Refrigeration Institute
ASHRAE	American Society of Heating, Refrigeration and Air Conditioning Engineers
CC-ASHP	Cold-climate air-source heat pump
CCTF	Climate controlled test facility
CE-O	CanmetENERGY-Ottawa
COP	Coefficient of performance
CSA	Canadian Standards Association
EER	Energy Efficiency Ratio
EXP	“Express Publication”
HSPF	Heating Seasonal Performance Factor
LCT	Lowest catalogued temperature
NRCan	Natural Resources Canada
$Q_C(95)$	Cooling capacity at 95°F
$Q_H(47)$	Heating capacity at 47°F
RAT	Room air temperature
$SCOP_C$	Cooling seasonal coefficient of performance
$SCOP_H$	Heating seasonal coefficient of performance
SEER	Seasonal Energy Efficiency Ratio
SHR	Sensible heat ratio
SPE	“Special Publication”
T_{ID}	Target indoor temperature
T_{ZL}	Zero-load temperature
UUT	Unit under test

1.0. Introduction

1.1. Background to the Development of Load-Based, Climate Specific Testing

Natural Resources Canada (NRCan) worked with partners across Canada and the United States to develop load-based testing and climate-specific rating procedures beginning in 2015. The goal was to develop a testing procedure that would require heat pumps to operate in all operating modes, such that performance results would encompass steady-state and part-load cyclical operation while also taking into consideration defrost cycles and stand-by energy consumption. The seasonal results would take into consideration the full range of outdoor temperatures encountered in climate regions across the United States and Canada. The test procedure would utilise the manufacturer's controls in the same way they would be installed at the point of use, thus attempting to directly represent how equipment would behave when installed in the real-world. The product of this work was CSA EXP-07, published in 2019.

Extensive lab testing using EXP-07:19 and SPE-07:23 beginning in 2018 revealed substantial differences between the fixed-speed ratings and load-based tests and provided evidence that these differences are largely driven by the behavior of built-in equipment control algorithms that are in use during the load-based testing. Stakeholders have suggested that more in-depth understanding of the differences between the two test procedures would require understanding of the reproducibility and repeatability of both methods.

In 2020 and 2021, Purdue University undertook a study of reproducibility and repeatability based on CSA EXP-07:19 and AHRI 210/240(2020). [1] A total of five units were tested. Some were tested multiple times in the same lab using both test procedures, to assess repeatability. Some were tested in multiple labs using both test procedures to assess reproducibility.

For reproducibility assessment, the same unit is tested in one lab, then shipped to another lab (or multiple labs) and the test procedure is repeated. The variation in results (using the seasonal efficiency metric) is analyzed statistically to determine the test-to-test variability, generally at a 95% confidence interval. Audit testing for enforcement purposes are nearly always conducted by a different lab than tests used for rating purposes, so large variations in reproducibility are undesirable, because they require ratings to be reported much more conservatively in order to ensure that enforcement tests remain within the allowed tolerances.

Generally, the CSA EXP-07:19 tests had repeatability variations of less than 5% (at 95% confidence) and were similar to the steady-state test repeatability, in instances where the unit was not broken down and re-installed in the same lab between tests. The reproducibility, however, was less consistent, with 6% to over 10% average variation on some units using CSA EXP-07, compared with 1.5% to 4.5% variation using AHRI 210/240. Generally, a 5% variation would be considered "good", 5-10% would be considered "fair", and over 10% would be "poor". A root-cause analysis identified a number of concerns in the test procedure itself that were considered in the update from CSA EXP-07:19 to CSA SPE-07:23 (hereafter referred to as SPE-07), with the belief that these updates would likely improve reproducibility by reducing the likelihood of variations in setup and implementation of the process between test labs. In the report, it was recommended to repeat an assessment of reproducibility using the SPE-07 procedure to estimate the extent of improvement, if any.

The Reproducibility testing in this project was conducted to address these concerns.

1.2. Introduction to "3Rs" Evaluation of CSA SPE-07

SPE-07 was published in 2023, and the Canadian Standards Association (CSA) filed a Notice of Intent for the procedure to become nationally accredited through the Standards Council of Canada (SCC) and the American National Standards Institute (ANSI). Before SPE-07 could become an accredited standard, demonstration of the "3 R's" (Representativeness, Repeatability and Reproducibility) would be needed.

Although the development and assessment of fixed-speed testing over past decades has been focused on repeatability and reproducibility, there had previously been no direct investigation of whether fixed-speed or load-based testing would more closely approximate real-world performance. A project to evaluate the representativeness of SPE-07 was subsequently coordinated by the Northeast Energy Efficiency Partnerships with support from member organisations of the Advanced Heat Pump Coalition (including NRCan). The project began with field testing of six different heat pumps in Lincoln, Nebraska through the winter and summer seasons ending in early 2023. These heat pumps were then shipped to Lab A where they underwent testing to both SPE-07 and AHRI 210/240 procedures. While this project included an evaluation of both representativeness and repeatability, it lacked data on the reproducibility (lab to lab results) of SPE-07. Reproducibility testing was therefore required to complete the "3R's" needed for SPE-07 to become a nationally accredited test standard and is the subject of this report.

NRCan's CanmetENERGY-Ottawa (CE-O) committed to contributing to reproducibility testing via use of its Climate Controlled Test Facility (CCTF). To prepare the facility for load-based testing according to the SPE-07 procedure, CE-O undertook extensive refurbishments to the CCTF (the CCTF had been previously used for exploratory testing according to EXP-07:19, the predecessor to SPE-07).

1.3. Evaluation of the Representativeness of CSA SPE-07:23

Phase 1 to evaluate representativeness was conducted in Lincoln, Nebraska, USA from August 2022 to March 2023. The heat pumps were installed in three identical, unoccupied mobile homes in a residential neighborhood. Each test house was outfitted with two heat pumps, one ducted and the other ductless, the use of which was alternated each week to provide space cooling and heating throughout the study period. All the heat pumps were variable-speed, cold-climate units except for one of the ducted systems (Unit B) which was a 2-speed, non-cold-climate system. The heat pumps operated in cooling mode from mid-August 2022 and were subsequently changed over to heating mode in late October. In March 2023 the tested systems were uninstalled and shipped to the laboratory for Phase 2. The field testing included a wide range of load conditions for both heating and cooling operation, with hourly outdoor temperatures ranging from -25 to 25.6 °C (-13 to 78 °F) during the heating test period and from 5.6 to 39.4 °C (42 °F to 103 °F) during the cooling period.

The test houses consisted of three manufactured homes of identical design, built in 2021 at the same factory, sited in the same neighborhood and orientation to the compass. Each house has

113 m² (1,220 ft²) floor area on one story (Figure 1), with low-slope vaulted ceilings and are insulated to current code requirements for manufactured homes.

The test houses were not occupied. They were outfitted with heaters to provide scheduled sensible internal gains during the entire field study, and humidifiers for latent internal gains (in the cooling season only) to approximate occupied conditions.

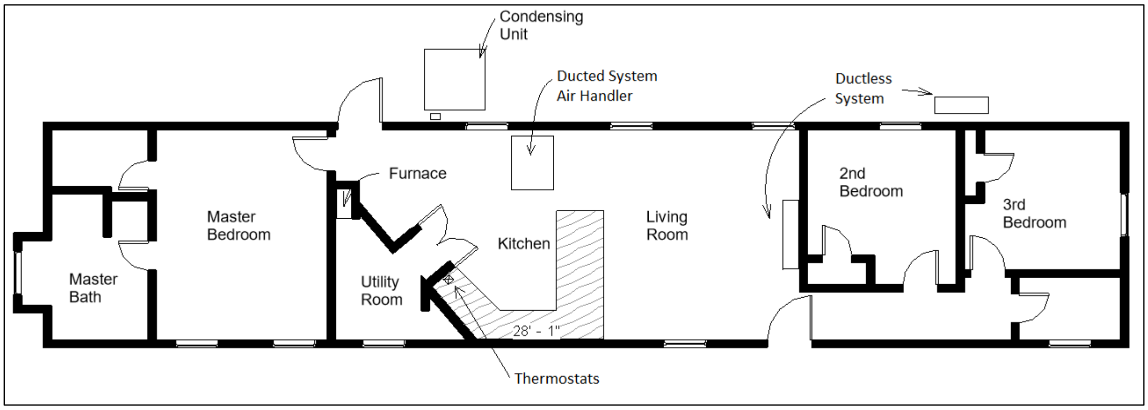


Figure 1. Field-test house layout

The intention was that the houses have consistent heating and cooling loads among them, and to be reasonably close to the SPE-07 and AHRI 210/240 heating and cooling load lines for the selected variable-speed units. There was some concern that the heating loads might not be enough to fully utilize the test units' low temperature capacity, so the heat loss coefficients were estimated using electric resistance heating tests. The target UA value for SPE-07 was 199 W/K (377 Btu/(hr·°F)). Measurements of the three houses ranged from 160 to 169 W/K (303 to 320 Btu/(hr·°F)), so modifications were made to the homes to increase the heat transfer. In each house, four of the double-glazed, low-e window sashes were replaced temporarily with single sheets of galvanized steel, and some of the homes had part of the insulation and air barrier materials, as well as skirting around the perimeter, removed on their undersides.

Furthermore, because the homes were unfurnished, it was felt that the shallow thermal mass capacitance (which is a parameter used in the load-based test procedure) should be estimated with field measurements. The mass capacitance target was 317 Wh/K (601 Btu/°F), and estimates of the houses were made before heat pump installation, using the cycling behavior of the native furnace. The resulting estimates ranged from 282 to 298 Wh/K (535 to 565 Btu/°F). Additional mass, consisting of 32 half-sheets of drywall, was added based on the estimated mass and on measurements made in a full-sized furnished home, from a previous study on load-based testing (Figure 2). [2]

Figure 2. Half-sheets of drywall used to increase shallow mass capacitance at the test house

Each of the houses had nearly 100 sensors, delivering data in real time to a cloud-based database. The sensors provided a comprehensive set of measurements, including within the house, and of the heat pumps. The measured temperatures, humidity, electrical power, and other parameters facilitated a thorough characterization of the capacity and efficiency of the heat pumps throughout the test period. Although atypical in field studies, for this effort the heat pump capacities were characterized both on the airside and on the refrigerant side. Air side capacity was calculated by measuring the air flow rate and the temperature and humidity of the air entering and leaving the indoor coil, as is typical. Also, Coriolis mass flow sensors were installed to measure refrigerant mass flow rate and capacity was calculated by characterizing the enthalpy entering and leaving the indoor coil with refrigerant temperature and pressure measurements (Figure 3). For each heat pump, this provided redundant measurements of capacity, a key variable, and turned out to be extremely important in obtaining accurate capacity measurements.

More details about phase I, including the home calibration process and instrumentation are published in Harley et al., 2023 and NEEP, 2023. [3] [4]

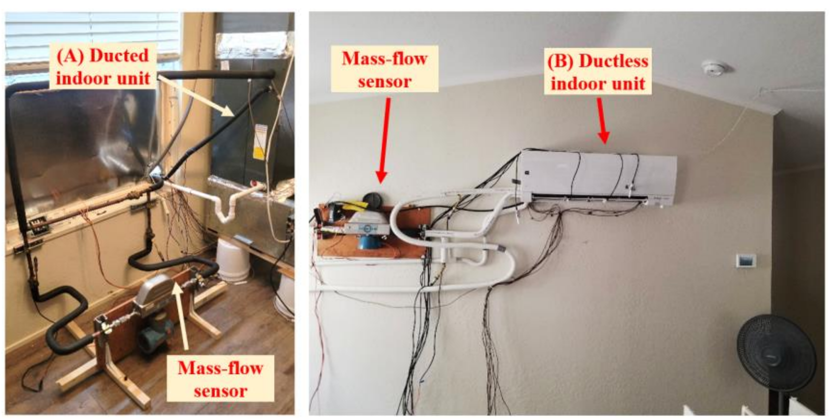


Figure 3. Ducted (A) and ductless (B) indoor units, showing refrigerant mass-flow sensors

At the conclusion of the field testing, the units were shipped to Lab A for Phase 2 testing using both SPE-07 and AHRI 210/240.

More details about the SPE-07 test procedure used in the lab testing are provided in section 2.0 of this report.

The full report describing the lab testing phase 2 and the results of the representativeness study (that also assessed repeatability) are provided in NEEP, 2024. [5]

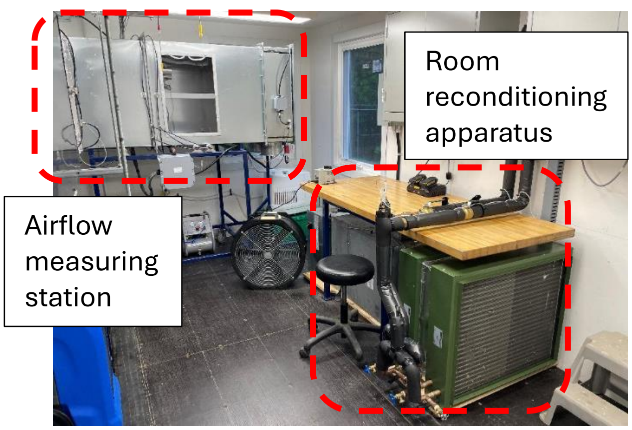
1.4. Laboratory Testing Overview

NRCan conducted SPE-07 testing at the Climate Controlled Test Facility (CCTF), which is a laboratory designed for HVAC and water-heating equipment testing located at the CanmetENERGY complex in Ottawa, Canada. It is equipped with an outdoor and indoor environment chamber with conditioning equipment capable of simulating building loads up to approximately 10.5 kW at temperatures as low as -30 °C and as high as 40 °C.

The outdoor chamber is also equipped with a secondary mechanical refrigeration system that is activated when either the primary system, or the unit under test (UUT), is undergoing a defrost operation, in order to maintain the outdoor temperature setpoint. The outdoor chamber humidity ratio is maintained by warm-air humidifiers located in the main laboratory area and distributed into the outdoor chamber by a variable speed fan. The two mechanical refrigeration systems in place remove moisture from the air.

The CCTF indoor test chamber is equipped with an ANSI/ASHRAE Standard 37[1] compliant airflow station, as shown in Figure 4, capable of measuring airflows from 60 to 2000 CFM for testing ducted and ductless systems at 0 – 1 inH_2O external static pressure. [6]

The indoor chamber reconditioning apparatus responsible for imposing the virtual building load is comprised of two chilled glycol coils with a blower and a 15-kW electric heater to precisely control the dry-bulb temperature as shown in Figure 4 (a). Circulation fans are strategically located in the chamber to ensure the air is properly mixed and uniform. A steam humidifier with 1% RH precision and a 300 CFM blower pack, as shown in Figure 4 (b), as well as a mechanical dehumidifier is used to control the humidity ratio.



(a)

(b)

Figure 4. (a) Reconditioning apparatus and airflow measuring station in the indoor test chamber, and (b) steam humidifier with blower pack installed in the indoor test chamber

[1] ANSI/ASHRAE 37-09 (R2019) was used at the time of testing conducted in this report. An updated version was released in 2024.

The CCTF is operated and monitored from a central laboratory control station as shown in Figure 5. The virtual building load emulator is programmed on the lab computer, which also monitors, data logs and checks for SPE-07 convergence during test intervals.

(a) (b)

Figure 5. (a) Main laboratory control station for the Climate Controlled Test Facility (CCTF) and (b) thermostat environment emulator (TEE) installed in the main laboratory area

A thermostat environment emulator (TEE) designed and built by Purdue University and shown in Figure 5 (b), was used to precisely control the temperature and air velocity experienced by the UUT's thermostat. The TEE was installed in the main laboratory space which is maintained at approximately 23°C year-round. The TEE's onboard conditioning equipment and controls ensure that the thermostat housed within it received the same room air temperature (RAT)[2] from the virtual building load emulator software that was maintained in the indoor test chamber and experienced by the UUT. [7]

[2] The RAT is the dry-bulb temperature target for the indoor room reconditioning system. It is a value calculated by the virtual building load emulator based on the equations provided in SPE-07.

The CCTF was designed to use the tunnel air enthalpy method as described in ASHRAE 37 to obtain the air-side measurements of UUT capacity. [6]

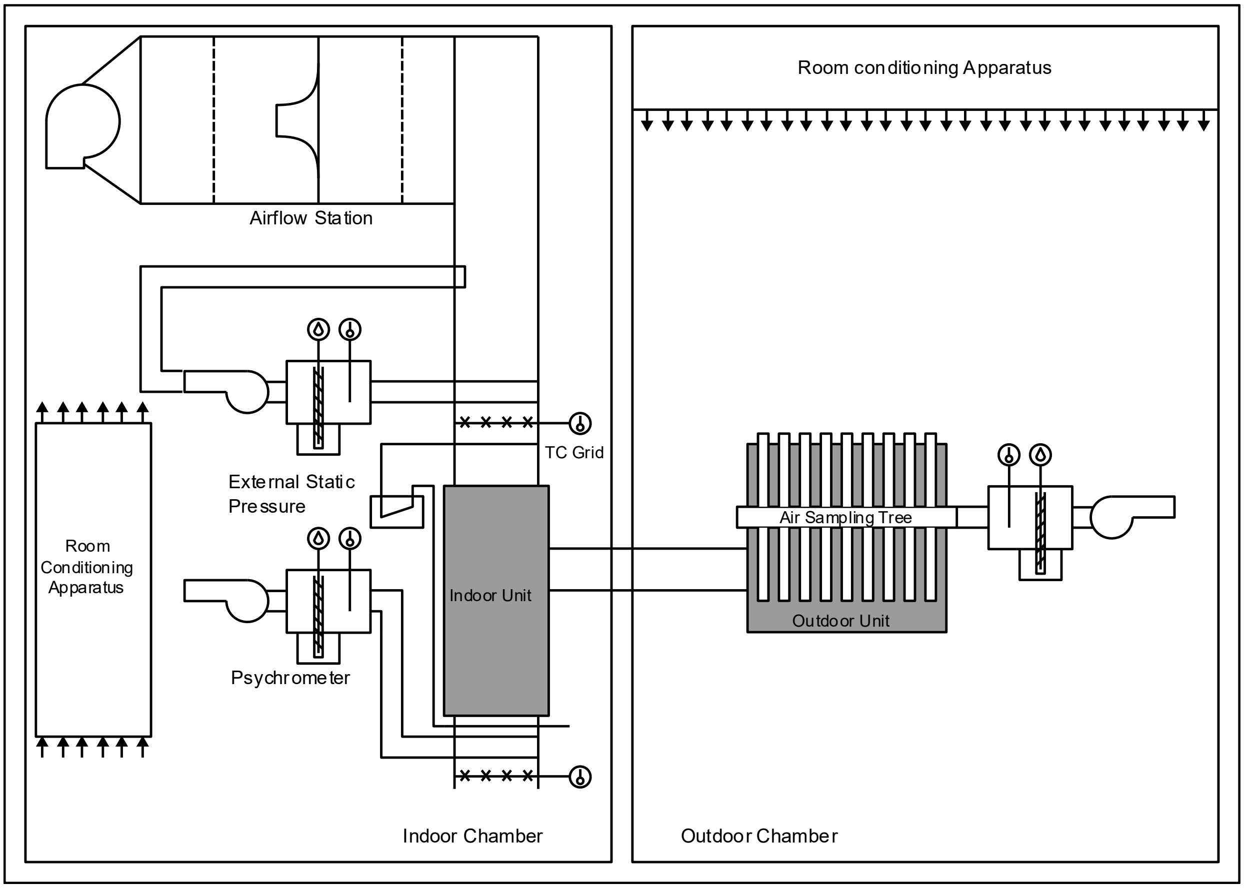


Figure 6. ASHRAE 37 tunnel air enthalpy arrangement at the CCTF

For SPE-07 testing, the CCTF used a thermopile made of type T SLE thermocouples, calibrated to ±0.2 °C, to obtain the temperature difference between the indoor unit's return and supply air streams. It also used nine individual Type T SLE thermocouples, calibrated to ±0.1 °C, to measure the return air dry-bulb temperature. The return temperature grid was used as the "indoor temperature" and was the basis for the room air temperature (RAT) control algorithm. A series of 1/8-inch Class A RTDs were also placed around the UUT to ensure the target indoor temperature was consistent.

The wet-bulb temperatures for the return and supply air streams of the indoor unit were measured by Class A accuracy ¼-inch RTD sensors placed in two aspirated psychrometers as shown in Figure 7. The outdoor chamber wet-bulb temperature was measured with the same class of RTD in a psychrometer equipped with an inline heater located in the main laboratory area. At temperatures below -15 °C, combined humidity and temperature sensors located in the air sampling trees were used. These sensors had an accuracy of 0.8% RH, and ±0.1 °C.

The supply and differential pressure measurements in the plenum and the airflow measuring station were taken by high accuracy pressure transducers with ±0.07% full-scale accuracy.

Unit A and C's indoor unit and Unit C's outdoor unit input power was measured by a power analyzer with an accuracy of ±0.04% of reading plus ±0.04% of range. Unit A's outdoor unit input power was measured by a power analyzer with an accuracy of ±0.015% of reading plus ±0.01% of range.

The outdoor unit dry-bulb temperature was taken by Class A ¼-inch RTDs located in the air sampling trees positioned around the outdoor unit. The air sampling trees were built according to the ASHRAE-37 standard specifications.

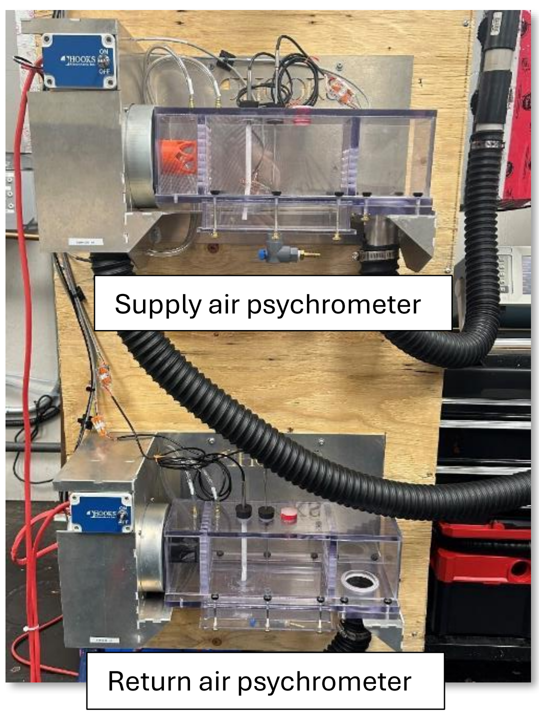


Figure 7. Supply and return psychrometers in the indoor test chamber

1.5. Overview of Unit A

Unit A was a 2-ton, variable speed, cold-climate heat pump with a vertically discharging outdoor unit, and a multi-position, centrally ducted, indoor unit.

Table 1. Unit A performance data – AHRI 210/240 2017 Addendum 1

$Q_C(95)$ A_2 Cooling Capacity [BTU/hr]	SEER	EER (A_2)	$Q_H(47)$ $H1_2$ Heating Capacity [BTU/hr]	HSPF (Region IV)
24,200	19.00	14.50	23,600	10.50

The indoor unit, shown in Figure 9, was installed in a horizontal left orientation due to height restrictions in the indoor chamber. Placing the indoor unit in a horizontal left orientation did not require making any changes except for making use of different condensate drip ports. Although manufacturers don't stipulate any differences in drainage between orientation for multi-position air handlers (given proper drain tubes be installed at the apex of the coil), it is not unreasonable to assume some differences in condensate drainage may have occurred given the shallower angles relative to the direction of gravity as shown in Figure 8. This change in orientation did not contravene SPE-07 but was not consistent with the installation orientation at Lab A. Furthermore, unlike Lab A, the indoor unit was not fitted with an inlet plenum at NRCan due to space restrictions. However, an inlet plenum is not required by SPE-07 and NRCan was compliant with all other requirements surrounding the return air sampling tree, instrumentation, and external static pressure targets.

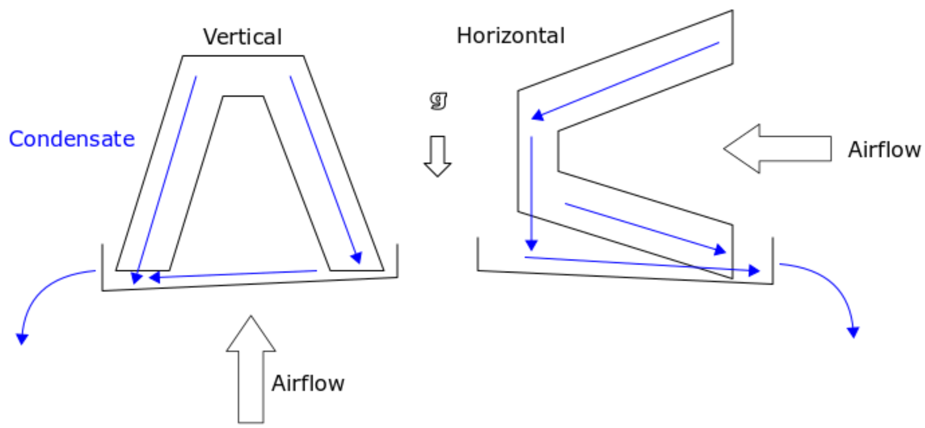


Figure 8. Indoor unit condensate drainage for vertical and horizontal discharge orientation

The indoor unit discharged into a plenum that was built according to ASHRAE-37 and connected to the airflow measuring station for air-side measurements. The plenum and inter-connecting ducting sections were insulated to R-20 with rigid foam boards and batt insulation, which was slightly higher than the R-19 required minimum stipulated in SPE-07.

The refrigerant flow meter used by Lab A for refrigerant mass flow measurements of Unit A was installed on the liquid line to maintain a similar pressure drop but was not used.

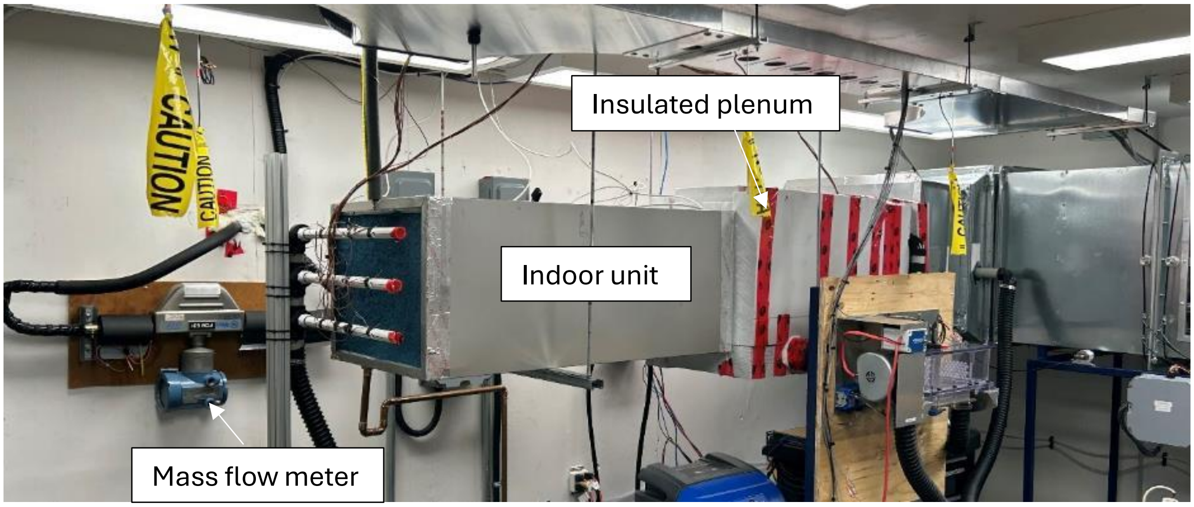


Figure 9. Unit A indoor unit installed in the indoor test chamber

The outdoor unit, shown in Figure 10, was not elevated for proper condensate drainage due to height restrictions in the outdoor chamber. The outdoor unit's fan was 1.5 meters away from the ceiling of the outdoor chamber. This contravened the requirement in ASHRAE-37 that stipulates any air discharging surface must be a minimum of 1.8 meters away from any other surface.

The outdoor unit was surrounded by four air sampling trees: one for each heat exchanging surface of the outdoor unit. The four sampling trees were connected to a common junction box where a single measurement of temperature and humidity was taken. Extra sensors including thermocouples and RTDs were installed on the air sampling trees to cross-check the measurement taken by the junction box temperature sensor. The junction box was connected to a psychrometer located in the main laboratory space and was calibrated to move air at a velocity of 900 ft/min, as per manufacturer recommendation.

The refrigerant line set was installed and insulated to match the installation at Lab A. However, the liquid line length used at Lab A of 8 feet could not be replicated in the chambers as the distance between the indoor and outdoor unit was too great. Instead, the liquid line length was increased to 13 feet. The extra line set was insulated using multiple layers of R-20 batt to minimize heat loss to the environment. Likewise, the line set between indoor and outdoor chambers was insulated with multiple layers of R-20 batt to minimize heat loss since the CCTF indoor and outdoor chambers do not share a wall.

The same refrigerant charge of 12.07 lbs was used even with the longer liquid line length as this charge level was adequate for a refrigerant line set length of up to 80 feet according to the manufacturer. The vapour line was 16 feet and 6 inches at both test facilities.

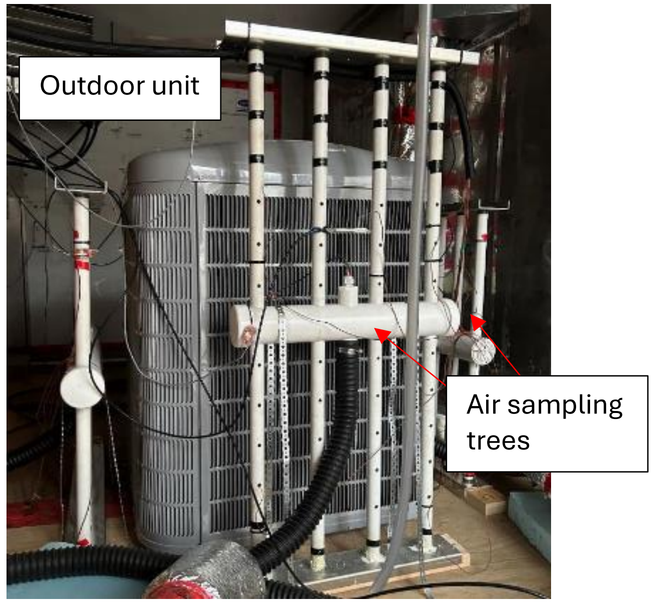


Figure 10. Unit A outdoor unit installed in the outdoor test chamber with air sampling trees

Unit A's thermostat, shown in Figure 11, was a communicating thermostat with onboard temperature and relative humidity sensors. It was mounted in the TEE in the main laboratory area and connected to the indoor unit in the indoor chamber via communication wire. The thermostat was received from Lab A with a fan airflow speed set to High. After consulting with Lab A, the thermostat was set to "Auto" fan airflow mode for the cooling tests, and "High" fan airflow mode for the heating tests, to match the settings used for the same tests conducted at Lab A.

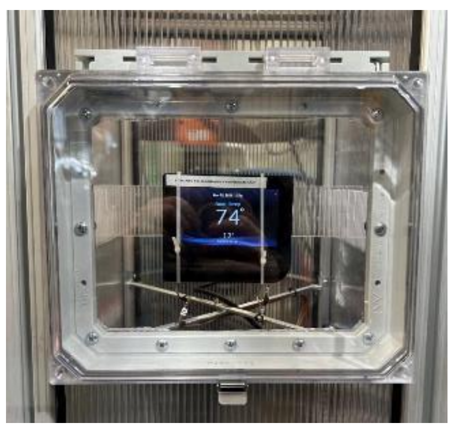


Figure 11. Unit A communicating thermostat placed in the thermostat environment emulator

1.6. Overview of Unit C

Unit C was a 2-ton variable speed cold-climate heat pump with a horizontally discharging outdoor unit, and a multi-position, centrally ducted, indoor unit.

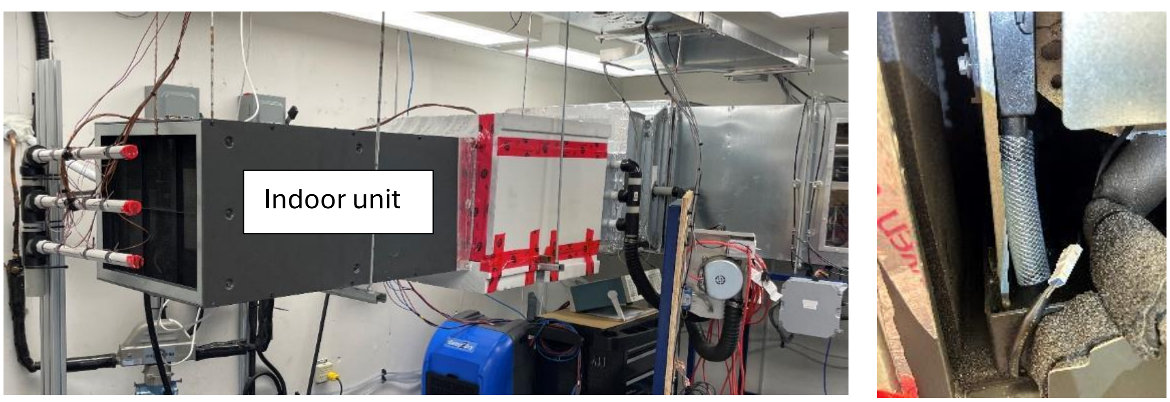
Table 2. Unit C performance data – AHRI 210/240 2024

Q_C(95) A_{full} Cooling Capacity [BTU/hr]	SEER2	EER2 (A_{full})	Q_H(47) $H1_{full}$ Heating Capacity [BTU/hr]	HSPF2 (Region IV)
18,000	18.70	12.40	21,600	9.20

The indoor unit, shown in Figure 12, was also installed in a horizontal left position due to height restriction, which required flipping the A-coil and blower 180 degrees. Two condensate discharge tubes, shown in Figure 12, also had to be installed at the end of the A-coil to ensure the condensate would drain correctly from the coil. The indoor unit was also modified to be powered independently of the outdoor unit with a manufacturer-provided module. This was done to enable independent power measurements.

As was done for Unit A, the indoor unit discharged into a plenum that was built according to ASHRAE-37 and connected to the airflow measuring station for air-side measurements. The plenum and inter-connecting ducting sections were insulated to R-20 with rigid foam boards and batt insulation, which was slightly higher than the R-19 required minimum stipulated in SPE-07.

The refrigerant flow meter used by Lab A (and in the field during Phase 1 testing) for refrigerant mass flow measurements of Unit C was installed on the liquid line to maintain a similar pressure drop.



(a) (b)

Figure 12. (a) Unit C indoor unit in the indoor test chamber, and (b) condensate tube attached to the A-coil following the change of discharge orientation

The outdoor unit, shown in Figure 13, was raised off the ground using a stock AC stand. The unit was positioned in the chamber to ensure it met the minimum 1.8 m distance requirement for air discharging surfaces away from other surfaces. However, one air intake surface had to be positioned close to the wall to maintain the same refrigerant line set length used in Lab A. This contravened ASHRAE-37's minimum requirement of 0.9 m, by 0.3 m.

The outdoor unit was surrounded by two air sampling trees: one for each heat exchanging surface of the outdoor unit. The two sampling trees were connected to a common junction box where a single measurement of temperature and humidity was taken. Extra sensors including thermocouples and RTDs were installed on the air sampling trees to cross-check the measurement taken by the junction box temperature sensor. The junction box was connected to a psychrometer located in the main laboratory space and was calibrated to move air at a velocity of 900 ft/min, as per manufacturer recommendation.

The total refrigerant line set lengths on both the liquid and vapour lines were the same as Lab A. However, due to the size and placement of the two chambers, NRCan could not match the length of refrigerant line set installed outdoors versus indoors and had to increase the amount of line set placed in the outdoor chamber relative to the indoor chamber. Therefore, the line set that would otherwise have been in the indoor chamber was insulated to R-19 using batt insulation. The refrigerant charge was the same as Lab A at 3.00 lbs of R410a.

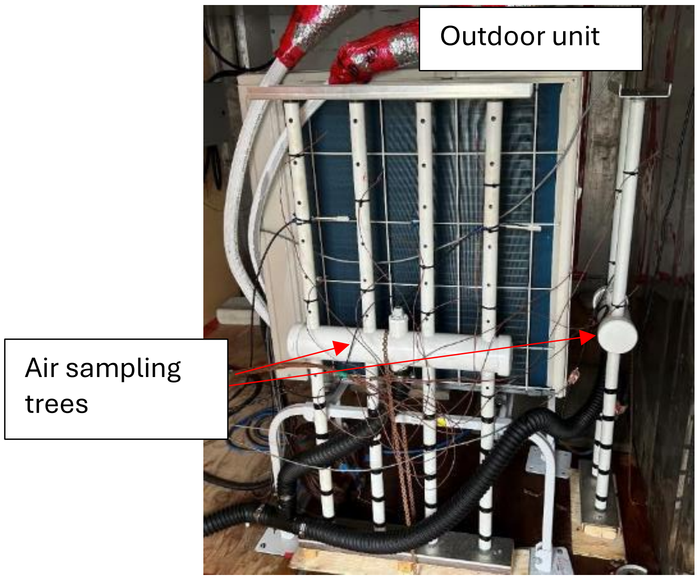


Figure 13. Unit C outdoor unit in the outdoor test chamber with air sampling trees

Unit C's thermostat, shown in Figure 14, was a communicating thermostat with an onboard temperature sensor. It was mounted in the TEE in the main laboratory area and connected to the indoor unit in the indoor chamber via communication wire. The thermostat was received from Lab A and no changes were made.

Figure 14. Unit C communicating thermostat in the thermostat environment emulator

2.0. Methodology

SPE-07 is a load-based test method that uses a sequence of simulated heating and cooling loads that are imposed in the indoor psychrometric chamber. The simulated "virtual" loads are controlled by the software that operates the test chamber, a setup that is also referred to as a load "emulator". The loads are scaled with each outdoor temperature condition to represent an appropriate heating or cooling requirement at that temperature, for a typical residential building in which a heat pump of that size would be installed. The tested system is set up to heat or cool the indoor chamber using its own controls in response to the simulated loads, while its heating or cooling capacity and power input are measured in real time.

At each test condition shown in Table 3 and Table 4, the outdoor room temperature is held constant, while the indoor room reconditioning equipment is programmed to respond to the current performance of the tested system. The virtual building model compares the heating or cooling capacity of the UUT with the target heat loss or cooling gain and adjusts the indoor chamber room conditions to reflect the difference. For example, in a heating test, if the capacity of the equipment is smaller than the load, the room temperature will slowly drop, as it would in a real building. This allows the tested unit's thermostat to sense that condition and increase its capacity in response. The virtual building load includes a term for realistic mass capacitance, which slows the room temperature response, helps stabilize operation during the test, and eliminates test chamber capacitance from the response.

Table 3. SPE-07 test room conditions for $SCOP_C$ rating test series

	Humid test conditions			Dry test conditions		
	Outdoor dry-bulb temperature [°C]	Indoor dry-bulb temperature [°C]	Indoor humidity ratio [-]	Outdoor dry-bulb temperature [°C]	Indoor dry-bulb temperature [°C]	Indoor humidity ratio [-]
CA	N/A	23.33	0.010	45	26.11	0.087
CB	40			40		
CC	35			35		
CD	30			30		
CE	25			25		

Table 4. SPE-07 test room conditions for $SCOP_H$ rating test series

	Continental outdoor conditions		Marine outdoor conditions		Indoor conditions	
	Dry-bulb temperature [°C]	Humidity ratio [-]	Dry-bulb temperature [°C]	Humidity ratio [-]	Dry-bulb temperature [°C]	Humidity ratio [-]
HL	LCT		N/A	N/A	21.1	0.0092 (Max)
HB	-15	0.0008				
HC	-8.3	0.0013	N/A	N/A		
HD	1.1	0.0031	1.1	0.0035		
HE	8.3	0.0042	N/A	N/A		
HF	12.2	0.0045	N/A	N/A		

The tested equipment is thus subjected to conditions that closely represent real-world operation, while it responds using its native controls. By contrast, in conventional rating tests, the indoor and outdoor rooms are held at fixed conditions, and the compressor and fan flows are locked in steady-state operation using proprietary test modes based on manufacturer discretion, within parameters set by the procedure for each test condition. During fixed-speed testing, the logic of the unit's native controls is bypassed entirely, and in some cases the compressor and/or fan(s) may operate at speeds that are not available during operation under the system's normal control sequences. Also, the SPE-07 tests incorporate defrost operations for heating, and compressor cycling under low load conditions for heating and cooling across all tests, rather than accounting for those effects using mathematical adjustments. Experience has shown that some units' behavior (and rated efficiency) can vary dramatically with load-based testing versus fixed-speed testing. [8] [9] [10]

In the normal calculation process of SPE-07, the results of the tests are mapped into a set of climate models based on bins (tranches of outdoor temperature ranges) to create a set of seasonal ratings for eight heating and seven cooling climates. The $SCOP_H$ and $SCOP_C$ presented in this report were calculated using the reporting template spreadsheet available with the CSA SPE-07:23 procedure.

2.1. Convergence Criteria Evaluation

During each test condition, a series of criteria are evaluated in order to determine when the results have been adequately stable to allow computation of the efficiency for that test condition and proceed to the subsequent test condition. The criteria are assessments of convergence, showing that the COP is consistent over time, and they vary across a number of parameters that are outlined in sections 2.1.1 and 2.1.2. In each case, if convergence is reached or the procedure otherwise directs, the net capacity, power input, and indoor and outdoor conditions are recorded for the specified time period and the COP is calculated as the ratio of capacity to input. If the indoor dry-bulb temperature condition goes out of a specified range for a specified time period, the unit is considered to be loaded beyond its variable-speed capacity, and the test is re-started as a “full load” test with indoor room conditions fixed and the unit running at maximum speed under its own controls.

2.1.1. Cooling Rating Test Convergence

The following criteria govern convergence of cooling test conditions. Evaluation of the criteria depend first on the outdoor and indoor dry-bulb temperatures and the indoor humidity ratio falling within the allowed tolerances for a minimum of 15 minutes at the given test condition.

2.1.1.1. Criterion 1 (part-load, variable speed operation)

If the tested system is running in “regular variable speed” mode (the minimum outdoor unit power is not less than 80% of the maximum, and the unit does not cycle off), the indoor dry-bulb temperature and measured COP are evaluated on a rolling basis across three successive 20-minute periods. Convergence is met when the COP for periods 2 and 3 are within 1% of the COP for periods 1 and 2 respectively, and the average indoor dry-bulb temperature (which is expected to vary because of the nature of the virtual building load) across the three periods is not more than 1.11 °C higher than the target value shown in Table 3.

2.1.1.2. Criterion 2 (part-load, on/off cycling)

If the tested system is running in “regular cycling” mode (the compressor cycles on and off for total cycle times that do not vary by more than ±20%), the indoor dry-bulb temperature and measured COP are evaluated on a rolling basis across two successive on/off cycles (or multiple cycles as needed for two successive periods of at least 20 minutes). Convergence is met when the COP for the two cycles (or periods of multiple cycles) are within 1%, and the average indoor dry-bulb temperature across the two periods is not more than 1.11 °C higher than the target value shown in Table 3.

2.1.1.3. Criterion 3 (full-load operation)

If the average indoor dry-bulb temperature remains more than 1.11 °C higher than the target value shown in Table 3 (evaluated on a rolling 15-minute average) for 90 minutes, the test condition is considered to be above the unit’s capacity and the test condition is restarted as a full-load test. All subsequent test conditions at higher loads are conducted only as full-load tests. In the case of a full-load test, convergence is reached when two successive (rolling) 20-minute periods have measured COPs within 1%.

2.1.1.4. 4h Test (CE and irregular system behaviour)

If the other cooling test criteria are not met, the net capacity, power input, and indoor and outdoor conditions are recorded across a 4-hour period, regardless of modulating or cycling behavior. The COP is calculated as the ratio of capacity to input across the entire period. Test condition CE (the lowest load) is always evaluated using this criterion, regardless of system behavior.

2.1.2. Heating Rating Test Convergence

The following criteria govern convergence of heating test conditions. Evaluation of the criteria depend first on the outdoor and indoor dry-bulb temperatures and the outdoor humidity ratio falling within the allowed tolerances for a minimum of 15 minutes at the given test condition.

2.1.2.1. Criterion 1 (part-load, variable speed operation)

If the tested system is running in “regular variable speed” mode, the indoor dry-bulb temperature and measured COP are evaluated on a rolling basis across three successive 20-minute periods where no defrost occurs. Convergence is met when the COP for periods 2 and 3 are within 1% of the COP for periods 1 and 2 respectively, and the average indoor dry-bulb temperature (which is expected to vary because of the nature of the virtual building load) across the three periods remains at a value that is not more than 1.11 °C below the target value shown in Table 4.

2.1.2.2. Criterion 2 (part-load, on/off cycling)

If the tested system is running in “regular cycling” mode, the indoor dry-bulb temperature and measured COP are evaluated on a rolling basis across two successive on/off cycles (or multiple cycles as needed for two successive periods of at least 20 minutes) where no defrost occurs. Convergence is met when the COP for the two cycles (or periods of multiple cycles) are within 1%, and the average indoor dry-bulb temperature across the two periods remains at a value that is not more than 1.11 °C below the target value shown in Table 4.

2.1.2.3. Criterion 3 (part-load, including defrost operation)

If a defrost occurs, the COP is evaluated across three complete frost/defrost cycles (that are a minimum of 20 minutes each). Convergence is met when the COP for cycles 2 and 3 are within 1% of the COP for cycles 1 and 2 respectively, and the average indoor dry-bulb temperature (which is expected to vary because of the nature of the virtual building load) across the three cycles remains at a value that is not more than 1.11 °C below the target value shown in Table 4. If only one defrost occurs and another does not occur within 3 hours, or during test interval HE, criteria 1 or 2 can be used with the single defrost period excluded from the evaluated time period.

2.1.2.4. Criterion 4 (full-load operation)

If the average indoor dry-bulb temperature remains greater than 1.11 °C below the target value shown in Table 4 (evaluated on a rolling 15-minute average) for 90 minutes, or for at least two complete frost/defrost cycles, the test condition is considered to be above the unit’s capacity and the test condition is restarted as a full-load test. All subsequent test conditions at higher loads are conducted only as full-load tests. In the case of a full-load test, convergence is reached when two successive (rolling) 20-minute periods have measured COPs within 1%, or if defrost occurs, when three successive defrosts have measured COPs within 1%.

2.1.2.5. 4h Test (HF and irregular system behaviour, no defrost)

If the other heating test criteria are not met and defrost does not occur, the net capacity, power input, and indoor and outdoor conditions are recorded across a 4-hour period, regardless of modulating or cycling behavior. The COP is calculated as the ratio of capacity to input across the entire period. Test condition HF (the lowest load) is always evaluated using this criterion, regardless of system behavior. If during test HF a defrost cycle occurs, the defrost cycle is excluded and the time is extended to evaluate the full 4 hours of operation.

2.1.2.6. 6h Test (HD and irregular system behaviour, with defrost)

If the other heating test criteria are not met and defrost does occur, the net capacity, power input, and indoor and outdoor conditions are recorded across a 6-hour period, regardless of modulating or cycling behavior, except that when multiple defrosts occur, the evaluation period includes only full frost/defrost cycles. The COP is calculated as the ratio of capacity to input across the entire evaluation period. Test condition HD (the highest likelihood of significant defrosting) is always evaluated using this criterion, regardless of system behavior.

2.1.2.7. 7h Test (irregular system behaviour, with two defrost cycles and a third that will likely occur within the next hour)

If during the 6-hour test, two defrost cycles have been completed and it is apparent that a third will likely occur before the 7-hour time limit, the 6-hour test is extended by one hour. If the third defrost cycle is completed by then, it is included; if it is not completed the evaluation period is limited to the first two cycles.

2.2. Changes to the SPE-07:23 Load Line

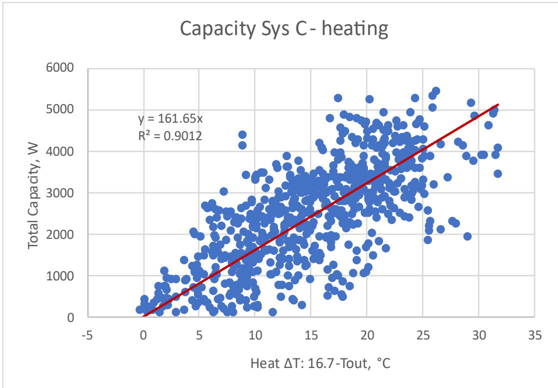
One feature of the SPE-07 test procedure that was modified for the purpose of this study is the load line, which is the relationship between outdoor temperature and the virtual building load. SPE-07 includes several default parameters that are intended to model a building in which the tested system would be installed, those being:

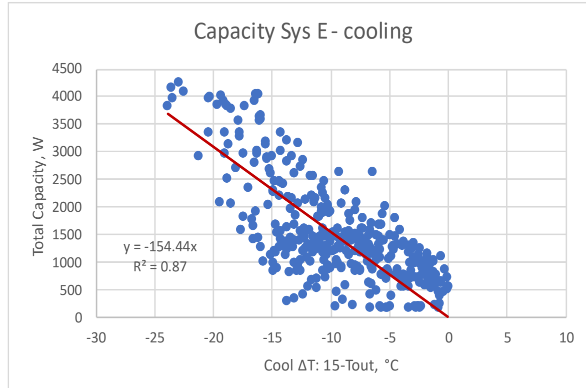
a) The heating and cooling load lines,
b) The load lines' relationship to the equipment size and to each other,
c) The assumed thermal mass time constant to be used during testing, and,
d) The sensible heat ratio of the equipment

These parameters need to be standardized for a test and rating procedure, but they will never precisely match any real house.

Prior to the field tests, the test houses were calibrated to be as close to the SPE-07 parameters for heating load line and mass capacitance as possible. The technical advisory committee decided that the best way to validate the load-based test method against real homes would be to match the loads that were measured in the homes during the field testing, even if those differed from the default assumptions in SPE-07. That is to say, the priority was to evaluate the test method by providing the fairest comparison possible between the field and lab testing, rather than to test the relevance of the SPE-07 default parameters for the set of installations. This was deemed to be particularly important for the cooling load lines, which (because of the dependence of cooling loads on solar radiation) could not be estimated during the calibration process, and which would not be expected to follow the assumptions about the relationship between heating and cooling design loads used in SPE-07.

Before the lab testing was conducted, the field data was analyzed to estimate the heating and cooling load lines for each house. Figure 15 shows examples of hourly data for one of the systems in the heating season (a) and one in the cooling season (b). The data does not neatly fit a linear pattern due to the many transients that affect a house from one hour to the next. Because a load line includes both a slope and a zero-load point (essentially an x-intercept) these data were also analyzed to estimate both; hence, the x-axes in the figure represent a temperature *difference* rather than an absolute temperature, where the zero point is 16.7 °C in chart (a) and 15 °C in chart (b), as shown on the axis labels. Because the systems' temperature controls were set in the field according to the controls' display and not calibrated to include any offset, the zero-load point varied from system to system and between heating and cooling.





(a) (b)

Figure 15. Typical load line regressions for heating (a) and cooling (b), from field data

The average heating design load at -15 °C for the tested systems in the test homes was 78% of the SPE-07 design load; the cooling load at 35 °C averaged 99% of SPE-07. The average zero-load temperature for heating was 15.7 °C (SPE-07 uses 15.6 °C); the average for cooling was 17.0 °C (SPE-07 uses 19.4 °C). These values for design load and zero-load point for each specific system were substituted for the SPE-07 default values for each heating and cooling test set conducted by Lab A and NRCan.

The mass capacitance during the field study could not be estimated from the heat pump operation data, so the default values from SPE-07 were used for the lab testing. The original measurements of the houses during the calibration period were only slightly below the SPE-07 values, and additional mass was added to the houses prior to the heat pump operation, so the capacitance values were not expected to impact the comparison significantly. Finally, the cooling sensible heat ratio (SHR) for the load-based lab testing for each unit was set at the average value from the field study cooling period for that unit. The average of field data for all homes was an SHR of 0.85, and the default for a humid climate in SPE-07 is 0.8. Changes to the SPE-07 default values for Unit A and C are presented in Section 7.2.

3.0. Results

The scope of reproducibility testing conducted in this present study was limited to the rating tests representative of an ASHRAE climate zone 6, thereby excluding the Very Hot CD_A (45°C) cooling test interval. Therefore, Lab A and NRCan attempted to conduct a complete series of learning tests and rating tests on both Unit A and Unit C and excluded tests outlined in Section B.2.4 of SPE-07 onwards, such as the standby power consumption test, which were not conducted at Lab A.

3.1. Unit A

3.1.1. Heating Rating Test Summary

A complete series of SPE-07 heating learning and rating tests was conducted at Lab A and NRCan's facilities on Unit A to evaluate reproducibility. The results of capacity, input power, COP, and seasonal COP (SCOP) are presented below with Lab A's test results serving as the reference against which NRCan's results are compared. Several tests are analyzed in greater detail to support discussion of differences in equipment behaviour and to investigate the limits of reproducibility. The green-shaded areas in figures represent the convergence period in instances where data is presented beyond the convergence period. Figures presenting results of indoor and outdoor dry-bulb temperature have a blue dashed line for the target indoor dry-bulb temperature and a red dashed line for the target outdoor dry-bulb temperature. For more detailed test results, see Section 7.1.1.

The heating learning series tests determined that tests HF through HC were part-load tests, whereas tests HB and HL (LCT) were full-load tests. This was consistent between Lab A and NRCan as were the thermostat setpoints as shown in Table 5. The learning series tests also calibrate the required thermostat setpoint for each test condition, to determine the setting that controls the indoor room temperature most closely to the target temperature. The HC thermostat setpoints differed between laboratories, which likely explains why the indoor room temperature in Lab A for test HC was noticeably lower than for NRCan; the consistent setpoints for all test conditions in Lab A suggest that they may not have spent as much effort matching the indoor target dry-bulb temperature for condition HC.

Table 5. Unit A thermostat setpoints in heating mode for NRCan and Lab A as determined by the heating learning series tests

Test	Outdoor dry-bulb temperature, °F (°C)	Indoor dry-bulb control target, °F (°C)	TSTAT, °F (°C) (NRCan)	TSTAT, °F (°C) (Lab A)
HF_C	54 (12.2)	69 (20.6)	69 (20.6)	69 (20.6)
HE_C	47 (8.3)	69 (20.6)	68 (20)	68 (20)
HD_C	34 (1.1)	69 (20.6)	68 (20)	68 (20)
HC_C	17 (-8.3)	69 (20.6)	70 (21.1)	68 (20)
HB_C	5 (-15)	Max	Max (88 (31.1))	Max (88 (31.1))
HL_C	-3 (-19.4)	Max	Max (88 (31.1))	Max (88 (31.1))
HD_M	34 (1.1)	69 (20.6)	68 (20.6)	68 (20)

Following the heating learning series tests, the heating rating tests were completed in order of lowest to highest load. Heating capacity results presented in Figure 16 reveal that tests HF and HE had higher integrated capacities at NRCan's test facility, test HB had lower integrated capacity, and other tests were more consistent. UUT integrated power consumption presented in Figure 17, on the other hand, was consistently higher at NRCan's test facility for all rating tests, with test HC's deviation from testing at Lab A being an outlier, with the lower thermostat setpoint and lower indoor dry-bulb temperature a likely contributor. The resulting rating test COPs shown in Figure 18 reveal that tests HD continental and marine, HC, and HB results diverged most substantially between test facilities. Rating tests HF, HD continental, HC and HB will be discussed in greater detail below.

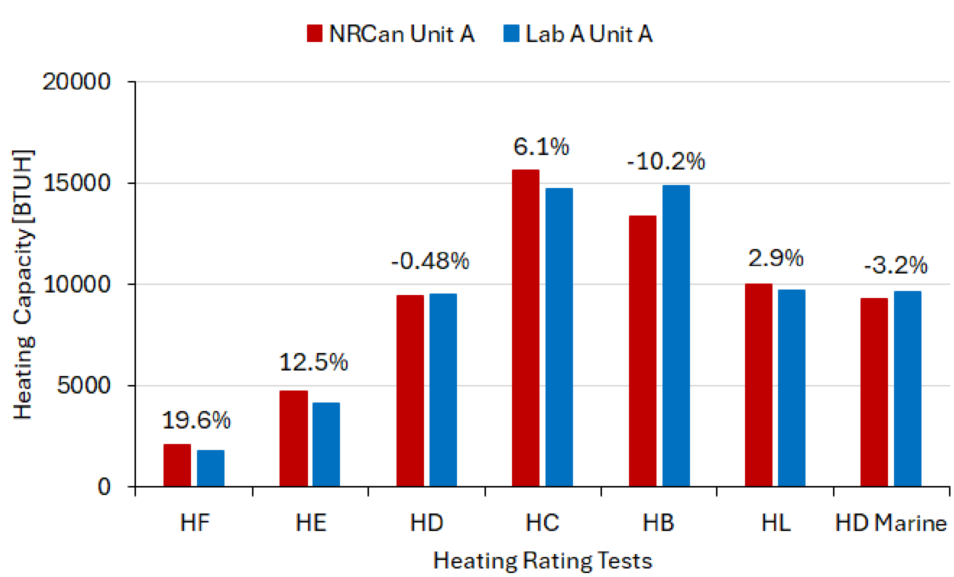


Figure 16. Heating capacity for Unit A, NRCan and Lab A results

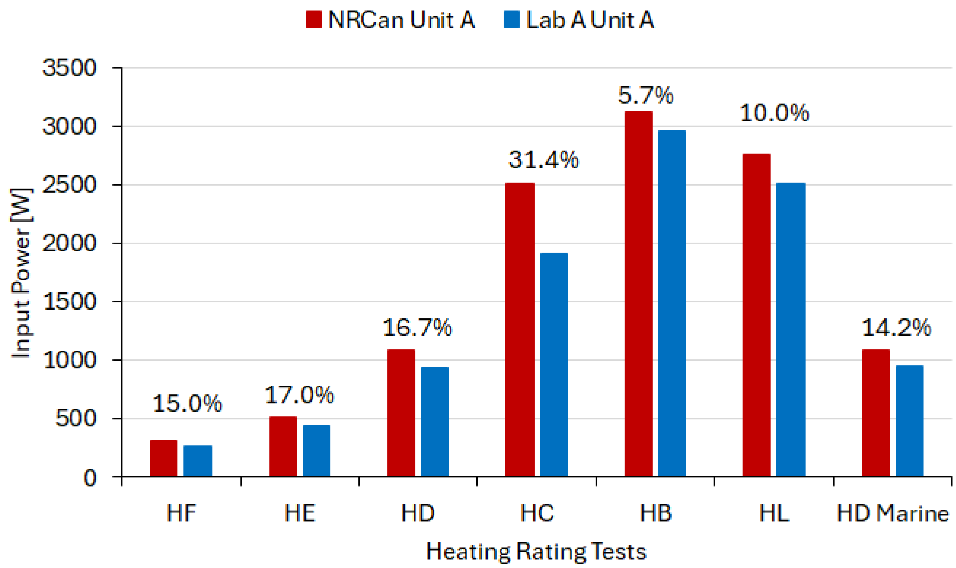


Figure 17. Heating input power for Unit A, NRCan and Lab A results

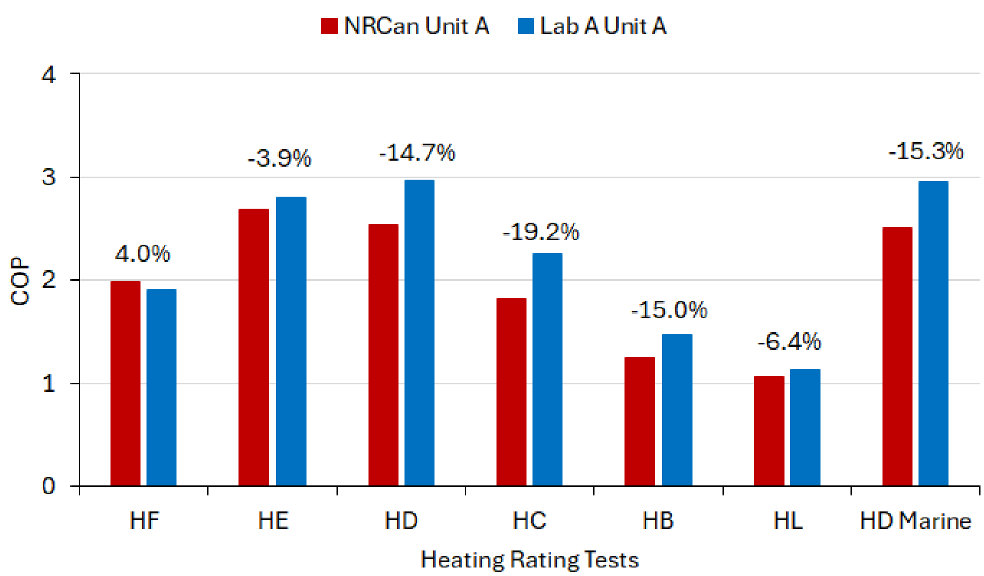


Figure 18. Heating COPs for Unit A, NRCan and Lab A results

Finally, the heating seasonal COPs ($SCOP_H$) for the heating bins are presented in Figure 19 and Table 6. The mean difference between Lab A and NRCan $SCOP_H$ across all climate zones is 8.8%, with a standard deviation of 3.6%. Overall, this shows fair reproducibility. The subarctic and very cold regions are considered “good” in reproducibility terms, while the rest of the regions are considered “fair” to use the terminology described in Purdue University’s AHRI report on repeatability and reproducibility of CSA EXP-07:19 (2022). [11]

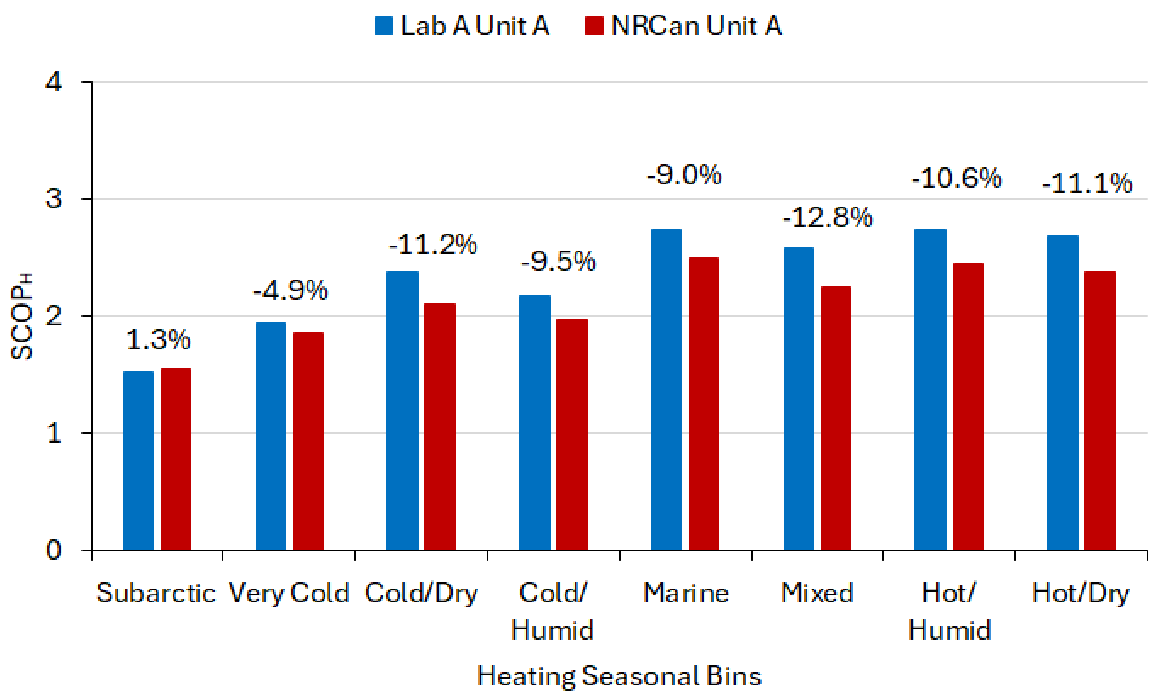


Figure 19. $SCOP_H$ for Unit A, NRCan and Lab A results

Table 6. $SCOP_H$ for Unit A, NRCan and Lab A results

	Subarctic	Very Cold	Cold/Dry	Cold/ Humid	Marine	Mixed	Hot/ Humid	Hot/Dry
Lab A	1.53	1.95	2.38	2.18	2.74	2.59	2.74	2.68
NRCan	1.55	1.85	2.11	1.97	2.49	2.25	2.45	2.38
$\Delta SCOP_H$ [%]	1.25	4.90	11.23	9.45	9.02	12.85	10.58	11.09

3.1.2. Individual Heating Test Results

3.1.2.1. Continental Heating Test Interval HF (12.2°C)

Figure 20 shows the behaviour of rating test interval HF (12.2°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was part-load and ran for 4 hours as required by SPE-07. The results demonstrate the reproducible nature of a load-based interval. Two short on cycles were captured at each test facility with relatively long off cycles. Unit A ramped up to approximately 4 kW of capacity and 1 kW of input power consumption. The heating capacity seemed to ramp up more slowly at Lab A than at NRCan potentially because of differences in instrumentation used for capacity measurements. Airflows were also relatively similar at around 700-750 CFM during off cycles, and 400-450 CFM during on cycles. The resulting difference in COP for this interval was 4.0%.

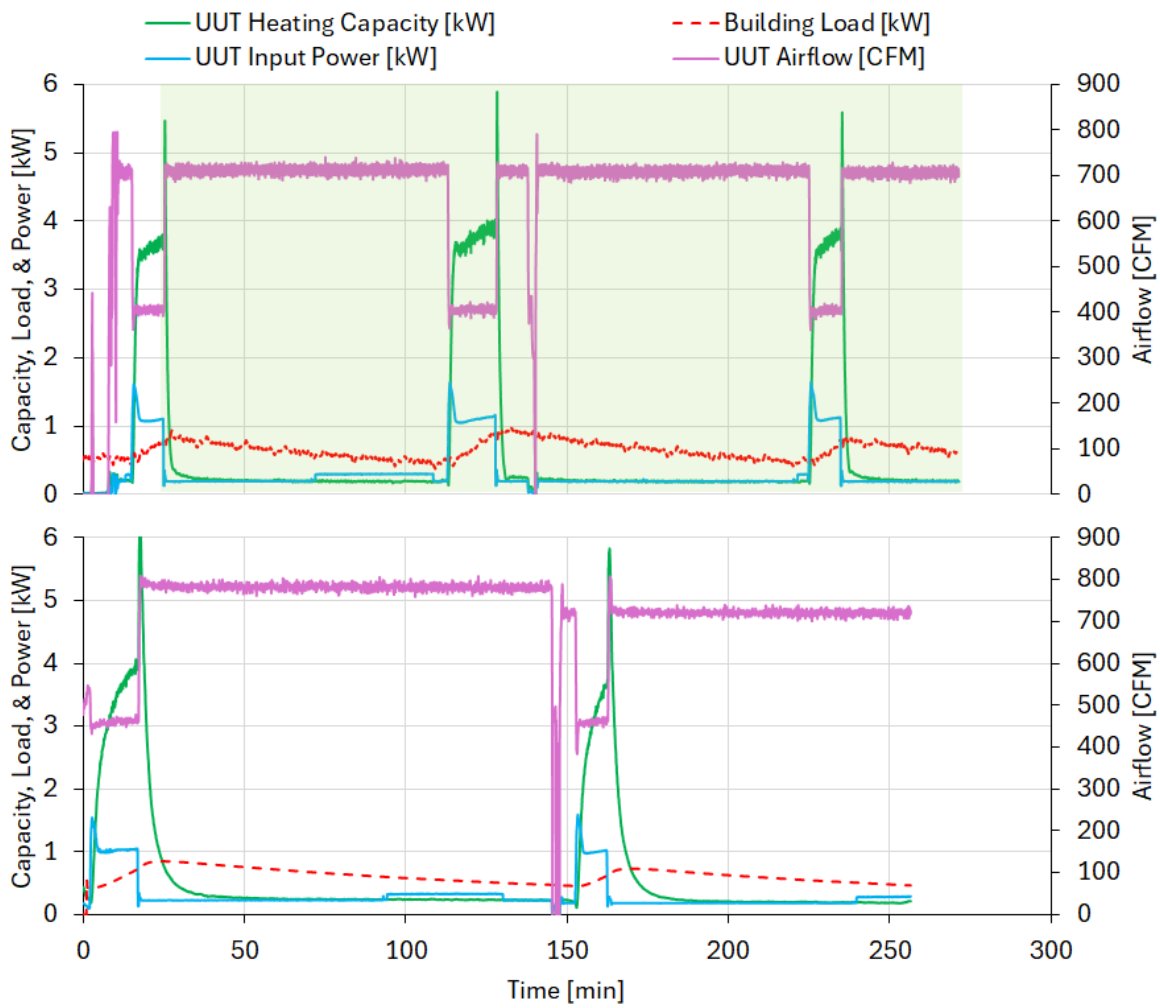


Figure 20. Unit A capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HF (12.2°C)

Figure 21 shows the indoor and outdoor dry-bulb temperature behaviour of test interval HF at NRCan and Lab A. The indoor dry-bulb temperature maintained at Lab A was slightly higher than at NRCan which could be due to different thermostat and equipment dynamic response for this interval. It should be noted that the test interval results were recorded at a 5-second interval at Lab A, while NRCan recorded data at a 1-second interval. This, along with different temperature sensing instrumentation, can partially explain the rounded peaks of the indoor dry-bulb temperature observed in Lab A's data, and the generally "smoother" capacity profiles.

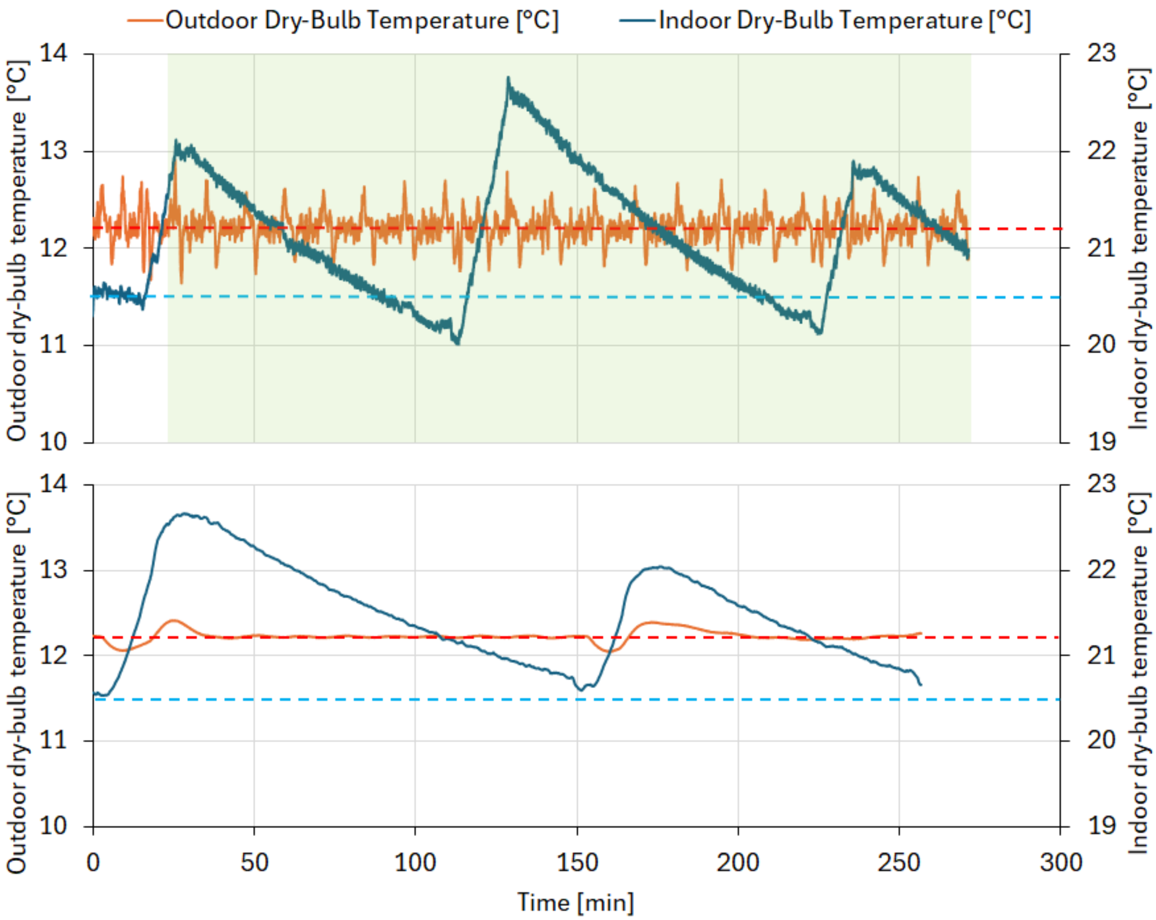


Figure 21. Unit A indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for heating continental test interval HF (12.2°C)[3]

[3] **Reminder!** The green-shaded areas in figures represent the convergence period in instances where data is presented beyond the convergence period. Figures presenting results of indoor and outdoor dry-bulb temperature have a blue dashed line for the target indoor dry-bulb temperature and a red dashed line for the target outdoor dry-bulb temperature.

3.1.2.2. Continental Heating Test Interval HD (1.1°C)

Figure 22 shows the behaviour of rating test interval HD – Continental (1.1°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was part-load and runs for 6 hours as required by SPE-07. Six cycles and one defrost operation were observed at both test facilities. The average ON cycle heating capacity was 6-7 kW for both tests, with airflows being 50-75 CFM higher on average at Lab A. The input power consumption diverged slightly between tests as well, with NRCan observing approximately 160 W more power. It should be noted that while both tests experienced one defrost operation, the cooling capacities during these defrost operations were not integrated the same way. This issue is explained in more detail in the HC test interval discussion below. The higher input power in NRCan's test interval resulted in a 14.7% lower COP.

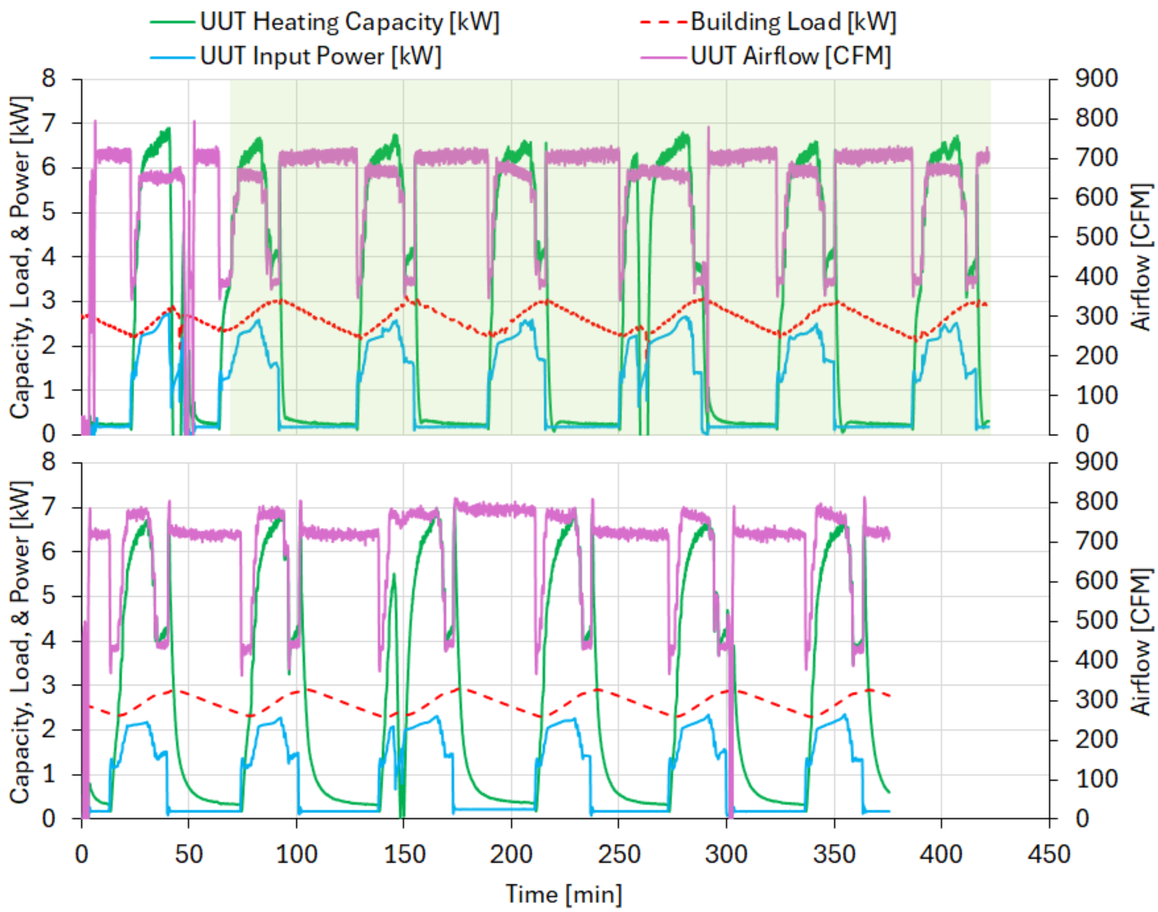


Figure 22. Unit A capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HD – Continental (1.1°C)

Figure 23 shows the indoor and outdoor dry-bulb temperature behaviour of test interval HD – continental at NRCan and Lab A. Unit A maintained the desired thermostat setpoint of 20.5°C at both facilities. However, as it had done in test interval HF, Unit A cycled ON at a lower indoor dry-bulb temperature at NRCan with an average start-of-cycle temperature of 18°C, versus 19°C at Lab A. The shut-off temperature was the same at both facilities at around 22°C. It is surmised that Unit A's thermostat exhibited a larger thermostat "deadband" during testing at NRCan as compared to tests conducted at Lab A.

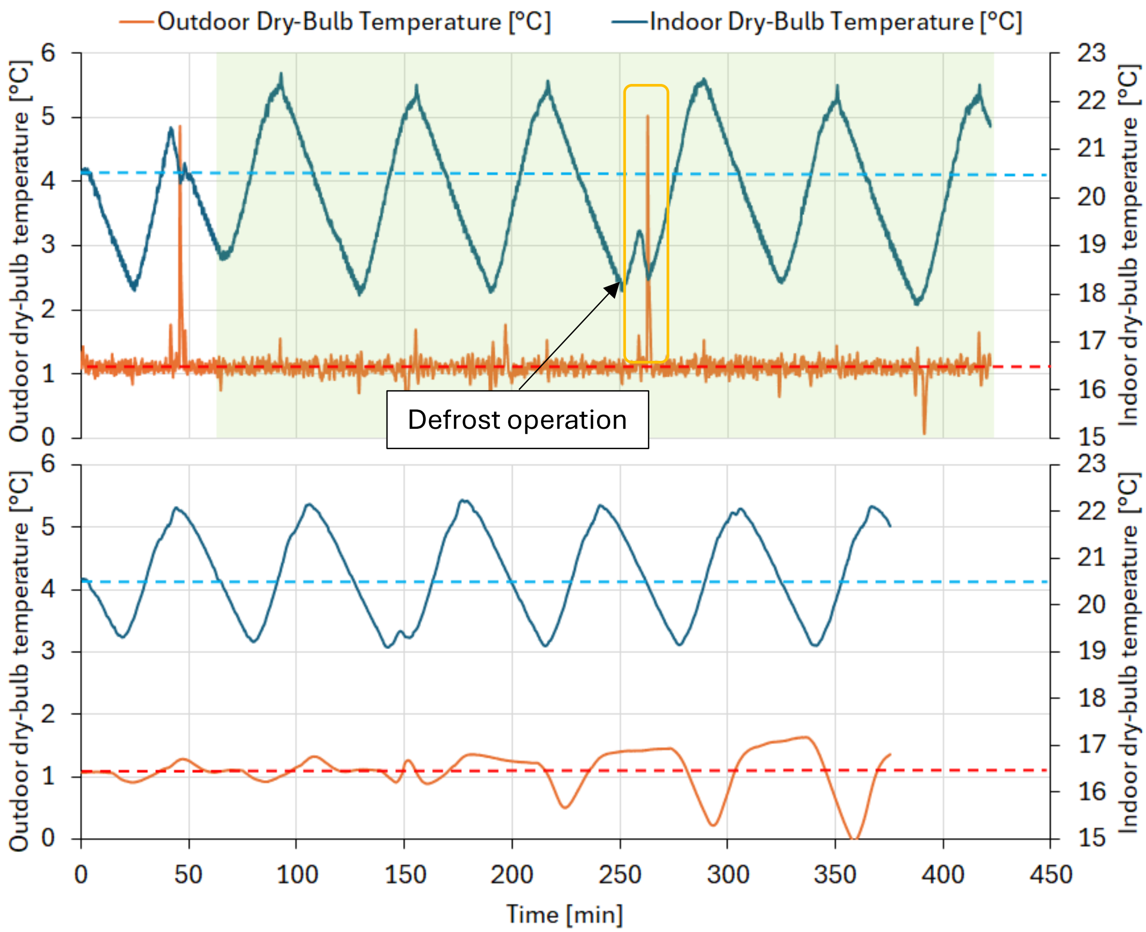


Figure 23. Unit A indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for heating continental test interval HD – Continental (1.1°C)

3.1.2.3. Continental Heating Test Interval HC (-8.3°C)

Figure 24 shows the behaviour of rating test interval HC (-8.3°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was part-load, but irregular operation prevented this unit from achieving convergence in this test interval, which means Unit A ran for 6 hours as required by SPE-07. Both facilities observed cyclic behaviour, with four defrost operations at NRCan and 2 at Lab A, and airflows being 50-75 CFM higher at Lab A. However, due to an issue in integrating the capacity while in defrost, the COP diverged significantly between tests. Because Unit A's indoor fan stays on during defrost operations, there is a cooling load introduced in the chamber that results in the RAT decreasing rapidly as can be seen by the indoor dry-bulb temperature in Figure 25 for NRCan.

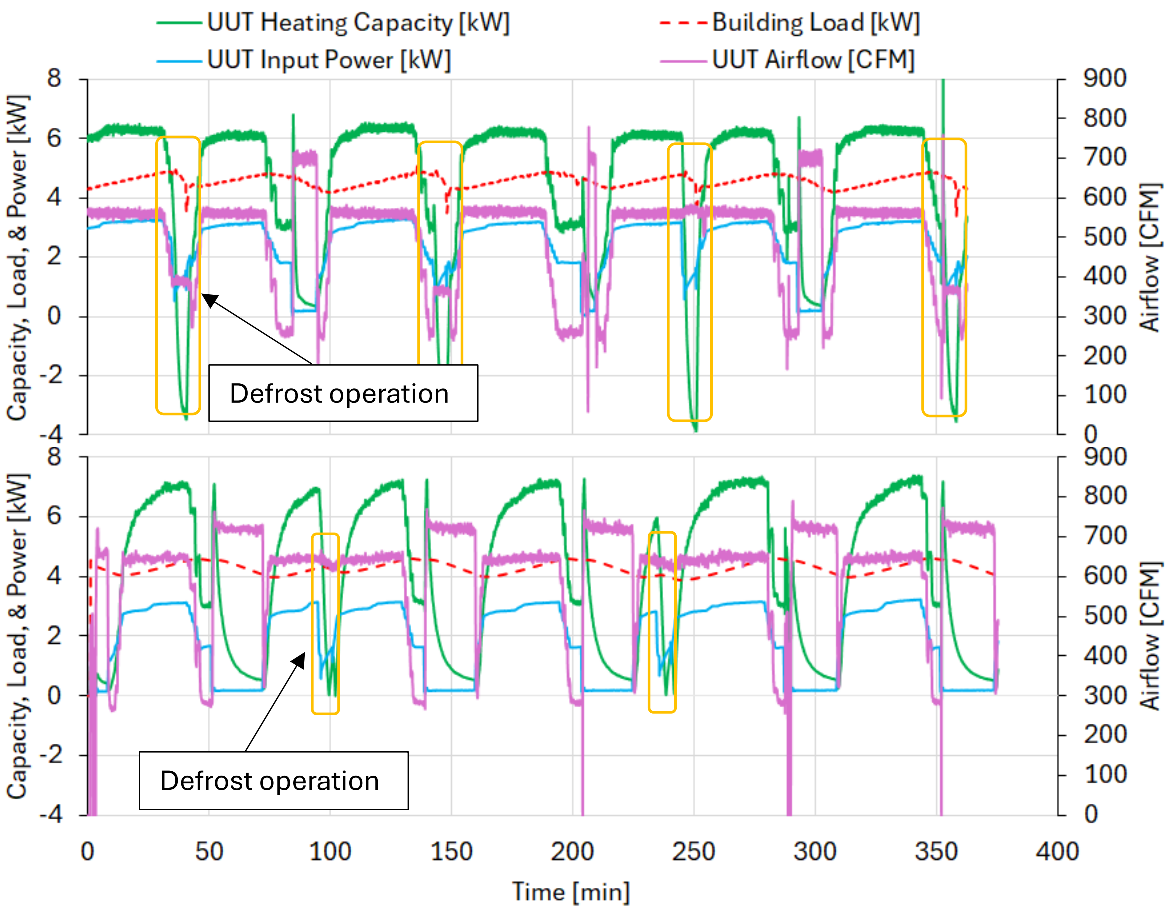


Figure 24. Unit A capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HC (-8.3°C)

This "cooling effect" was not included in Lab A's analysis and the measured heating capacity never dropped below 0 kW during a defrost operation. This led to a divergence in COP of 19.2%. Furthermore, the rate at which the heating capacity recovers at NRCan was faster than Lab A. This may be due to differences in instrumentation between laboratories, with Lab A using an RTD located in an aspirated psychrometer to measure dry-bulb temperature, and NRCan using the grid of thermocouples and thermopile in the air stream. It could also be due to the thermostat sensing a larger temperature drop during this period at NRCan and responding by ramping up the capacity more quickly.

Unit A otherwise behaved similarly between test facilities, cycling at a semi-regular interval and alternating between high airflows in off cycles, and modulating airflows in on cycles. Lab A observed higher peak capacities nearing the end of on cycles with approximately 7 kW of heating capacity, whereas it never output more than 6.2 kW of heating capacity at NRCan.

Figure 25 is a closeup focusing on one defrost cycle from each test facility and indicates how Lab A represented the absolute value of heating capacity as positive during the entire cycle. There are downstream consequences of this error in the implementation of the virtual building load over time (which is adjusted based on the measured capacity), so that the inverted sign in measured capacity affects the emulated room response.

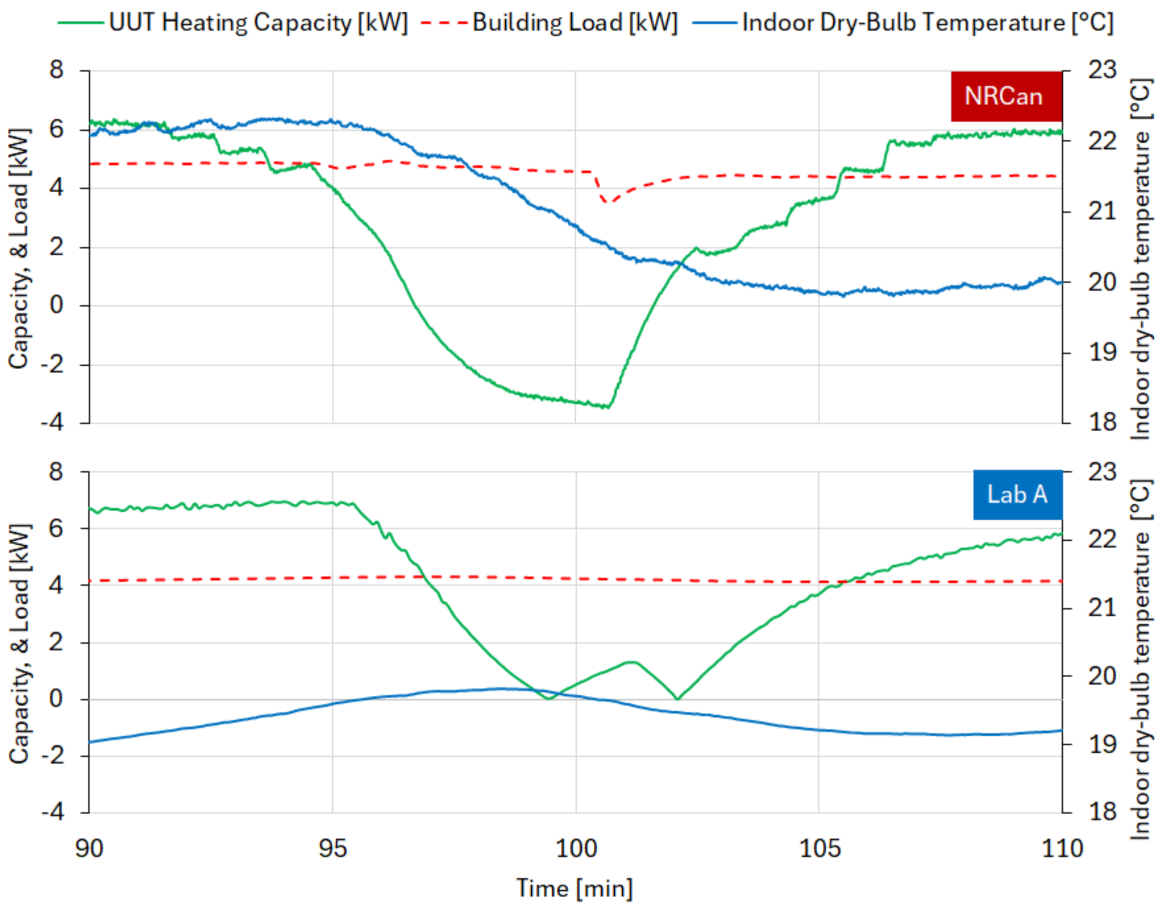


Figure 25. Unit A defrost operation during rating test interval HC

Figure 26 shows the indoor and outdoor dry-bulb temperature behaviour of test interval HC at NRCan and Lab A. Although both facilities maintained the indoor thermostat setpoint of 20.5°C within acceptable limits, Unit A maintained a slightly higher average temperature at NRCan. NRCan's outdoor chamber also experienced higher temperature fluctuations during defrost operations, which led to lower building loads during this time as is outlined in yellow in Figure 24. Interestingly, the pattern of similar end-of-cycle temperatures seen at other test conditions is not maintained in test interval HC, with the UUT ending its cycle at approximately 22°C at NRCan, and 21.5°C at Lab A.

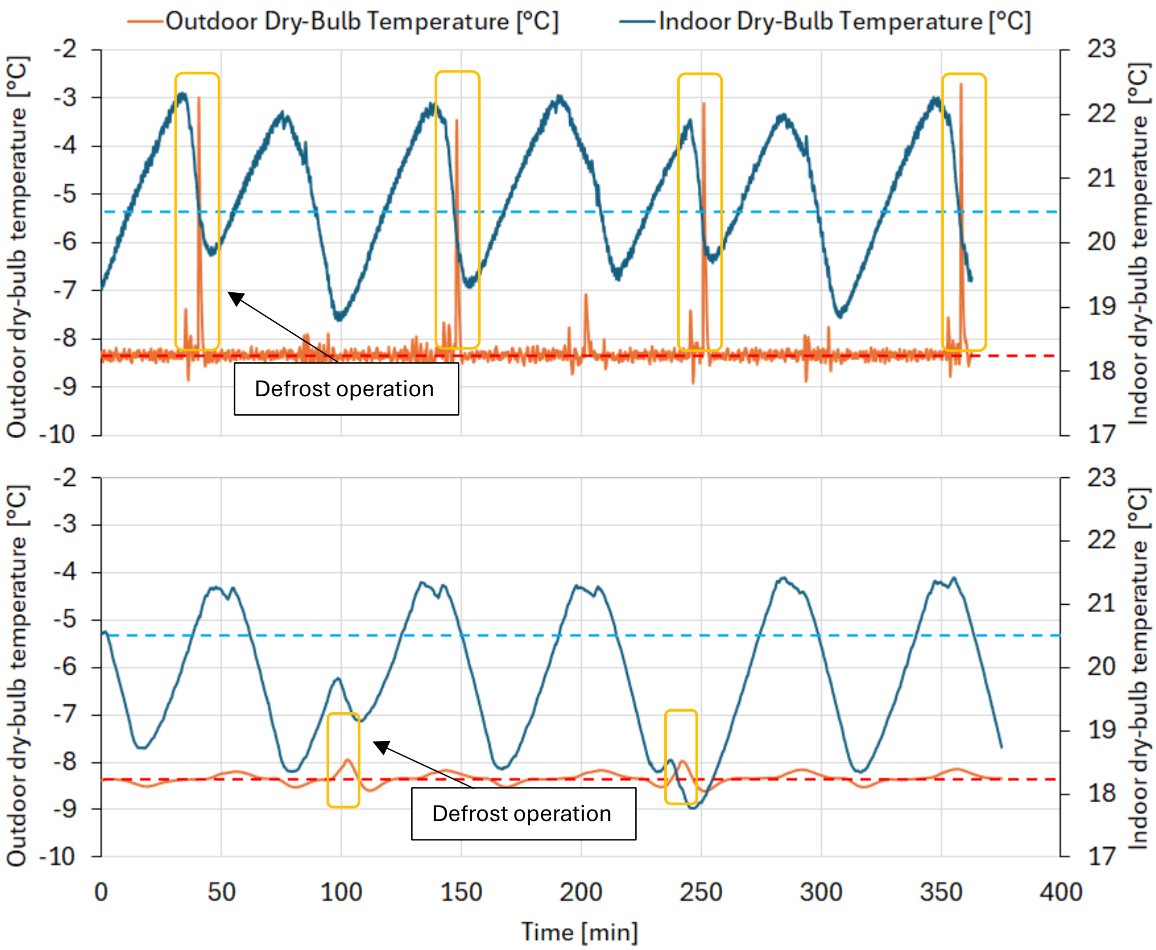


Figure 26. Unit A indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for heating continental test interval HC (-8.3°C)

3.1.2.4. Continental Heating Test Interval HB (-15°C)

Figure 27 shows the behaviour of rating test interval HB (-15.0°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was a full-load test, with NRCan having met convergence through two 20-minute period without a defrost operation, and Lab A using one complete defrost cycle as their convergence period.[4] Lab A observed the unit achieving a higher capacity by approximately 1 kW for the same input power consumption. This resulted in a 15% lower COP at NRCan. The airflow at Lab A was 55 CFM higher than NRCan and had a static pressure of 0.20 inH_2O versus 0.15 inH_2O at NRCan. The supply air temperature at Lab A was 40.4 °C and 40.3 °C at NRCan. The difference in capacity is therefore likely due to a different equipment and thermostat response which led to higher airflows at Lab A. It is also possible that the equipment could have initiated a higher capacity response following a defrost operation.

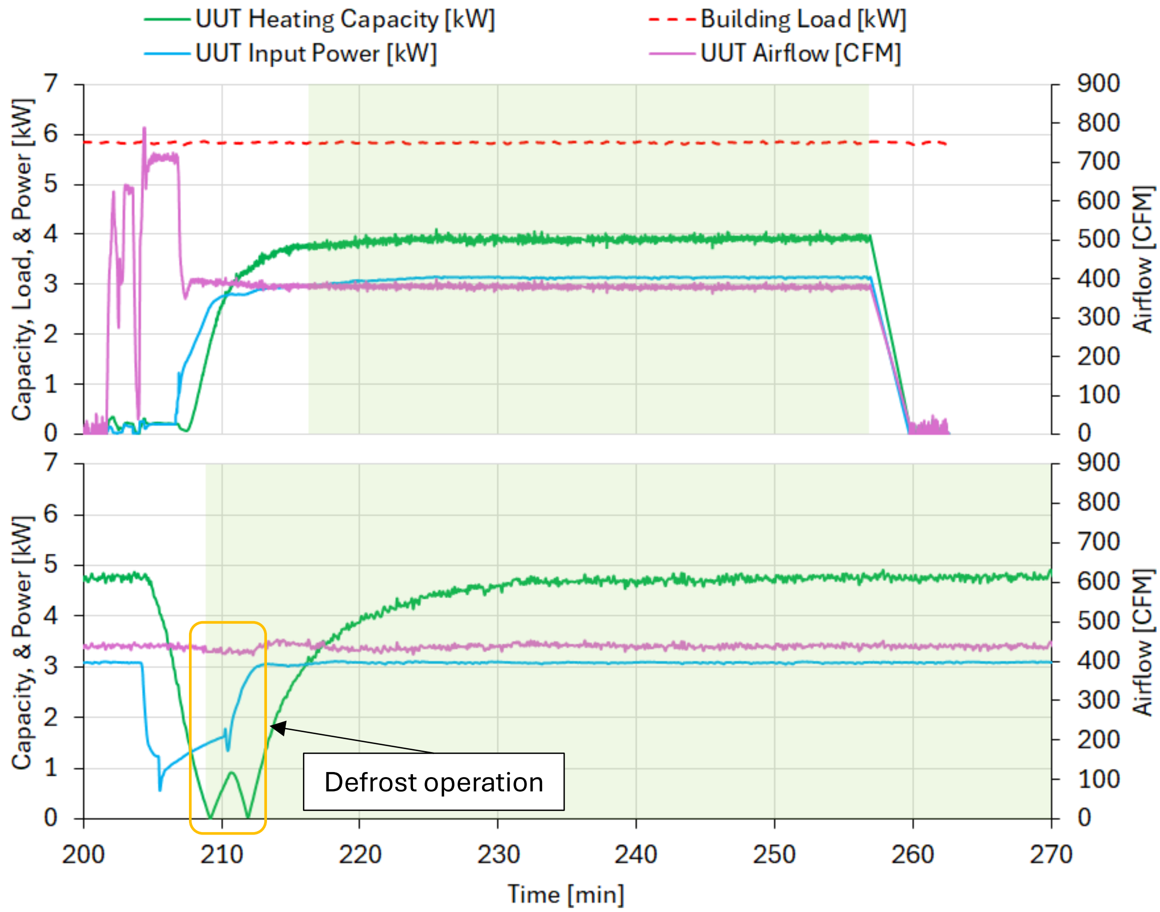


Figure 27. Unit A capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HB (-15.0°C)

[4] Loss of data at Lab A during the HB (-15°C) test interval resulted in no useable, "consecutive", data for multiple defrost cycles. The workaround used by Lab A was to include a single complete defrost cycle, although SPE-07:23 requires three consecutive defrost cycles for convergence.

Both facilities maintained the indoor and outdoor dry-bulb temperature setpoints during this full-load interval as shown in Figure 28.

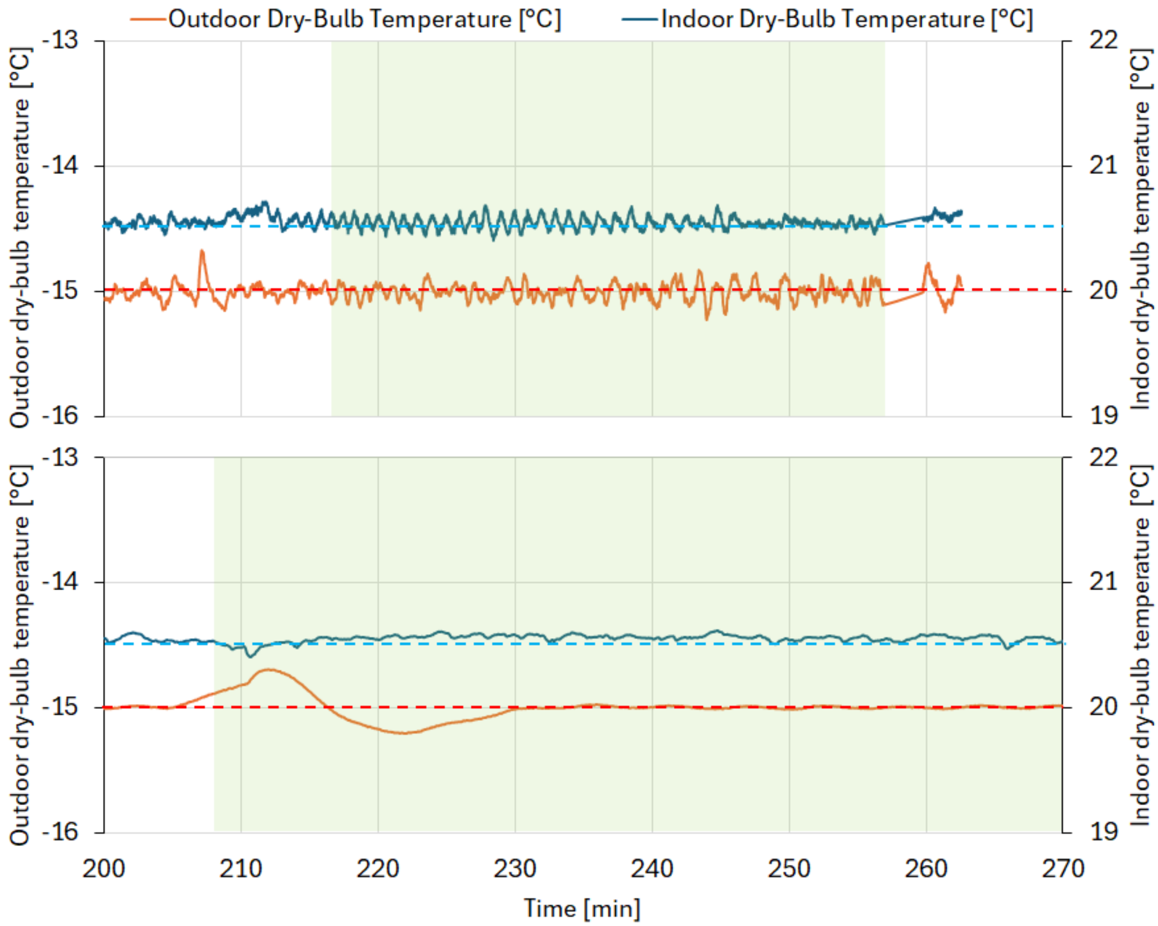


Figure 28. Unit A indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for heating continental test interval HB (-15.0°C)

3.1.3. Cooling Rating Test Summary

A complete series of SPE-07 cooling learning and rating tests (excluding the dry cooling test CA (45°C)) was conducted at Lab A and NRCan's facilities on Unit A to evaluate reproducibility. The results of capacity, input power, COP, and cooling $SCOP_C$ are presented below with Lab A's test results serving as the reference against which NRCan's results are compared. Several tests are analyzed in greater detail to support discussion of differences in equipment behaviour and to investigate the limits of reproducibility. The green-shaded areas in figures represent the convergence period in instances where more data is presented. Figures presenting results of indoor and outdoor dry-bulb temperature have a blue dashed line for the target indoor dry-bulb temperature and a red dashed line for the target outdoor dry-bulb temperature. For more detailed test results, see Section 7.1.1.

The cooling learning series tests determined that tests CE through CB in dry conditions, as well as tests CE through CC in humid conditions, were to be part-load tests, whereas test CB in humid condition was conducted as a full-load test. This was consistent between Lab A and NRCan as were the thermostat setpoints determined following the learning series tests as shown in Table 7. Thermostat setpoints determined in the learning test for conditions CD and CC in humid conditions differed slightly between laboratories, which led to consistently maintained indoor temperatures during testing, which is the intent of the thermostat calibration step during the learning tests.

Table 7. Unit A thermostat setpoints in cooling mode for NRCan and Lab A as determined by the cooling learning series tests

Test	Outdoor dry-bulb temperature, °F (°C)	Indoor dry-bulb control target, °F (°C)	TSTAT, °F (°C) (NRCan)	TSTAT, °F (°C) (Lab A)
CE_D	77 (25)	75 (23.9)	75 (23.9)	75 (23.9)
CD_D	86 (30)	75 (23.9)	75 (23.9)	75 (23.9)
CC_D	95 (35)	75 (23.9)	75 (23.9)	75 (23.9)
CB_D	104 (40)	75 (23.9)	74 (23.3)	74 (23.3)
CE_H	77 (25)	70 (21.1)	70 (21.1)	70 (21.1)
CD_H	86 (30)	70 (21.1)	70 (21.1)	69 (20.6)
CC_H	95 (35)	70 (21.1)	68 (20)	69 (20.6)
CB_H	104 (40)	Min	Min (52 (11.1))	Min (52 (11.1))

Following the cooling learning series tests, the cooling rating tests were completed in order of lowest to highest load. Total cooling capacity results, which includes latent and sensible cooling, are presented in Figure 29 and Figure 30 for dry and humid conditions respectively. The results indicate consistency between facilities, except for test CE in dry conditions.

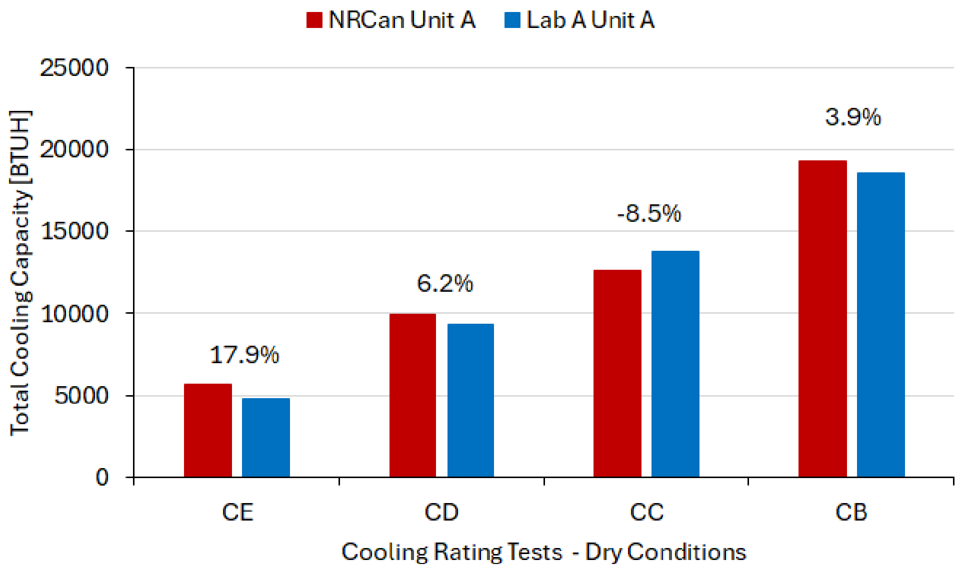


Figure 29. Total cooling capacity for Unit A, in dry conditions, NRCan and Lab A results

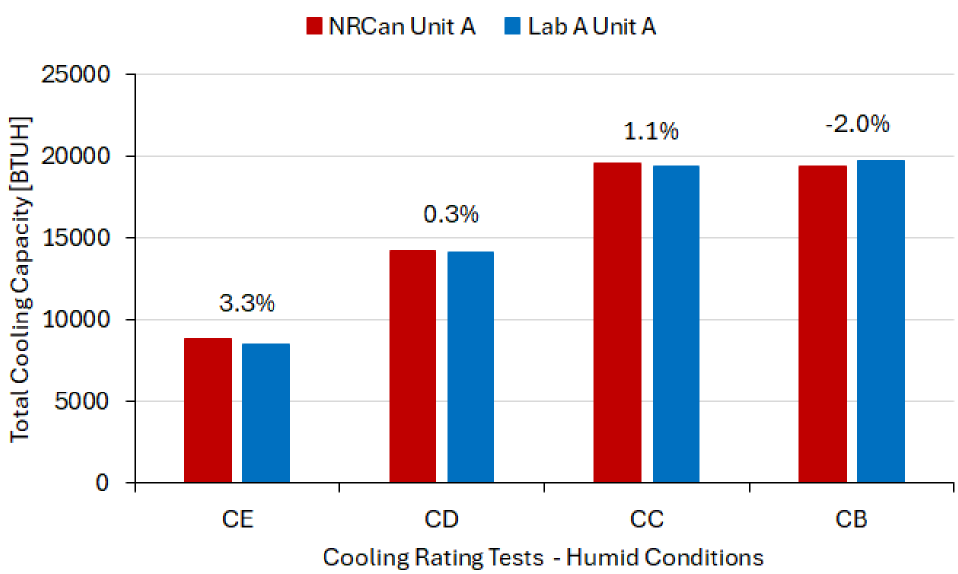


Figure 30. Total cooling capacity for Unit A, in humid conditions, NRCan and Lab A results

UUT integrated power consumption by test interval is presented in Figure 31 and Figure 32 for dry and humid conditions respectively. Here, we can observe a consistent trend towards more power consumption at NRCan across all test intervals.

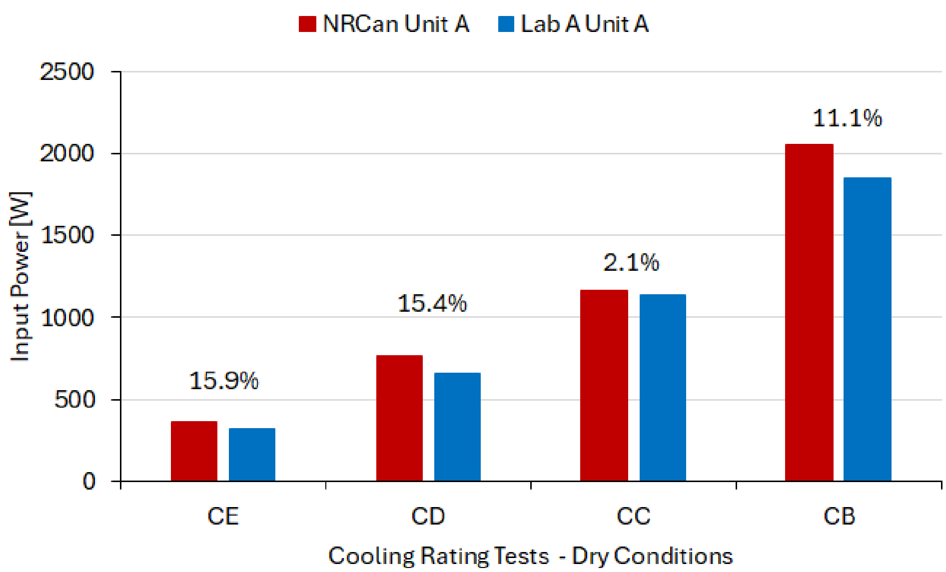


Figure 31. Cooling input power for Unit A, in dry conditions, NRCan and Lab A results

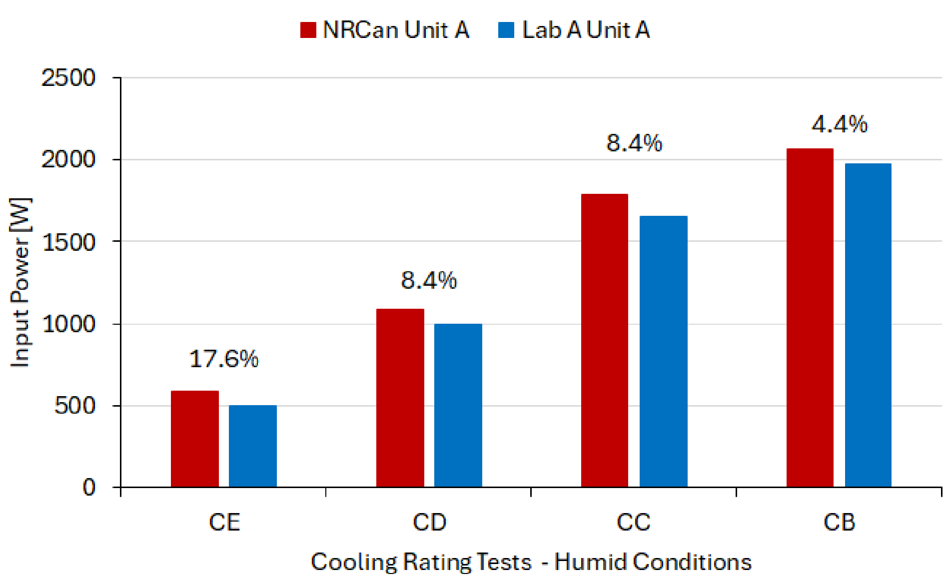


Figure 32. Cooling input power for Unit A, in humid conditions, NRCan and Lab A results

The resulting rating test COPs in Figure 33 and Figure 34 reveal that all tests had lower COPs at NRCan due to the higher input power, except for test CE in dry conditions.

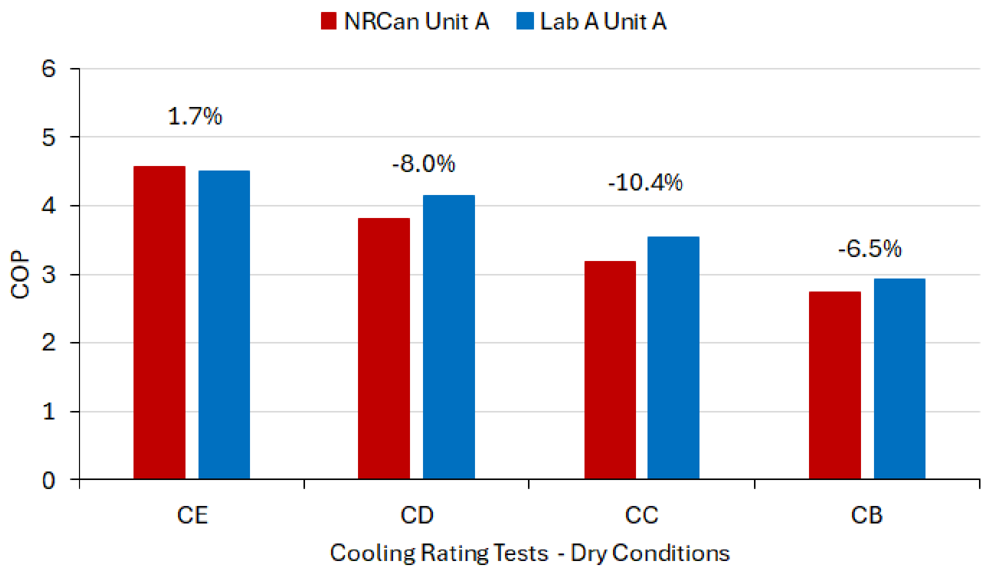


Figure 33. Cooling COP for Unit A, in dry conditions, NRCan and Lab A results

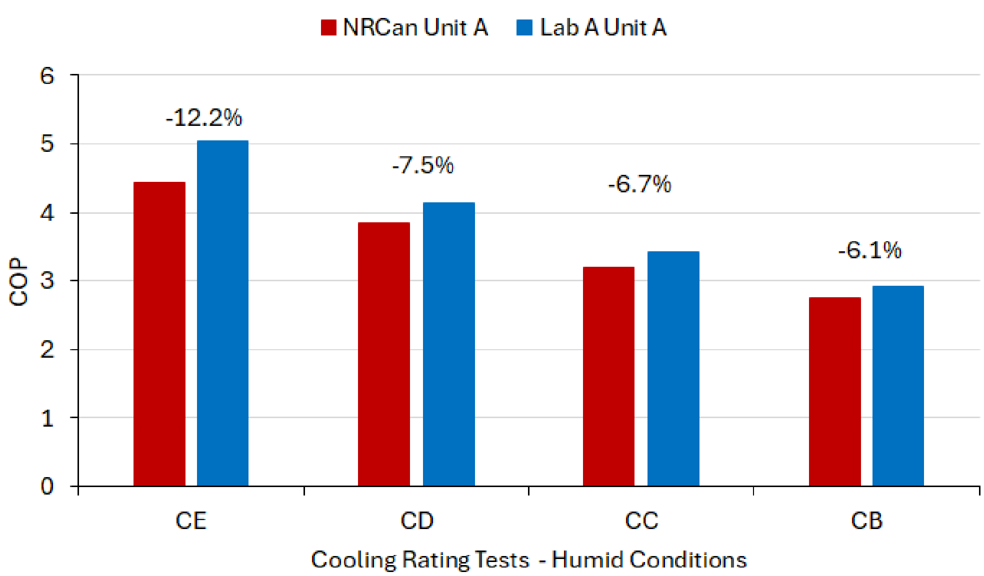


Figure 34. Cooling COP for Unit A, in humid conditions, NRCan and Lab A results

Finally, the seasonal COPs (SCOP) for the cooling bins are presented in Figure 35 and Table 8. The mean difference across all climate zones between Lab A and NRCan SCOPs was 8.6%, with a standard deviation of 1.4%. All climate zones were considered “fair” in terms of reproducibility with Lab A results. Rating tests CE humid, CC dry, and CB humid will be discussed in greater detail below.

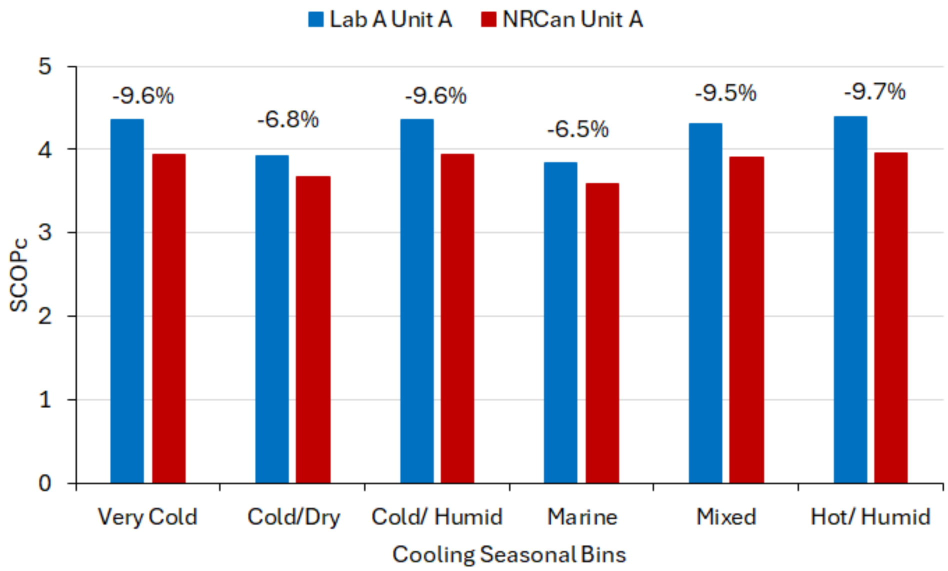


Figure 35. $SCOP_C$ for Unit A, NRCan and Lab A results

Table 8. $SCOP_C$ for Unit A, NRCan and Lab A results

	Very Cold	Cold/Dry	Cold/ Humid	Marine	Mixed	Hot/ Humid
Lab A	4.36	3.93	4.35	3.83	4.30	4.38
NRCan	3.94	3.66	3.93	3.58	3.89	3.96
ΔSCOP [%]	9.61	6.75	9.59	6.54	9.45	9.72

3.1.4. Individual Cooling Test Results

3.1.4.1. Humid Cooling Test Interval CE (25°C)

Figure 36 shows the behaviour of rating test interval CE humid (25°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was part-load and ran for 4 hours. Cyclic behaviour was observed at both facilities, with longer and more intense cycles at NRCan. The airflow during an on cycle was on average 100-200 CFM higher at NRCan with input power also being higher, while the total cooling capacities were similar at both facilities. Consequently, the off cycles were also longer at NRCan leading to two fewer cycles over the 4-hour test period. The integrated COP was 12.2% lower at NRCan.

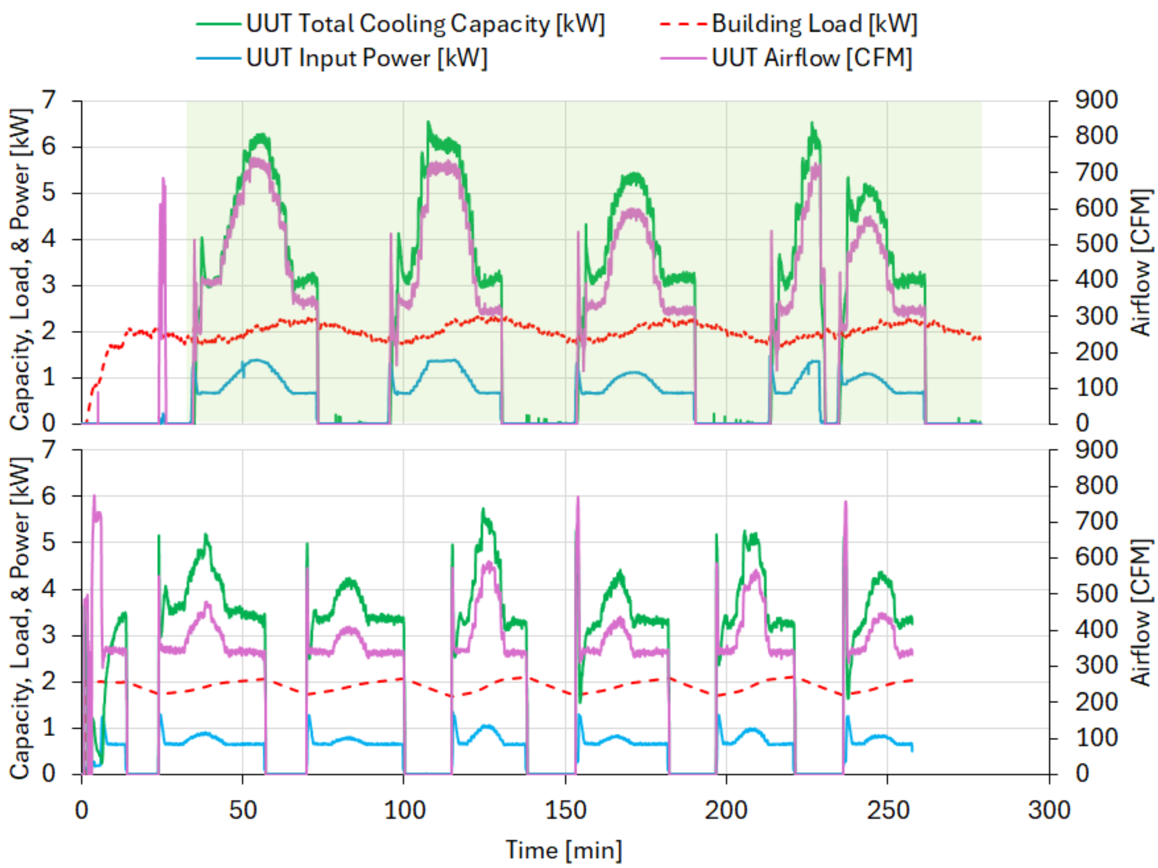


Figure 36. Unit A capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CE (25°C)

Figure 37 shows the indoor and outdoor dry-bulb temperature behaviour of test interval CE humid at NRCan and Lab A. The resulting indoor dry-bulb temperature was 0.5°C higher at Lab A. Indoor humidity ratios were similar with 0.0083 and 0.0080 at NRCan and Lab A respectively. The difference may be explained by different thermostat and equipment dynamic behaviour for this test interval because the indoor dry-bulb temperature never dropped more than 0.25°C below the thermostat setpoint at Lab A, while it dropped up to 1.0°C below the setpoint at NRCan.

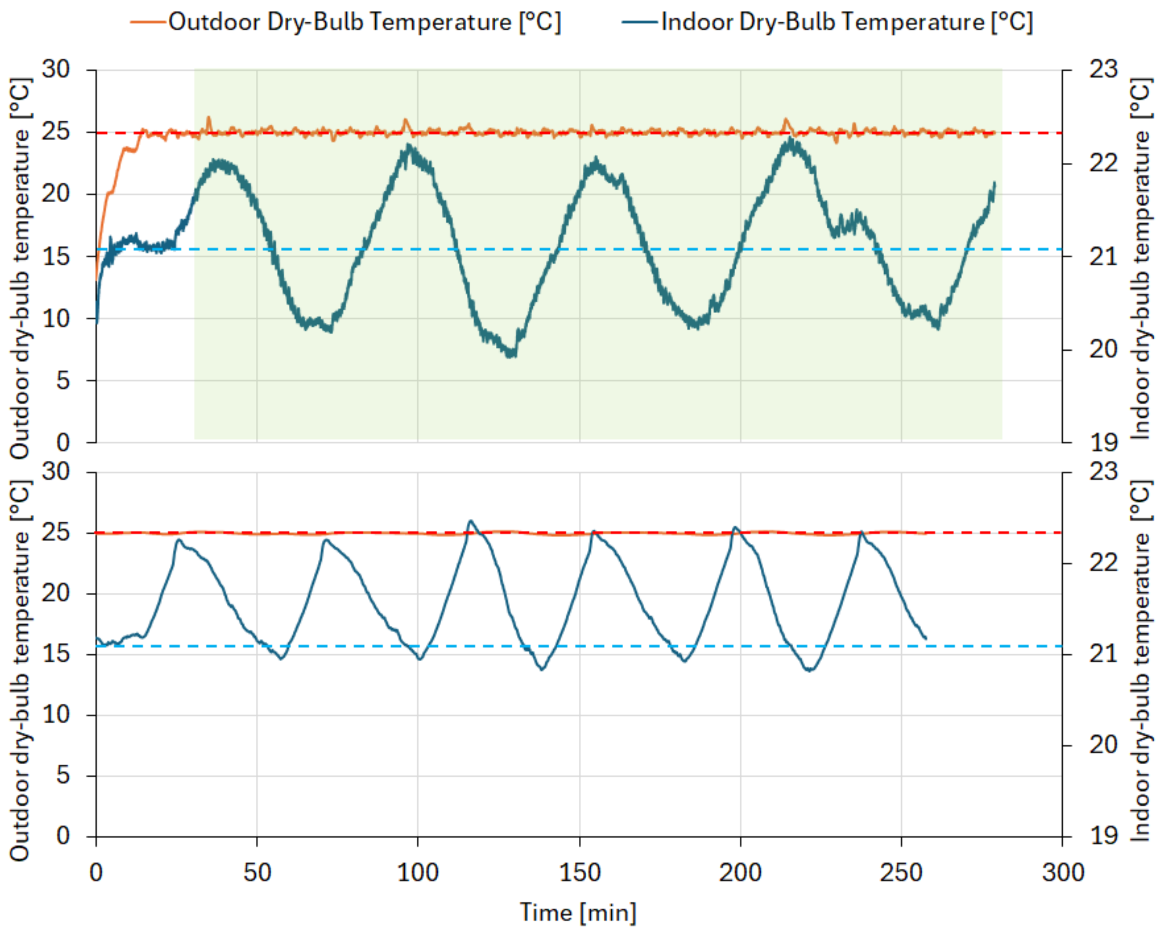


Figure 37. Unit A indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CE (25°C)

3.1.4.2. Dry Cooling Test Interval CC (35°C)

Figure 38 shows the behaviour of rating test interval CC dry (35°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was part-load and could converge based on regular variable speed or on/off cycling behaviour. However, convergence was not achieved at either facility, thereby extending the test interval to 4 hours as required by SPE-07. The UUT cycles were longer and more intense at NRCan than LabA, but also more consistent, while irregular cycling was observed at Lab A. The airflow during on cycles was approximately 50 CFM higher at Lab A. The total cooling capacity was higher on average at Lab A, but the steady-state capacity was similar in both tests at around 6 kW. The input power, however, was nearly identical with an integrated difference of only 25 W between the two tests. This resulted in a 10.4% lower COP at NRCan.

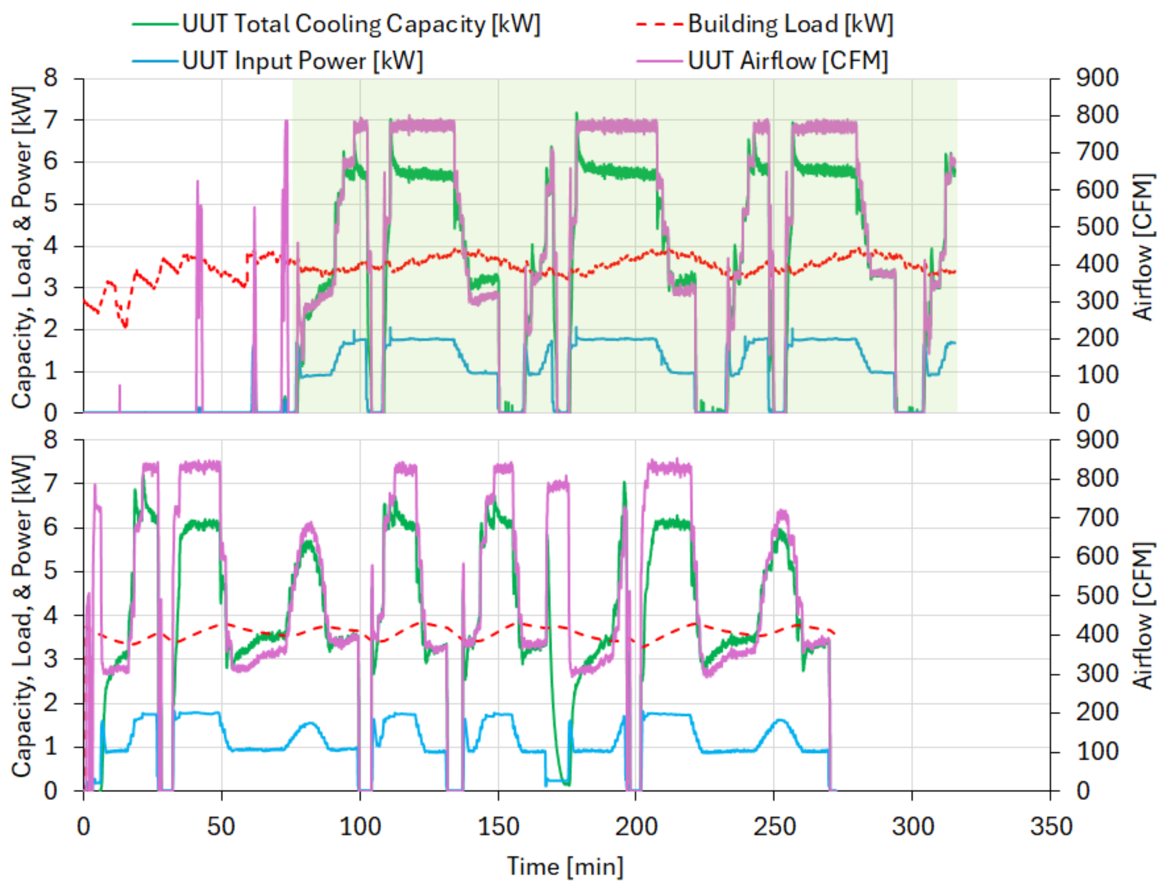


Figure 38. Unit A capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for cooling dry climate test interval CC (35°C)

Figure 39 shows the indoor and outdoor dry-bulb temperature behaviour of test interval CC dry at NRCan and Lab A. The integrated indoor dry-bulb temperatures were consistent between facilities with 24.7°C at NRCan and 24.5°C at Lab A. The UUT seemed to favour a large deadband with indoor temperatures reaching 0.5°C higher at UUT activation at NRCan than Lab A.

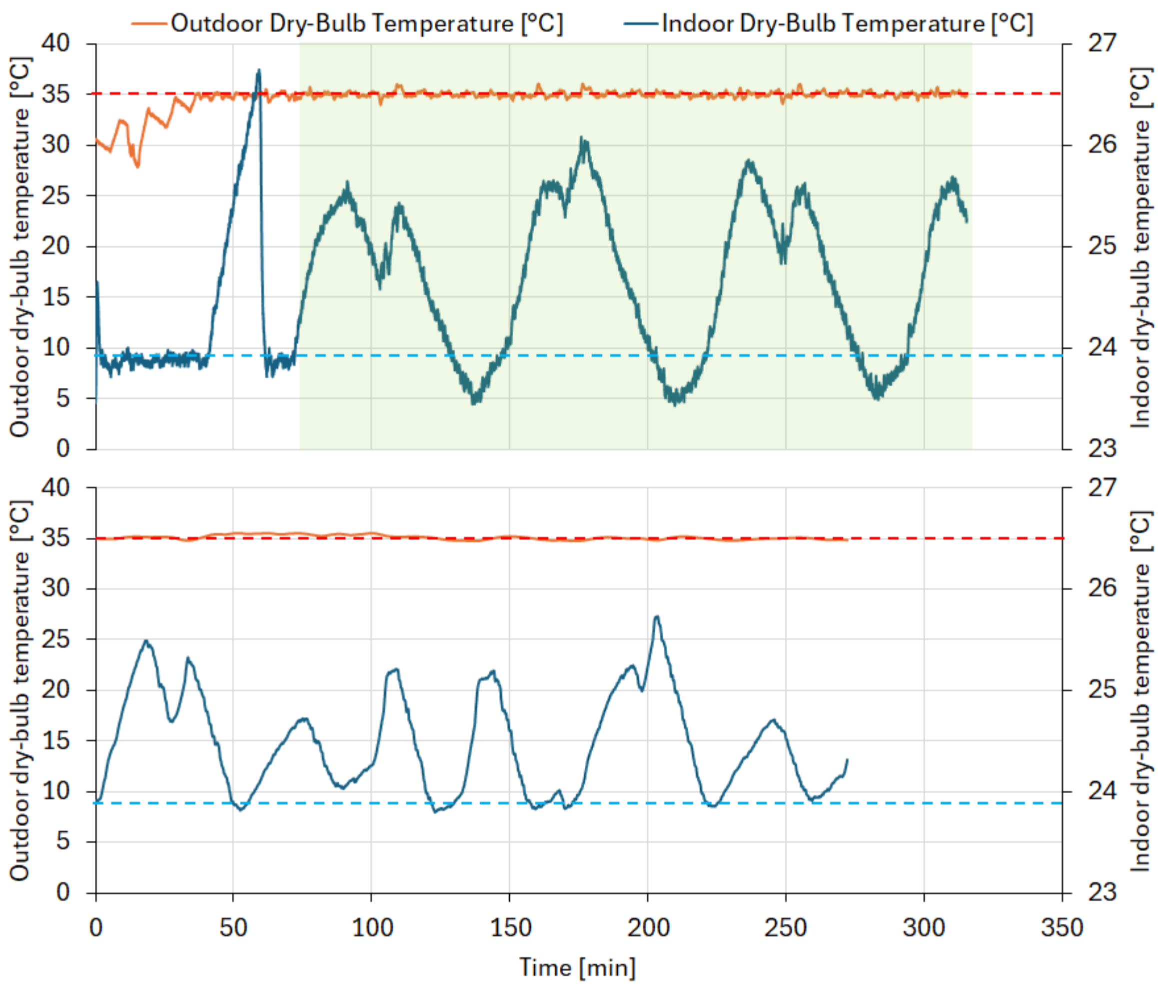


Figure 39. Unit A indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for cooling dry climate test interval CC (35°C)

The divergence in cooling capacity was not apparent in either latent or sensible capacities, with both facilities observing a similar split between the two as shown in Figure 40. The resulting integrated indoor humidity ratio was 0.0084 at NRCan and 0.0082 at Lab A.

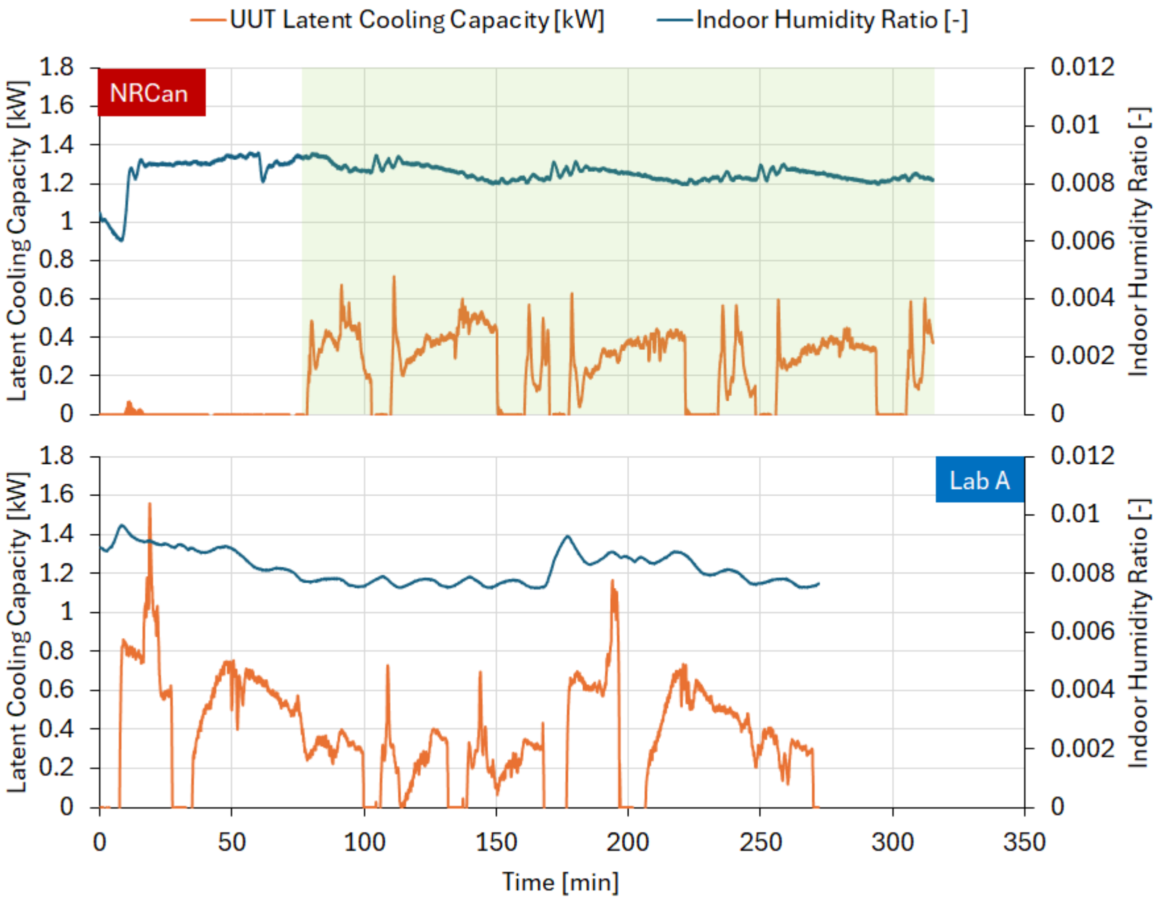


Figure 40. Unit A cooling dry climate test interval CC latent cooling capacity and indoor humidity ratio

3.1.4.3. Humid Cooling Test Interval CB (40°C)

Figure 41 shows the behaviour of rating test interval CB humid (40°C) for UUT capacity, building load and input power at NRCan and Lab A. The test interval was full load at both facilities with NRCan and Lab A achieving convergence after two consecutive 20-minute periods. The total cooling capacities were similar at both facilities with only 0.1 kW of difference, while the input power was approximately 80 W higher at NRCan, leading to a difference in COP of 6.1% between Lab A and NRCan. The split in input power difference was even between the indoor and outdoor units with 32 W and 50 W more power consumed at NRCan respectively. The airflows were the same between test facilities at 800 CFM, with identical indoor unit external static pressures of 0.75 inH_2O.

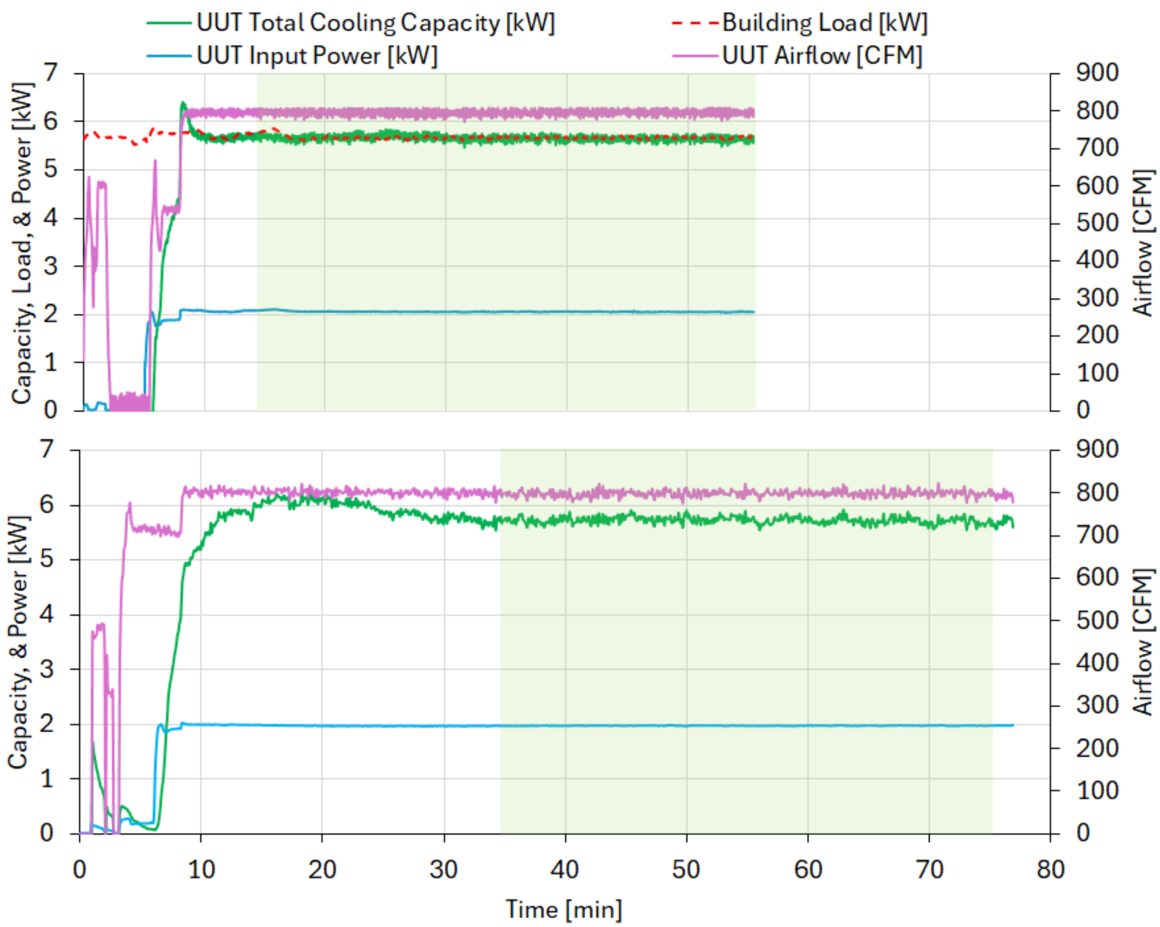


Figure 41. Unit A capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CB (40°C)

Figure 42 shows the indoor and outdoor dry-bulb temperature behaviour of test interval CB humid at NRCan and Lab A. Both facilities maintained a constant indoor dry-bulb temperature during the test as required in a full-load interval.

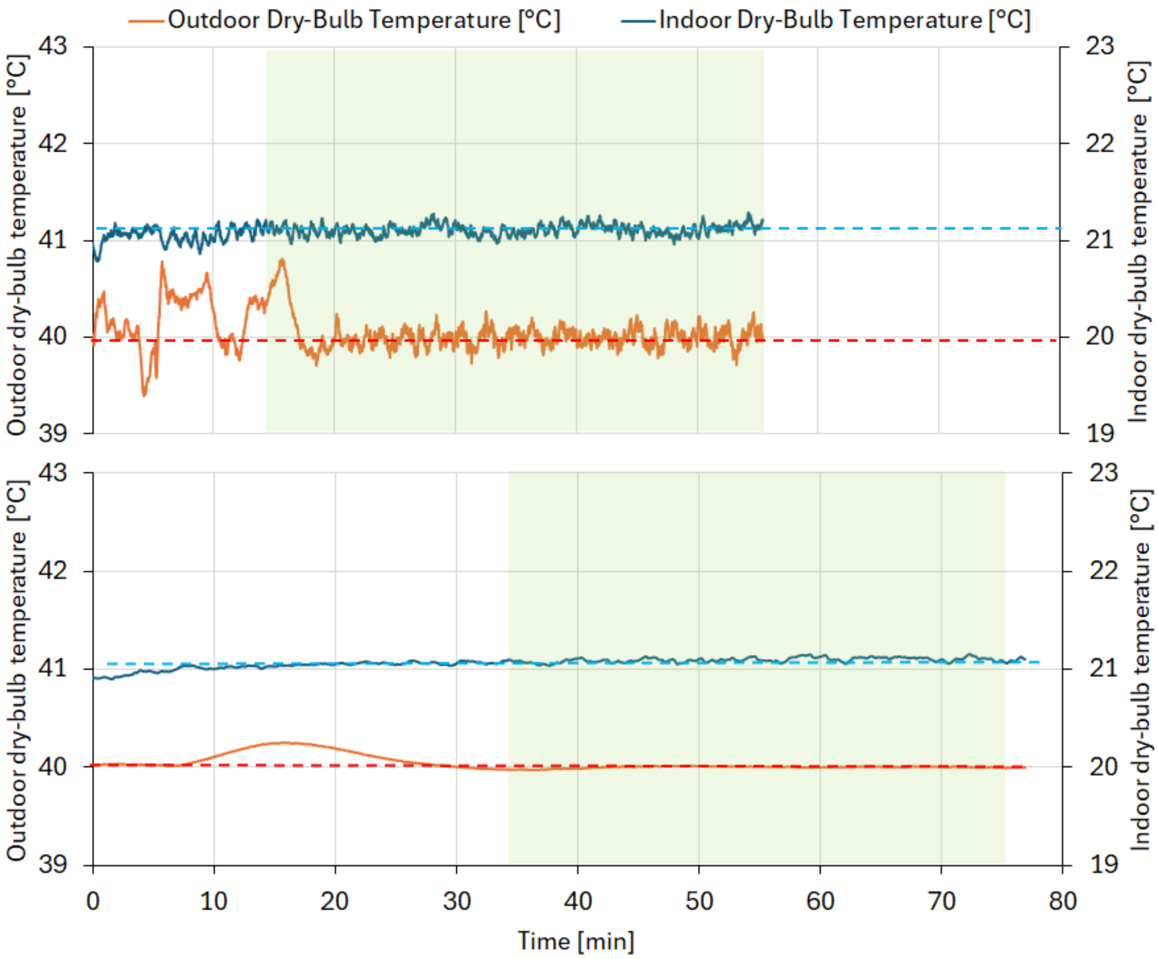


Figure 42. Unit A indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CB (40°C)

This test interval presents a case where the UUT behaved identically at both test facilities and still resulted in a COP difference of 6.1%. This result could be explained by differences in installation between test facilities that led to issues such as increased outdoor unit fan power consumption due to the height restrictions in the outdoor chamber at NRCan, as well as differences in instrumentation used for calculating the total cooling capacity such as RTDs used by Lab A and thermocouple grids used by NRCan.

3.2. Unit C

3.2.1. Heating Rating Test Summary

A complete series of SPE-07 heating learning and rating tests was conducted at Lab A and NRCan's facilities on Unit C to evaluate reproducibility. The results of capacity, input power, COP, and seasonal COP (SCOP) are presented with Lab A's test results serving as the reference against which NRCan's results are compared. Several tests are analyzed in greater detail to support discussion of differences in equipment behaviour and to investigate the limits of reproducibility. The green-shaded areas in figures represent the convergence period in instances where more data is presented. Figures presenting results of indoor and outdoor dry-bulb temperature have a blue dashed line for the target indoor dry-bulb temperature and a red dashed line for the target outdoor dry-bulb temperature. For more detailed test results, see Section 7.1.2.

The heating learning series tests determined that tests HF through HD were part-load tests, whereas tests HC through HL (LCT – Lowest Catalogued Temperature) were full-load tests. The tests load conditions were consistent between Lab A and NRCan. However, the thermostat setpoints were not consistent between facilities as shown in Table 9. Furthermore, the HL learning series test conducted at NRCan revealed that Unit C would not output its rated capacity at its lowest catalogued temperature and would intermittently cut-out with no defrost operations. This test interval was therefore removed from the SCOP calculations for the purposes of evaluating reproducibility, although Lab A's results are still presented in the Figures below.

Table 9. Unit C thermostat setpoints in heating mode for NRCan and Lab A as determined by the heating learning series tests

Test	Outdoor dry-bulb temperature, °F (°C)	Indoor dry-bulb control target, °F (°C)	TSTAT, °F (°C) (NRCan)	TSTAT, °F (°C) (Lab A)
HF_C	54 (12.2)	74 (23.3)	69 (20.6)	72 (22.2)
HE_C	47 (8.3)	74 (23.3)	69 (20.6)	72 (22.2)
HD_C	34 (1.1)	74 (23.3)	70 (21.1)	72 (22.2)
HC_C	17 (-8.3)	Max	Max (83 (28.3))	Max (83 (28.3))
HB_C	5 (-15)	Max	Max (83 (28.3))	Max (83 (28.3))
HL_C	-13 (-25)	Max	Max (83 (28.3))	Max (83 (28.3))
HD_M	34 (1.1)	74 (23.3)	70 (21.1)	72 (22.2)

Following the heating learning series tests, the heating rating tests were completed in order of lowest to highest load. Heating capacity results presented in Figure 43 reveal that tests HC and HB had higher integrated capacities at NRCan's test facility, whereas other tests were consistent between facilities. On the other hand, UUT integrated power consumption presented in Figure 44 was consistently higher at NRCan's test facility for all rating tests, with test HC and HB's deviation from results at Lab A being particularly significant. The resulting rating test COPs shown in Figure 45 reveal that all heating test intervals at NRCan were lower performing than Lab A. Rating tests HE, HC, and HL will be discussed in greater detail below.

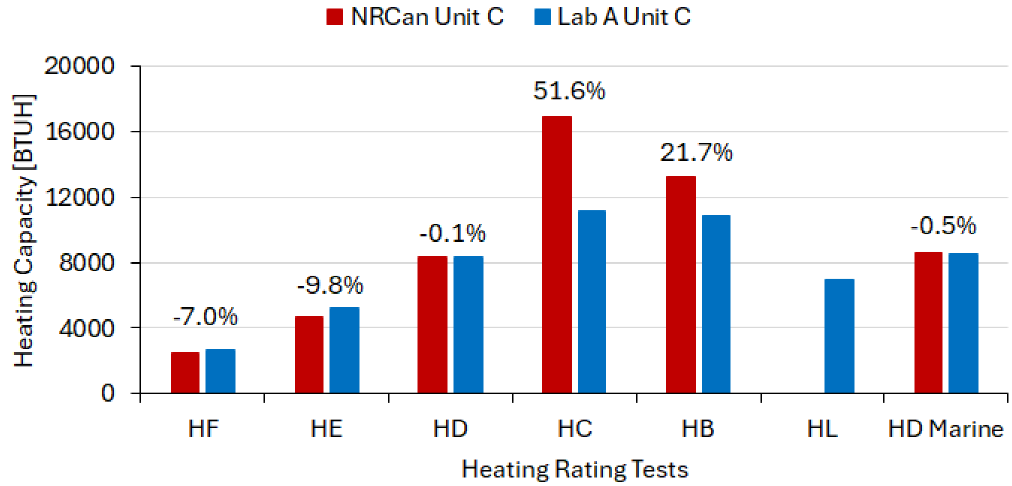


Figure 43. Heating capacity for Unit C, NRCan and Lab A results

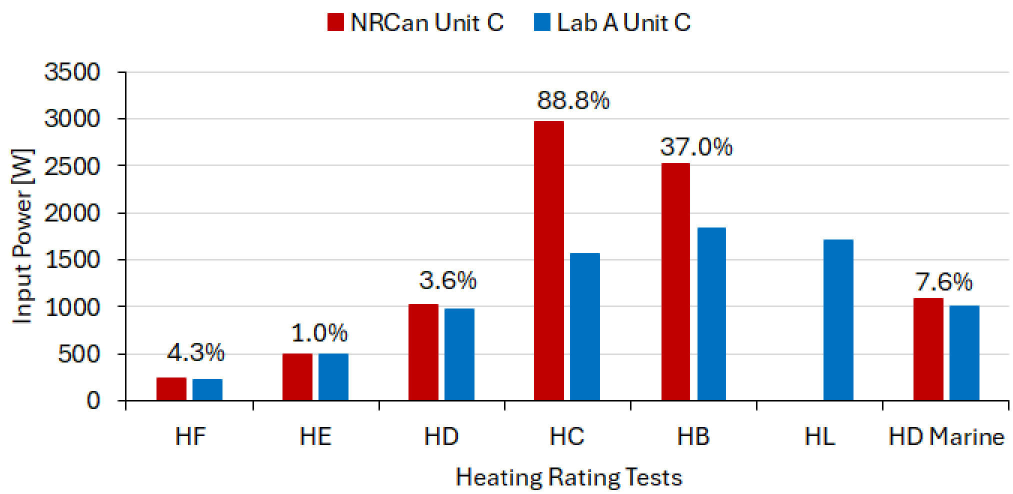


Figure 44. Heating input power for Unit C, NRCan and Lab A results

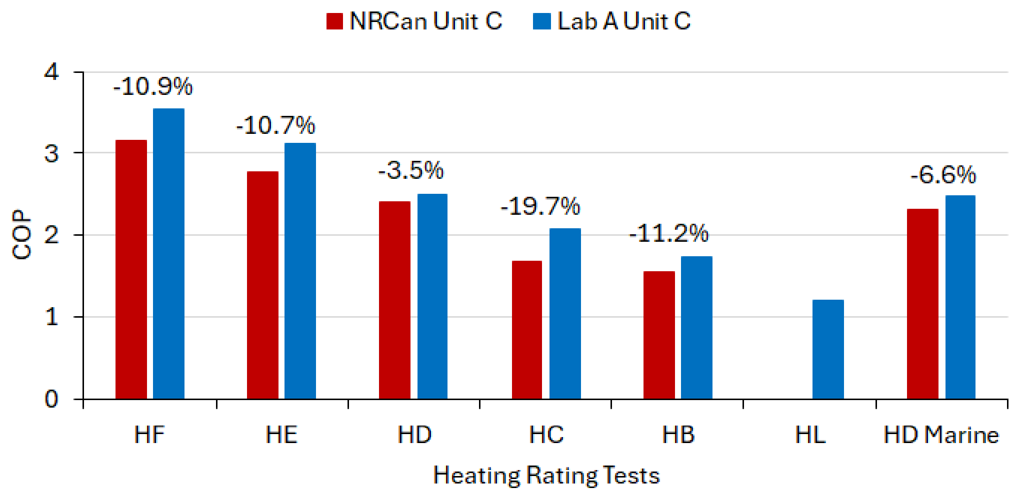


Figure 45. Heating COP for Unit C, NRCan and Lab A results

Finally, the heating seasonal COPs ($SCOP_H$) for the heating bins are presented in Figure 46 and Table 10. The mean difference across all climate zones between Lab A and NRCan $SCOP_H$ was 9.8%, with a standard deviation of 1.1%. All climate region SCOPs are "fair" in terms of reproducibility.

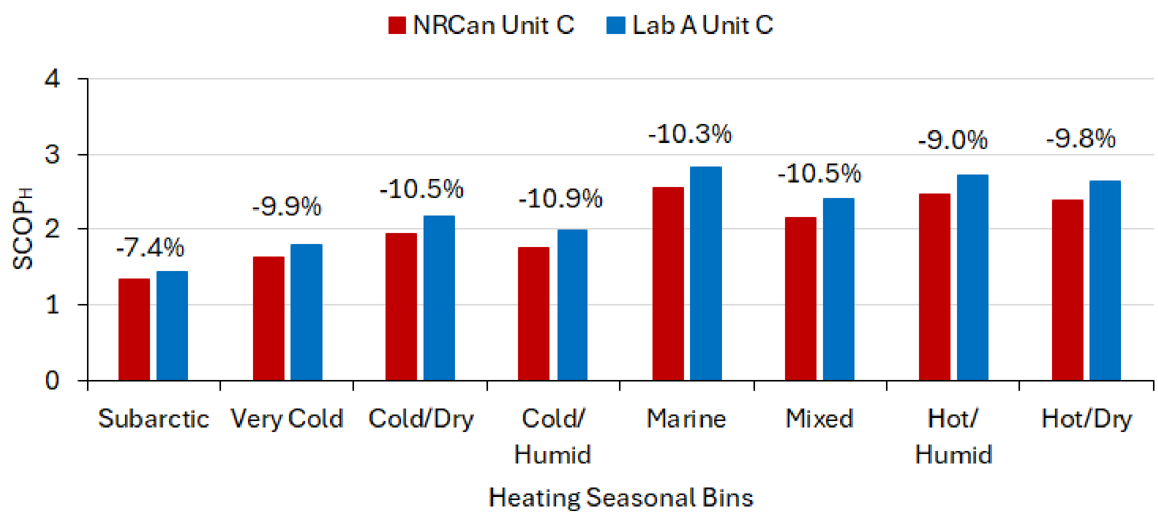


Figure 46. $SCOP_H$ for Unit C, NRCan and Lab A results

Table 10. $SCOP_H$ for Unit C, NRCan and Lab A results

	Subarctic	Very Cold	Cold/Dry	Cold/ Humid	Marine	Mixed	Hot/ Humid	Hot/Dry
Lab A	1.45	1.80	2.17	1.98	2.83	2.40	2.72	2.64
NRCan	1.34	1.62	1.94	1.77	2.54	2.15	2.47	2.38
ΔSCOP [%]	7.4	9.9	10.5	10.9	10.3	10.5	9.0	9.8

3.2.1. Individual Heating Test Results

3.2.1.1. Continental Heating Test Interval HE (8.3°C)

Figure 47 shows the capacity, building load, input power and airflow of Unit C for heating test interval HE (8.3°C). NRCan achieved convergence following two full cycles in part-load on/off cycling operation, while Lab A completed the test after 4 hours due to irregular system behaviour. Unit C's behaviour was similar at both test facilities with heating capacities reaching approximately 4.5 kW and input power topping off at 1.5 kW. However, the integrated capacity was higher by 0.5 kW in Lab A, while the integrated input power was the same between facilities. Airflows were higher at Lab A by 20 CFM. Cycles were slightly shorter on average at NRCan.

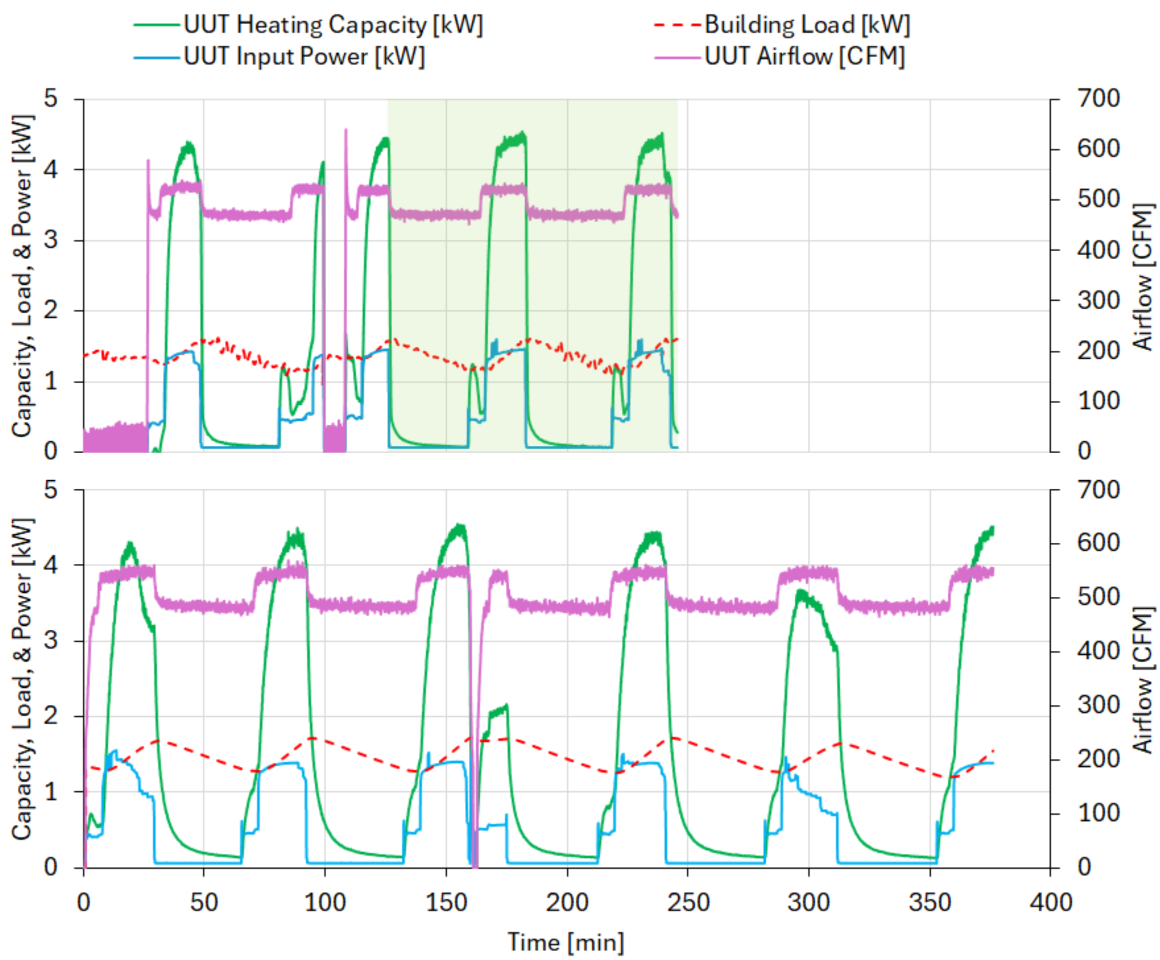


Figure 47. Unit C capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HE (8.3°C)

Figure 48 shows the indoor and outdoor dry-bulb temperatures of Unit C for test interval HE. The integrated indoor temperature was 74.3 °F at NRCan and 75.3 °F at Lab A. Unit A seems to have maintained the indoor temperature more closely to the thermostat setpoint of 74 °F at NRCan but overall equipment and thermostat behaviour was similar. This difference in maintained indoor dry-bulb temperature led to a 0.1 kW higher average building load at Lab A. The resulting integrated COP was 10.4% lower at NRCan.

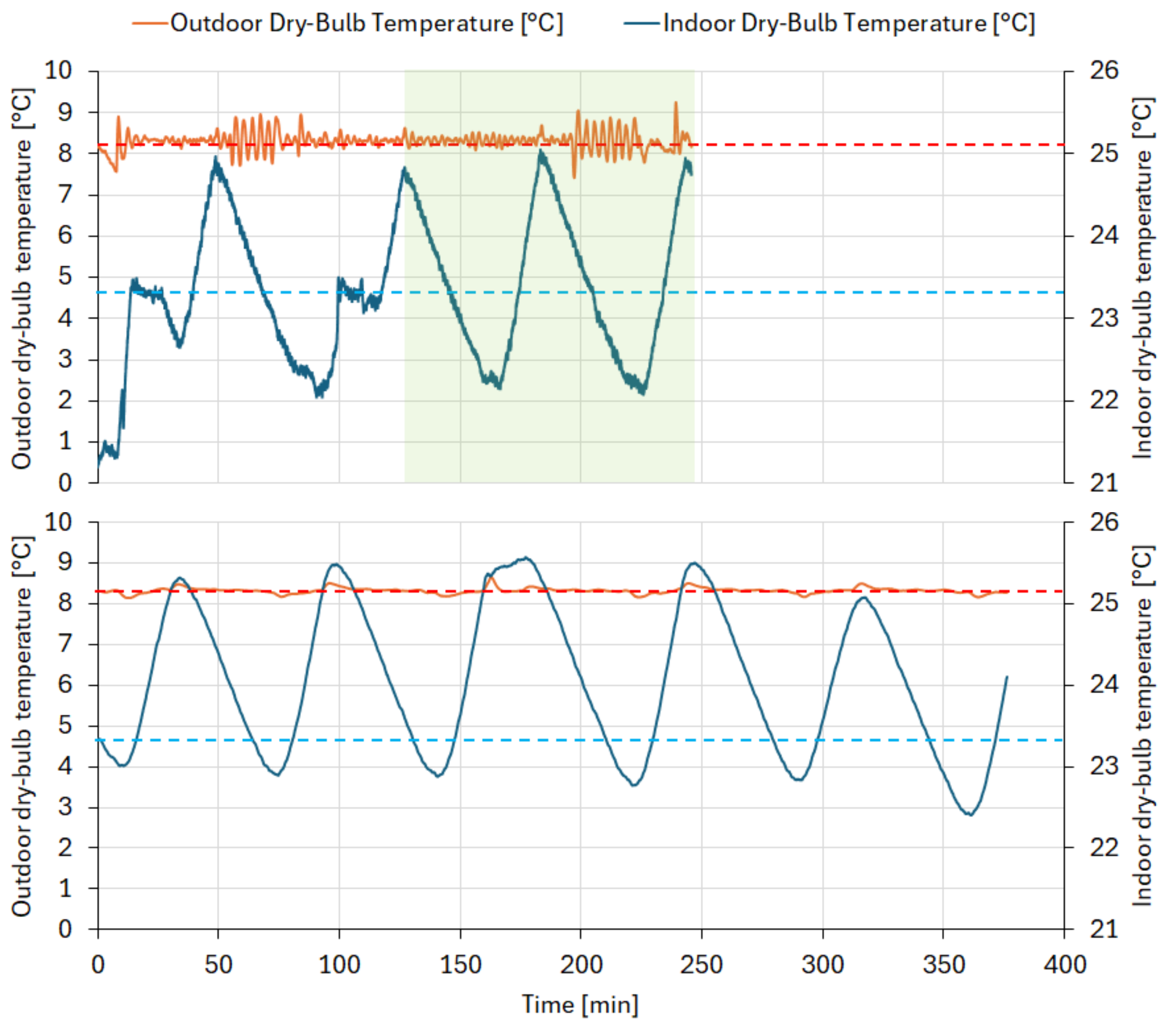


Figure 48. Unit C indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for heating continental test interval HE (8.3°C)

3.2.1.2. Continental Heating Test Interval HC (-8.3°C)

Figure 49 shows the capacity, input power and airflow of Unit C for heating test interval HC (-8.3°C). This test was a full-load interval at both facilities and convergence was achieved in steady-state operation without defrosts. However, Unit C output 5 kW of heat at NRCan and 3.3 kW at Lab A. Similarly, the input power was 1.5 kW higher at NRCan, while airflows were similar.

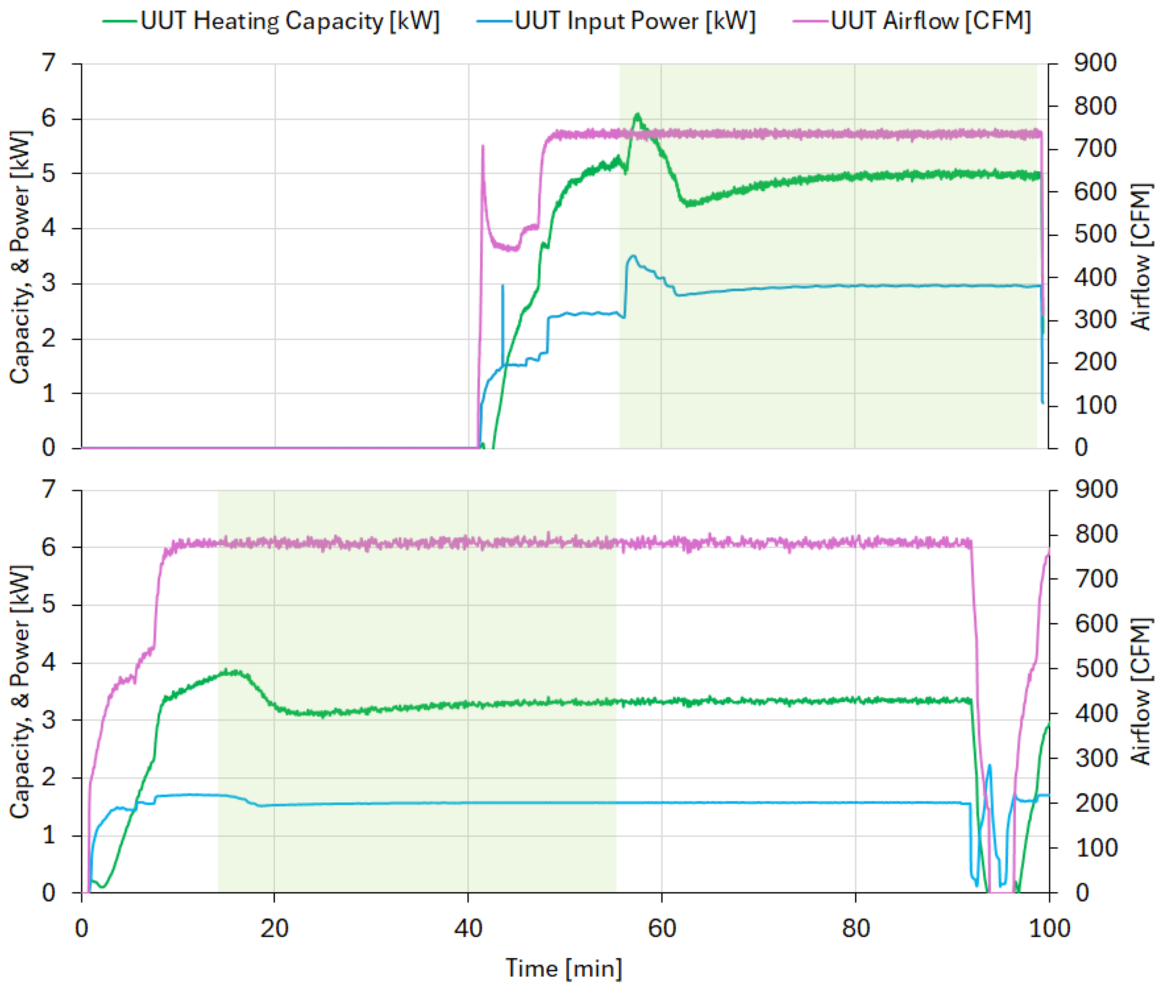


Figure 49. Unit C capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HC (-8.3°C)

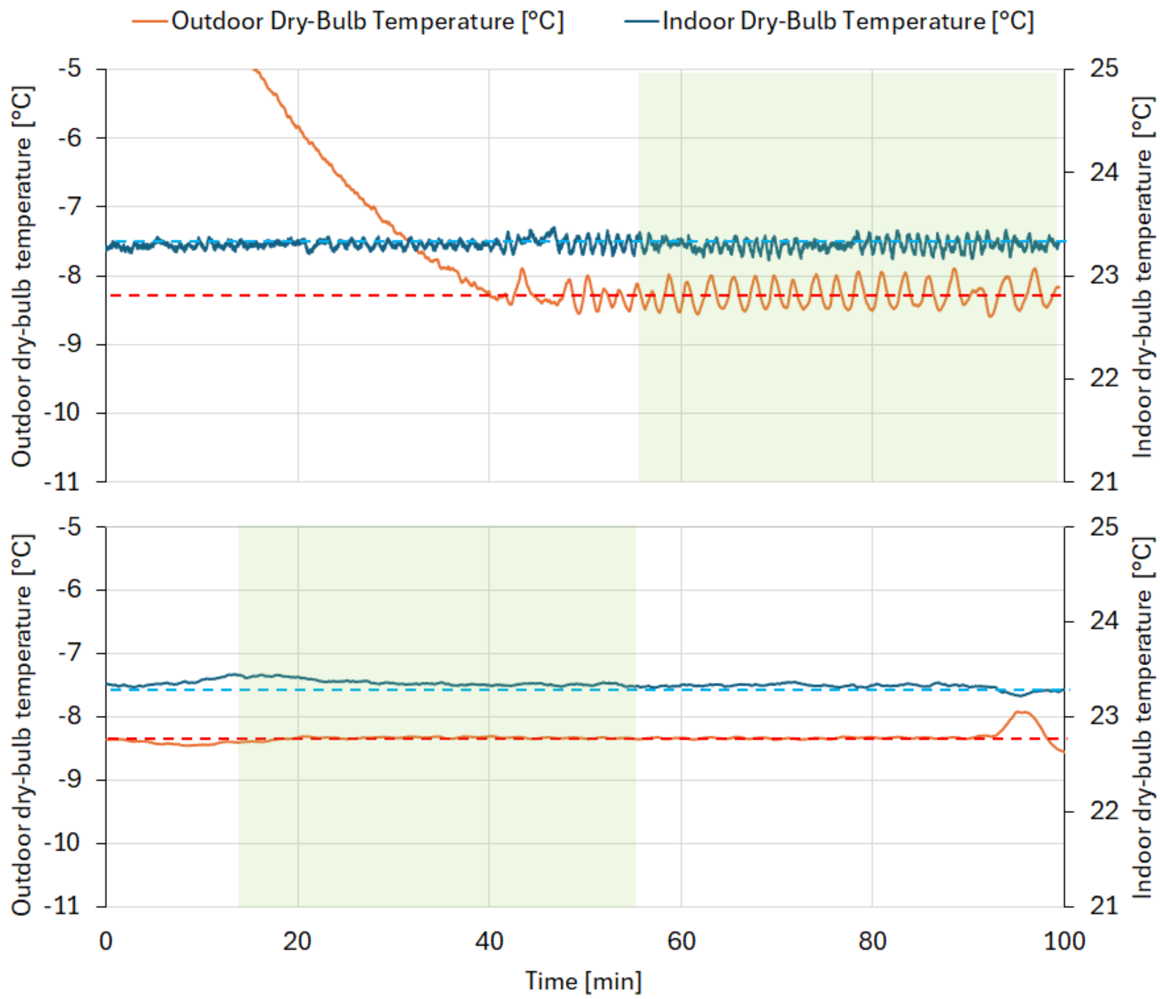


Figure 50. Unit C indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for heating continental test interval HC (-8.3°C)

NRCan and Lab A repeated this test three times with NRCan obtaining similar results and equipment behaviour each time as shown in Figure 51 through Figure 53. Lab A, however, observed a different equipment response in the third HC test conducted, which was more aligned with observed behaviour and performance at NRCan. This change in equipment behaviour is difficult to explain. Some thermostats have onboard demand control functions which could affect the equipment's performance, but no such features are known to have been active at either test facilities during testing.

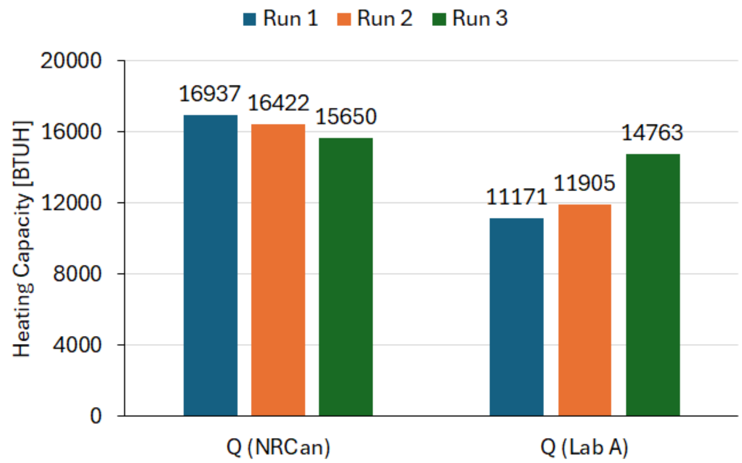


Figure 51. Unit C heating capacity results for test interval HC (-8.3°C) repeated three times at NRCan and Lab A

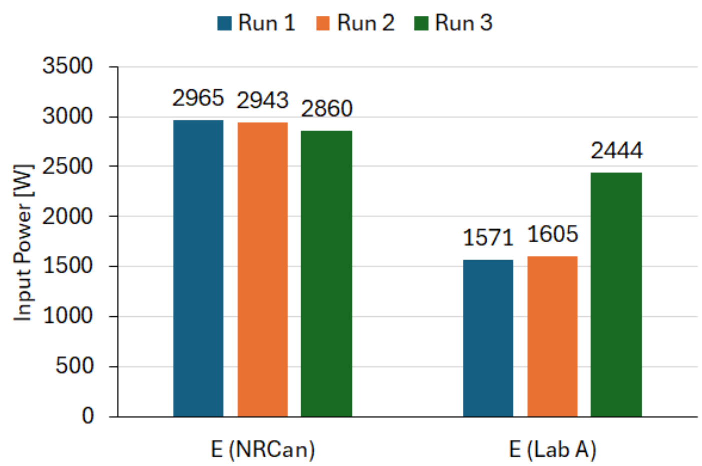


Figure 52. Unit C input power results for test interval HC (-8.3°C) repeated three times at NRCan and Lab A

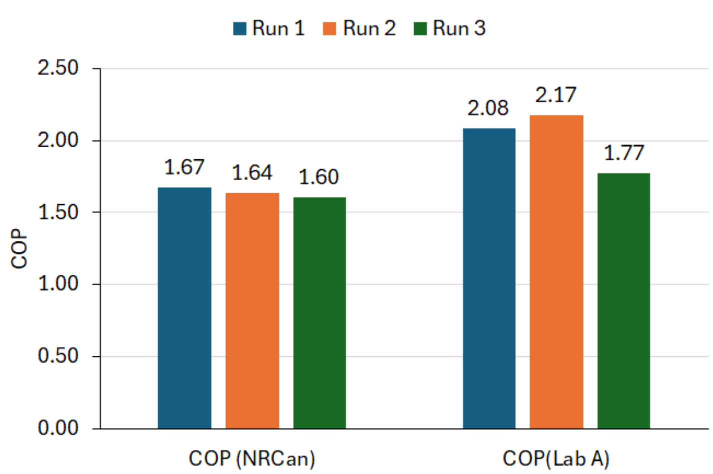


Figure 53. Unit C COP results for test interval HC (-8.3°C) repeated three times at NRCan and Lab A

3.2.1.3. Continental Heating Test Interval HL (-25°C)

Figure 54 shows the capacity, input power and airflow of Unit C for the test interval HL. Unit C's lowest catalogued operating temperature (LCT) was -25°C. However, Unit C did not operate at this temperature at NRCan's test facility. While the input power followed a similar ramp and peak, the heating capacity never climbed above 1.2 kW at NRCan while it reached a peak of 3 kW at Lab A. Unit C even exhibited negative capacity on startup indicating an issue with the compressor's ability to discharge high temperatures refrigerant.

Following repeated attempts to operate Unit C at the LCT at NRCan, it was decided to check the refrigerant charge to make sure none had been lost since the start of testing. With 3.00 lbs having been added at the start of testing, NRCan recovered 2 lbs 12 oz. Although this constitutes a slight loss of refrigerant, it is suspected that the remaining 4 ounces were still present in the unit in vapour form. NRCan then added a factory charge, or 3 lbs 12 oz, with the same resulting outdoor unit cut-off when approaching the LCT as shown in Figure 55.

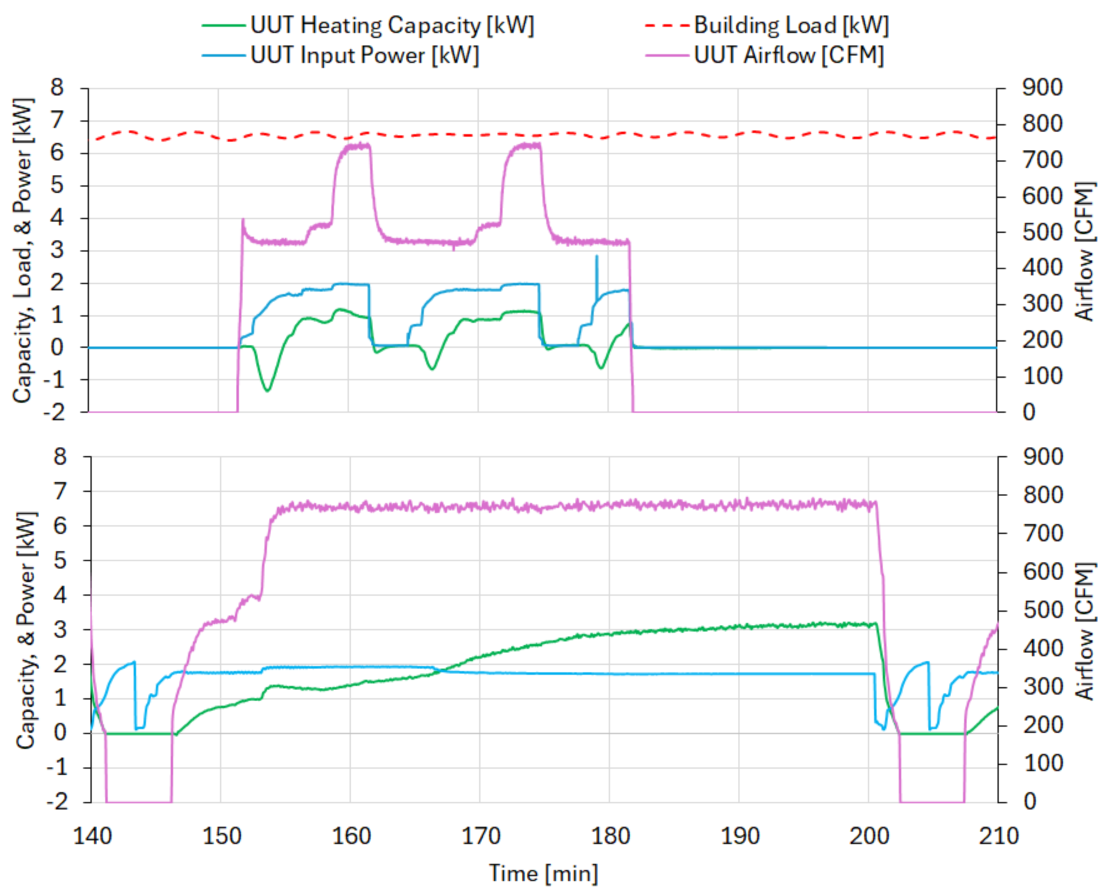


Figure 54. Unit C capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HL (-25.0°C)

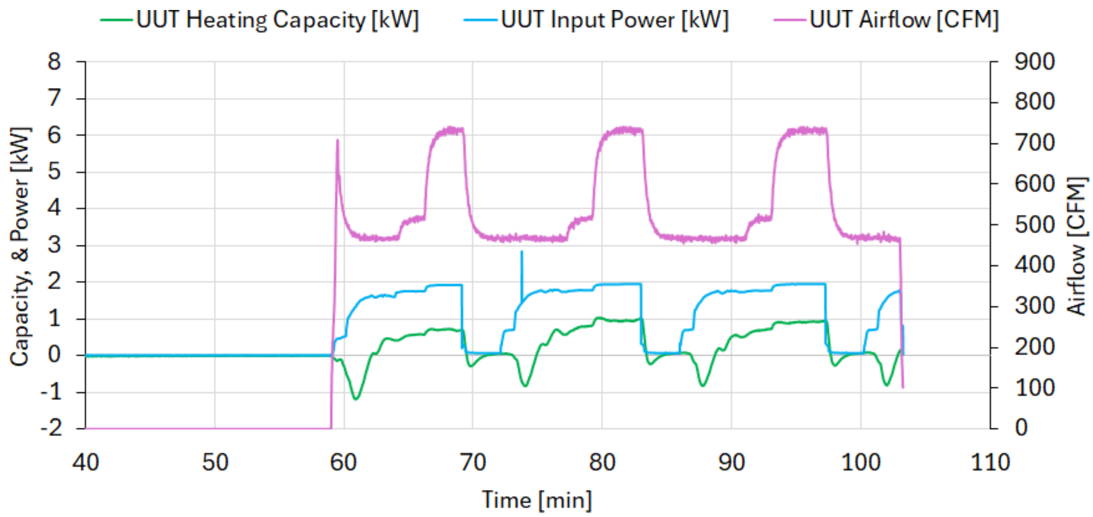


Figure 55. Unit C capacity, building load, and input power results at NRCan for heating continental test interval HL (-25.0°C) using a refrigerant factory charge of 3 lbs 12 oz

3.2.2. Cooling Rating Test Summary

A complete series of SPE-07 cooling learning and rating tests (excluding the dry cooling test CA (45°C)) was conducted at Lab A and NRCan's facilities on Unit C to evaluate reproducibility. The results of capacity, input power, cooling COP, and $SCOP_C$ are presented below with Lab A's test results serving as the reference against which NRCan's results are compared. Several tests are analyzed in greater detail to support discussion of differences in equipment behaviour and to investigate the limits of reproducibility. The green-shaded areas in figures represent the convergence period in instances where more data is presented. Figures presenting results of indoor and outdoor dry-bulb temperature have a blue dashed line for the target indoor dry-bulb temperature and a red dashed line for the target outdoor dry-bulb temperature. For more detailed test results, see Section 7.1.2.

The cooling learning series tests determined that tests CE through CC in dry conditions were to be load-following tests, whereas test CB in dry condition was a full-load test. The test load conditions were consistent between facilities in dry conditions. However, there were considerable differences between facilities in humid test conditions as presented in Table 11. Unit C was only able to maintain a load at the desired indoor dry-bulb temperature at test interval CE humid. It was determined that a significant thermostat temperature sensor drift had occurred, and no replacement thermostat could be provided in time for testing. No onboard temperature offset was available, which could have been used to correct this drift.

Table 11. Unit C thermostat setpoints in cooling mode for NRCan and Lab A as determined by the cooling learning series tests

Test	Outdoor dry-bulb temperature, °F (°C)	Indoor dry-bulb control target, °F (°C)	TSTAT, °F (°C) (NRCan)	TSTAT, °F (°C) (Lab A)
CE_D	77 (25)	79 (26.1)	78 (25.6)	80 (26.7)
CD_D	86 (30)	79 (26.1)	76 (24.4)	80 (26.7)
CC_D	95 (35)	79 (26.1)	76 (24.4)	80 (26.7)
CB_D	104 (40)	Min	Min (67 (19.4))	Min (67 (19.4))
CE_H	77 (25)	74 (23.3)	67 (19.4)	74 (23.3)
CD_H	86 (30)	74 (23.3)	Min (67 (19.4))	72 (22.2)
CC_H	95 (35)	74 (23.3)	Min (67 (19.4))	72 (22.2)
CB_H	104 (40)	Min	Min (67 (19.4))	Min (67 (19.4))

Following the cooling learning series tests, the cooling rating tests were completed in order of lowest to highest load.

The thermostat temperature sensor drift previously noted render the humid cooling test results, and consequently the $SCOP_C$, invalid. However, there are test results worth exploring for the purposes of evaluating and improving reproducibility.

Total cooling capacity results, which includes latent and sensible cooling capacities, are presented in Figure 56 and Figure 57 for dry and humid conditions respectively.

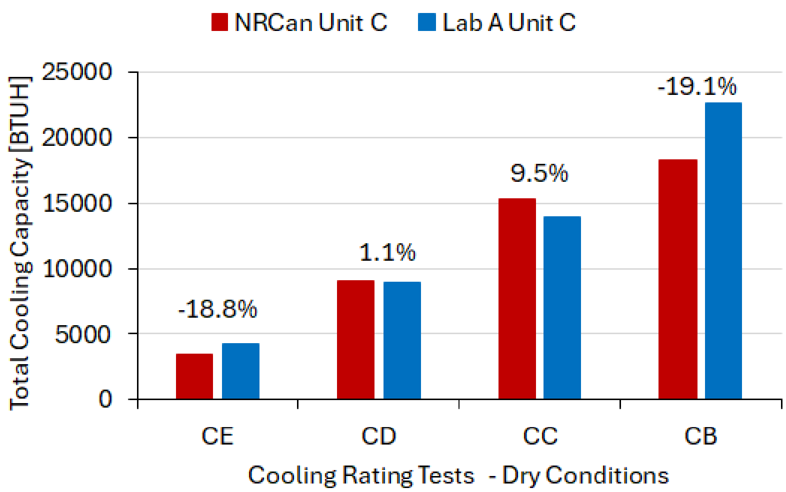


Figure 56. Total cooling capacity for Unit C, in dry conditions, NRCan and Lab A results

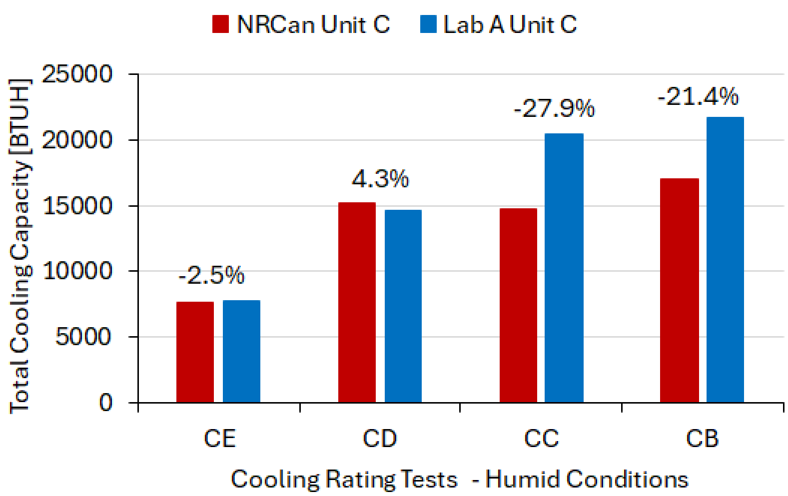


Figure 57. Total cooling capacity for Unit C, in humid conditions, NRCan and Lab A results

UUT integrated power consumption is presented in Figure 58 and Figure 59 for dry and humid conditions respectively.

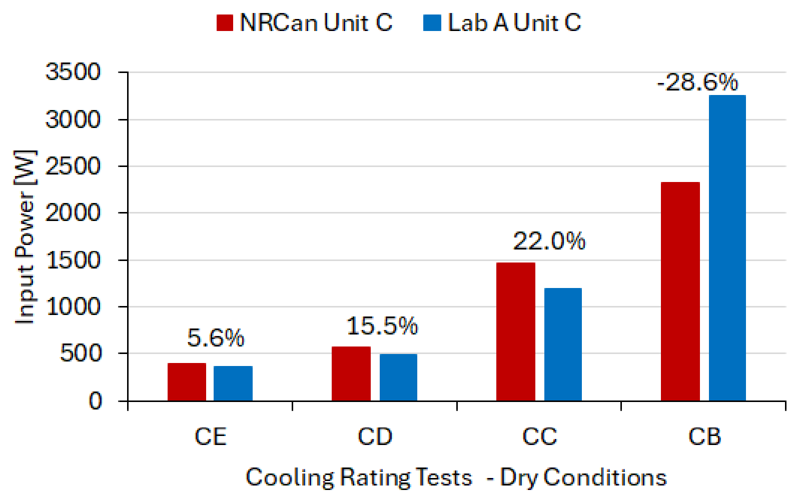


Figure 58. Cooling input power for Unit C, in dry conditions, NRCan and Lab A results

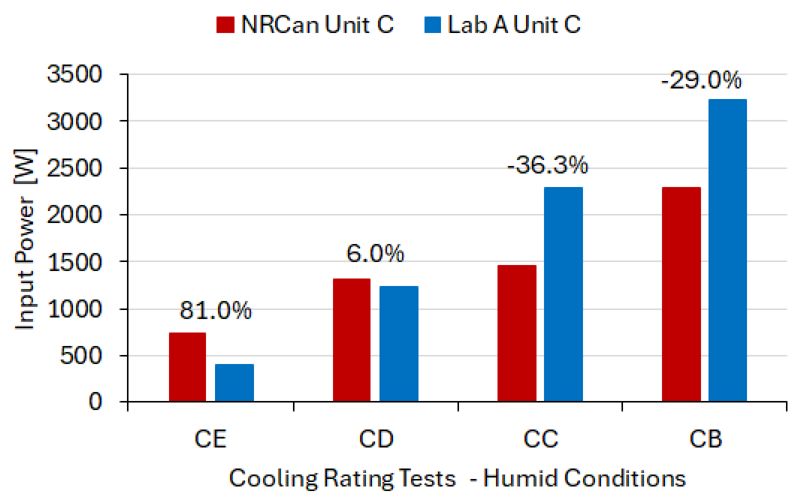


Figure 59. Cooling input power for Unit C, in humid conditions, NRCan and Lab A results

The resulting rating test COPs are presented in Figure 60 and Figure 61 for dry and humid climate conditions respectively.

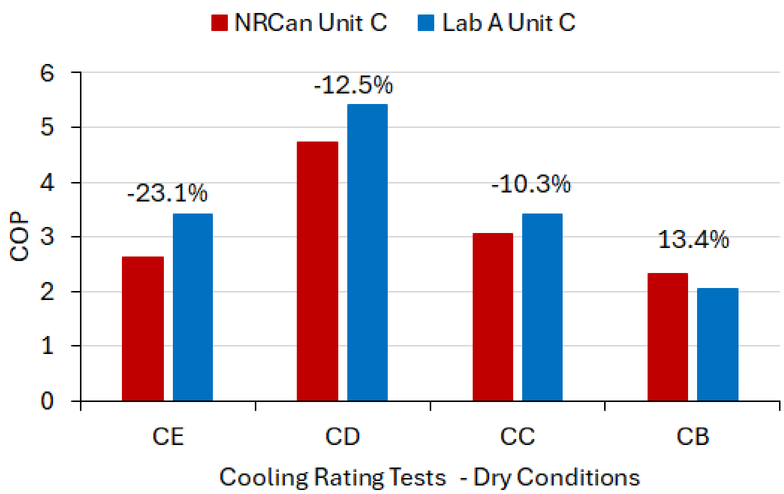


Figure 60. Cooling COP for Unit C, in dry conditions, NRCan and Lab A results

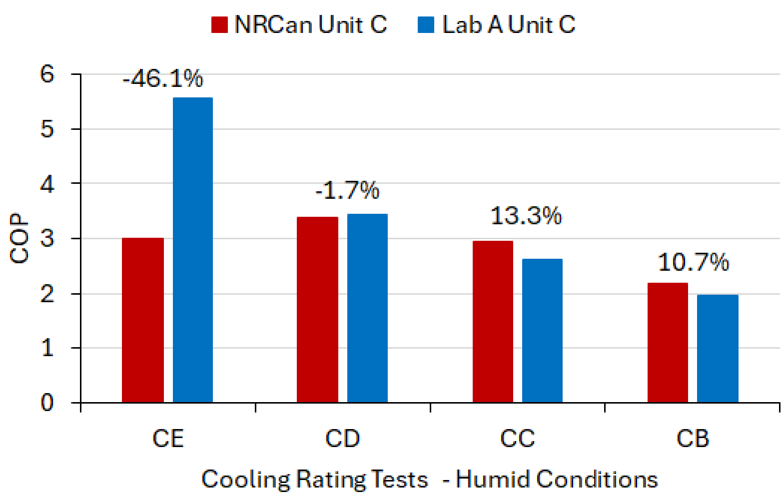


Figure 61. Cooling COP for Unit C, in humid conditions, NRCan and Lab A results

Finally, the $SCOP_C$ for the cooling bins are presented in Figure 62 and Table 12. The mean difference across all climate zones between Lab A and NRCan SCOPc was 15.8%, with a standard deviation of 5.1%. This demonstrated "poor" reproducibility for Unit C in cooling mode tests. Rating tests CE dry, CE humid, CC dry, and CB humid will be discussed in greater detail below.

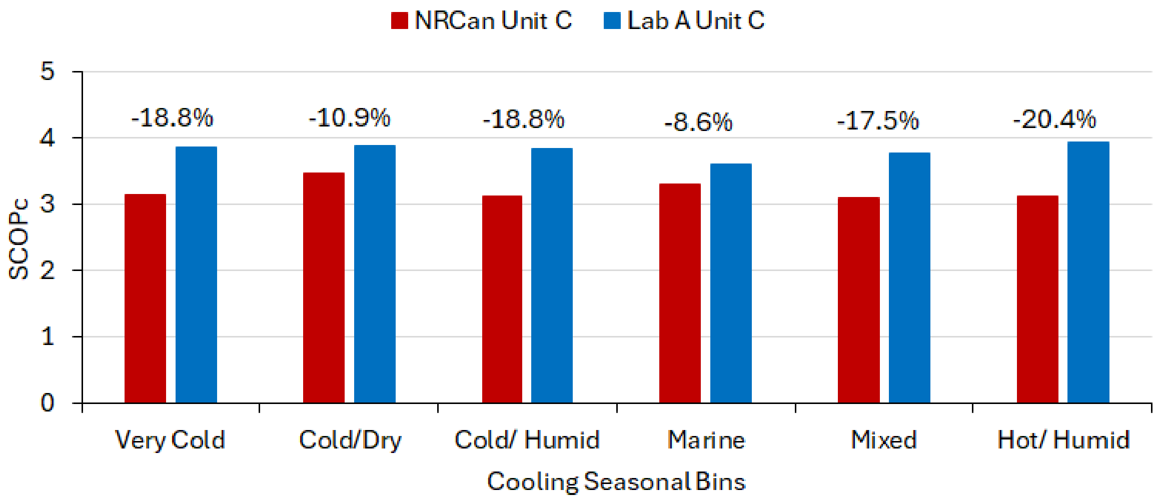


Figure 62. $SCOP_C$ for Unit C, NRCan and Lab A results

Table 12. $SCOP_C$ for Unit C, NRCan and Lab A results

	Very Cold	**Cold/Dry**	**Cold/ Humid**	**Marine**	**Mixed**	**Hot/ Humid**
Lab A	3.86	3.88	3.84	3.61	3.76	3.93
NRCan	3.13	3.46	3.12	3.30	3.10	3.13
ΔSCOP [%]	18.84	10.87	18.75	8.59	17.51	20.44

3.2.1. Individual Cooling Test Results

3.2.1.1. Dry Cooling Test Interval CE (25°C)

Figure 63 shows the behaviour of rating test interval CE dry (25°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was part-load and ran for 4 hours as required by SPE-07. Cyclic behaviour was observed at both facilities, with relatively similar cycle times of 20 minutes. However, the average cooling capacity during an on cycle, as shown in Figure 64, was 1.40 kW at Lab A, and 1.14 kW at NRCan, while the input power was 0.43 kW at Lab A, and 0.44 kW at NRCan. The difference in capacity observed in Figure 64 could be the result of Lab A using a ¼-inch RTD in an aspirated psychrometer to calculate capacity, whereas NRCan used a grid of thermocouple sensors in the air stream. This difference in instrumentation could lead to a measurement lag at Lab A resulting in lower amplitude "peaks" and "troughs". The result is 18.8% higher integrated cooling capacity at Lab A. The integrated COP was therefore 23.1% lower at NRCan.

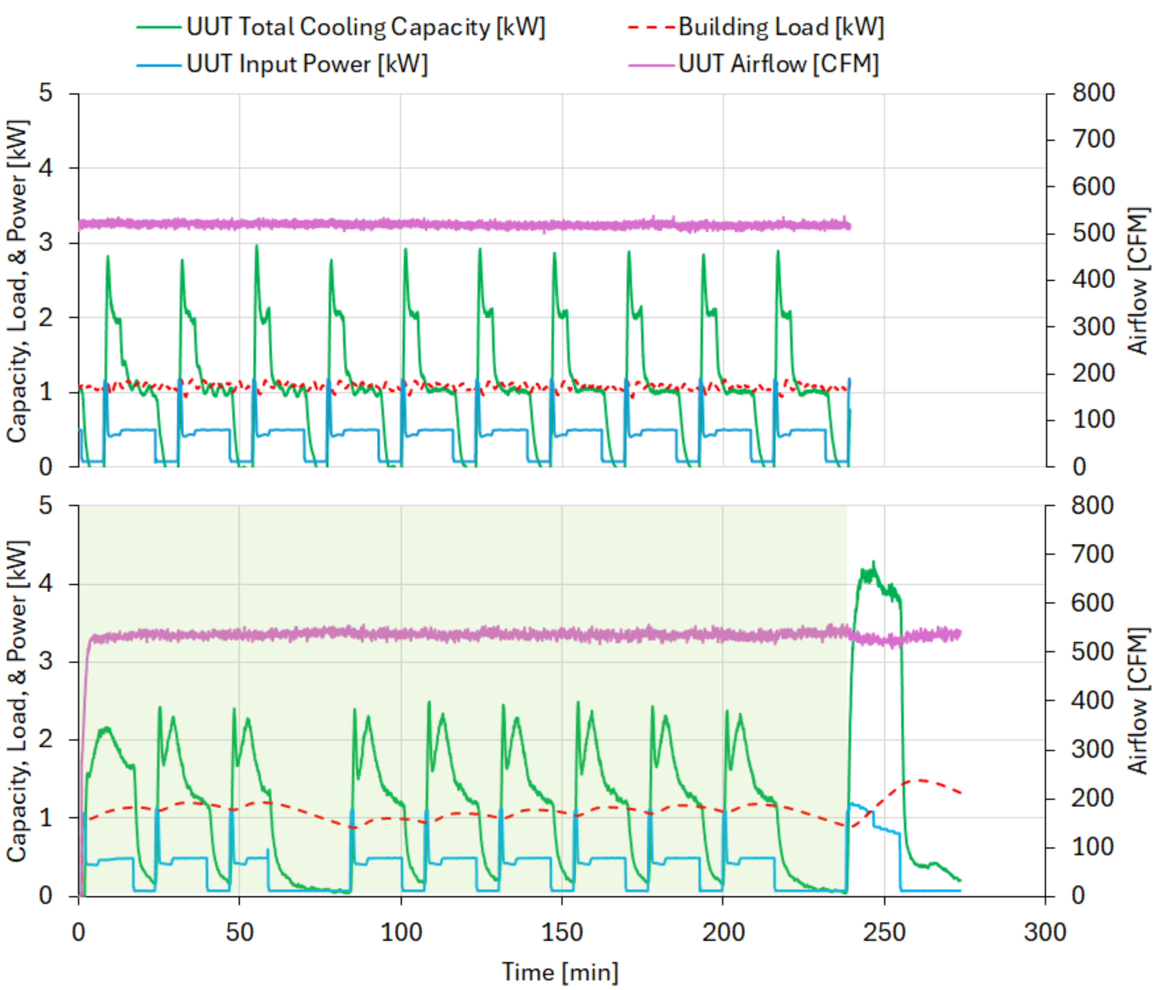


Figure 63. Unit C capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for cooling dry climate test interval CE (25°C)

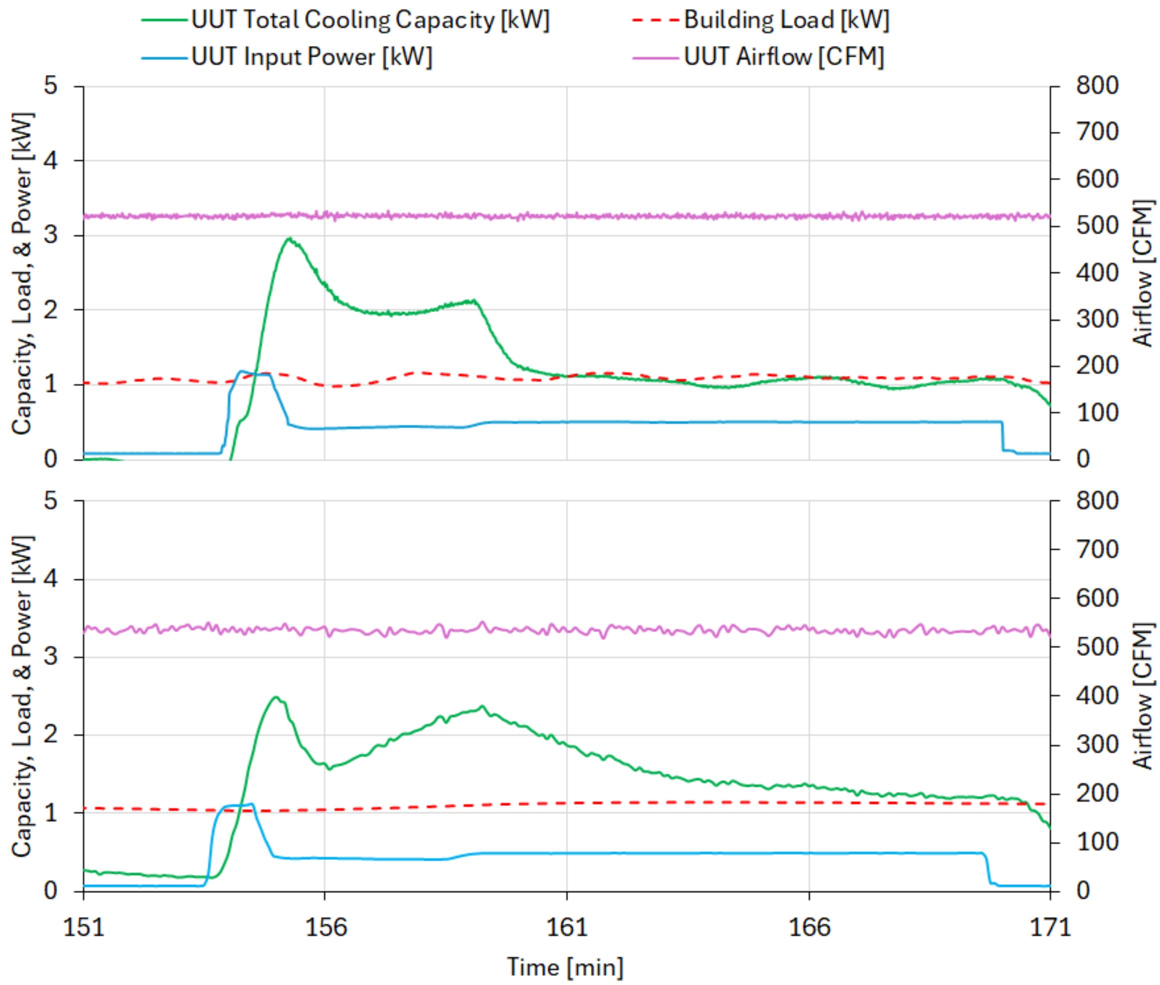


Figure 64. Unit C dry climate test interval CE on cycle capacity, input power, load, & airflow

Figure 65 shows the indoor and outdoor dry-bulb temperatures during the CE dry test interval for Unit C. In this test, the UUT was not overcoming a significant building load, and it was able to maintain the setpoint at both facilities.

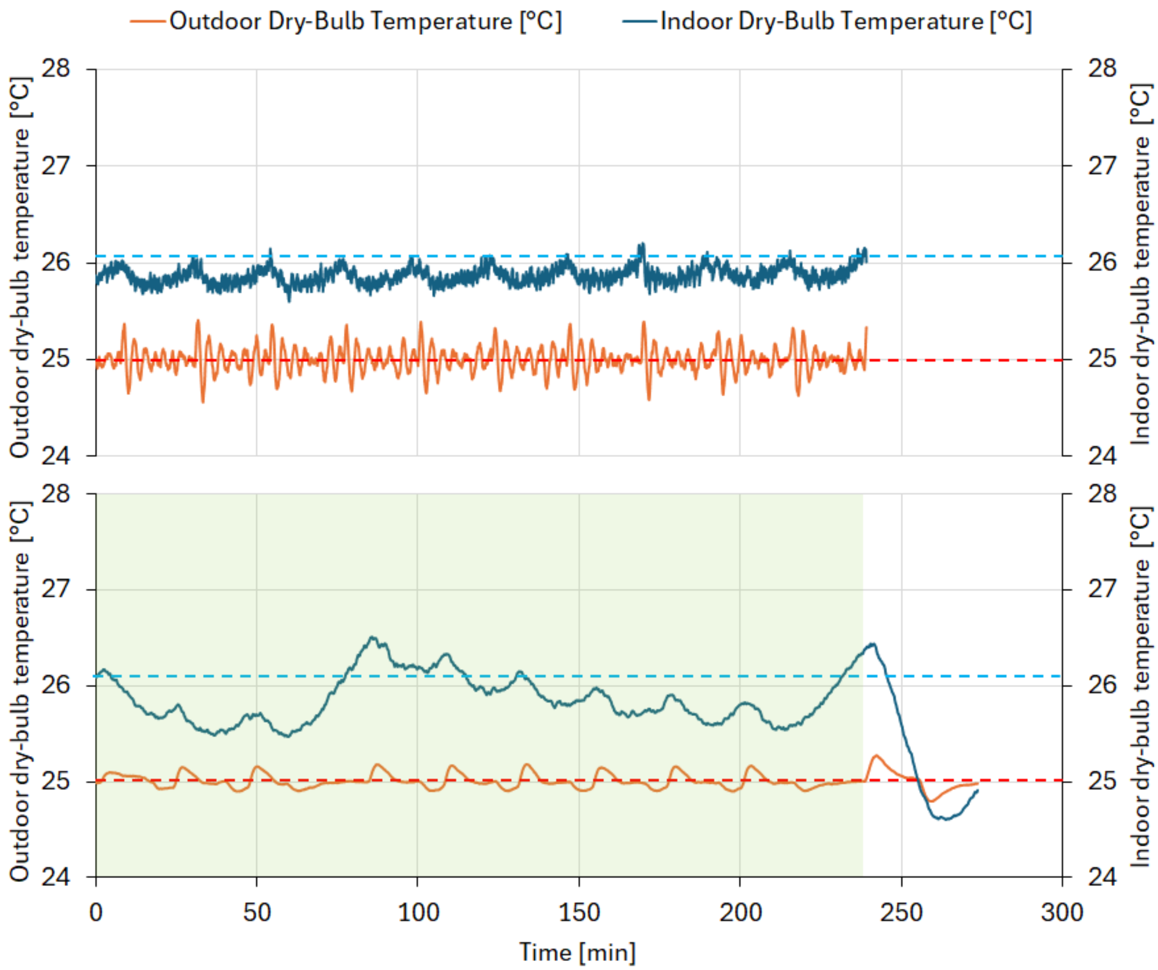


Figure 65. Unit C indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for cooling dry climate test interval CE (25°C)

3.2.1.2. Humid Cooling Test Interval CE (25°C)

Figure 66 shows the behaviour of rating test interval CE humid (25°C) for UUT capacity, building load and input power at NRCan and Lab A. This test interval was part-load and ran for 4 hours as required by SPE-07. NRCan observed regular cycling while Lab A observed a mix of steady-state and cyclic behaviour. This led to larger building load swings at Lab A. The airflow was significantly higher at NRCan with an average of 720 CFM versus 520 CFM at Lab A.

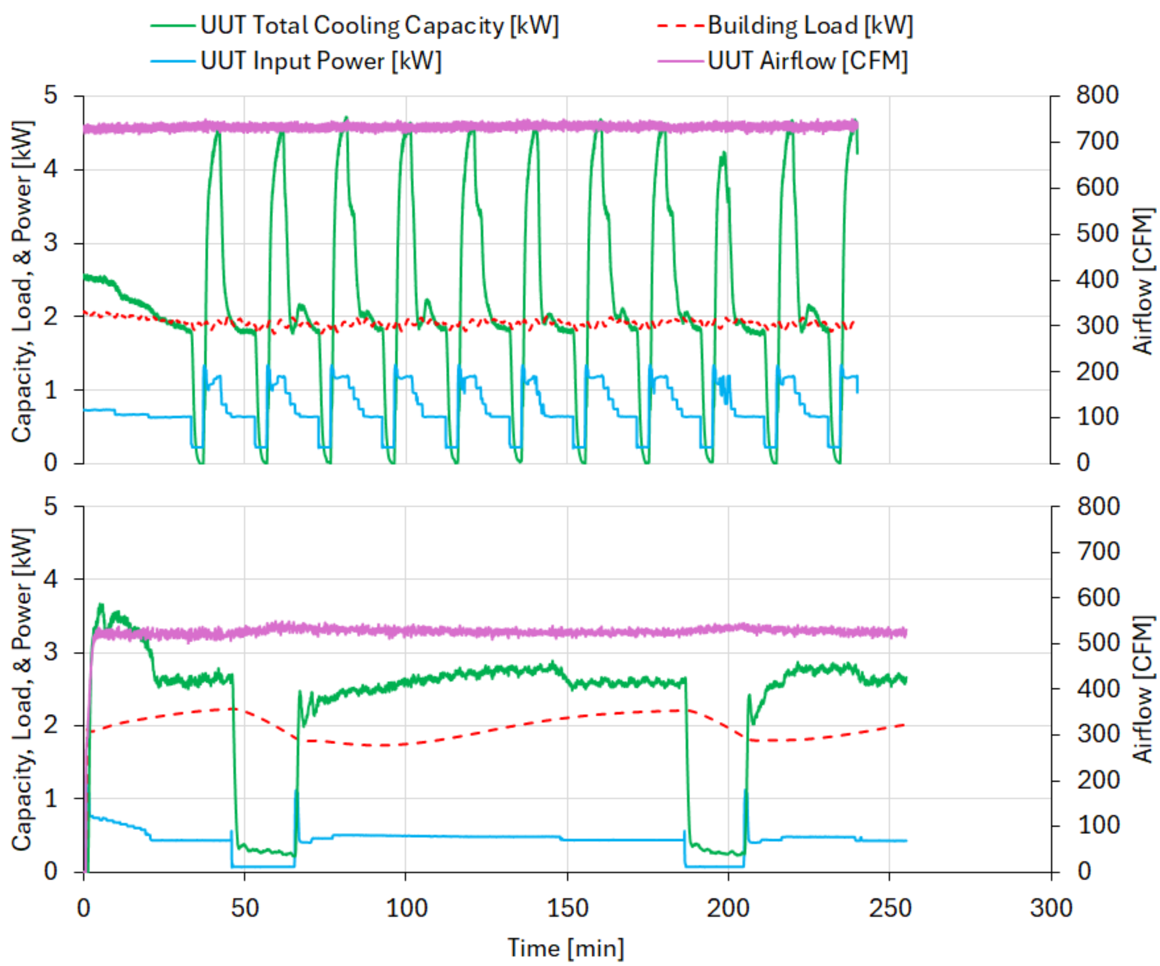


Figure 66. Unit C capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CE (25°C)

Figure 67 shows the indoor and outdoor dry-bulb temperatures during the CE humid test interval for Unit C. The UUT was able to maintain the thermostat setpoint at both facilities but did so in different ways as reflected in the indoor temperature data. NRCan observed the UUT cycling at short intervals, which led to a relatively steady indoor temperature. Lab A on the other hand, observed longer cycles, which led to more pronounced swings in indoor temperature as the capacity stayed above the building load line for longer. The difference in behaviour may be due to a different thermostat response, which, as was discussed earlier, encountered a significant temperature sensor drift at NRCan and therefore, required different setpoints to meet the same target indoor temperature of 23.3°C.

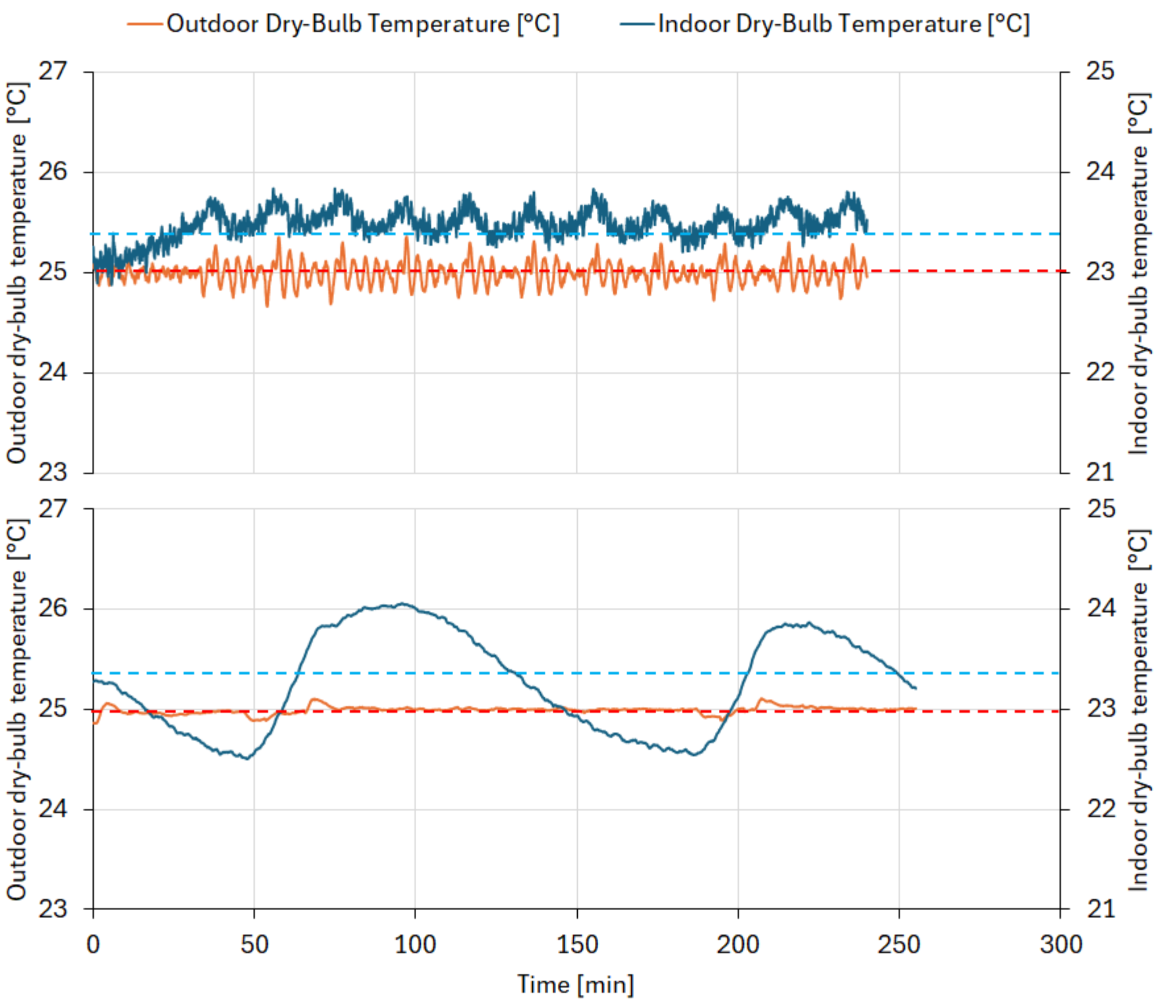


Figure 67. Unit C indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CE (25°C)

3.2.1.3. Dry Cooling Test Interval CC (35°C)

Figure 68 and Figure 69 show Unit C's behaviour during test interval CC dry (35°C) at NRCan and Lab A, respectively. NRCan conducted a learning series test that resulted in a thermostat setpoint of 76°F (24.4°C) to maintain the target indoor dry-bulb temperature of 26.1°C. However, during the rating test, the UUT was not able to maintain the indoor temperature setpoint and was converted to a full load test after 90 minutes. The full load test interval converged on two 20-minute periods and is presented in Figure 68. In this full load test, Unit C maintained a total cooling capacity of approximately 4.5 kW with an input power of 1.5 kW. Lab A, on the other hand, successfully conducted a part-load test with an integrated cooling capacity of 4 kW with an input power of 1.2 kW. This resulted in a 10.3% lower COP at NRCan.

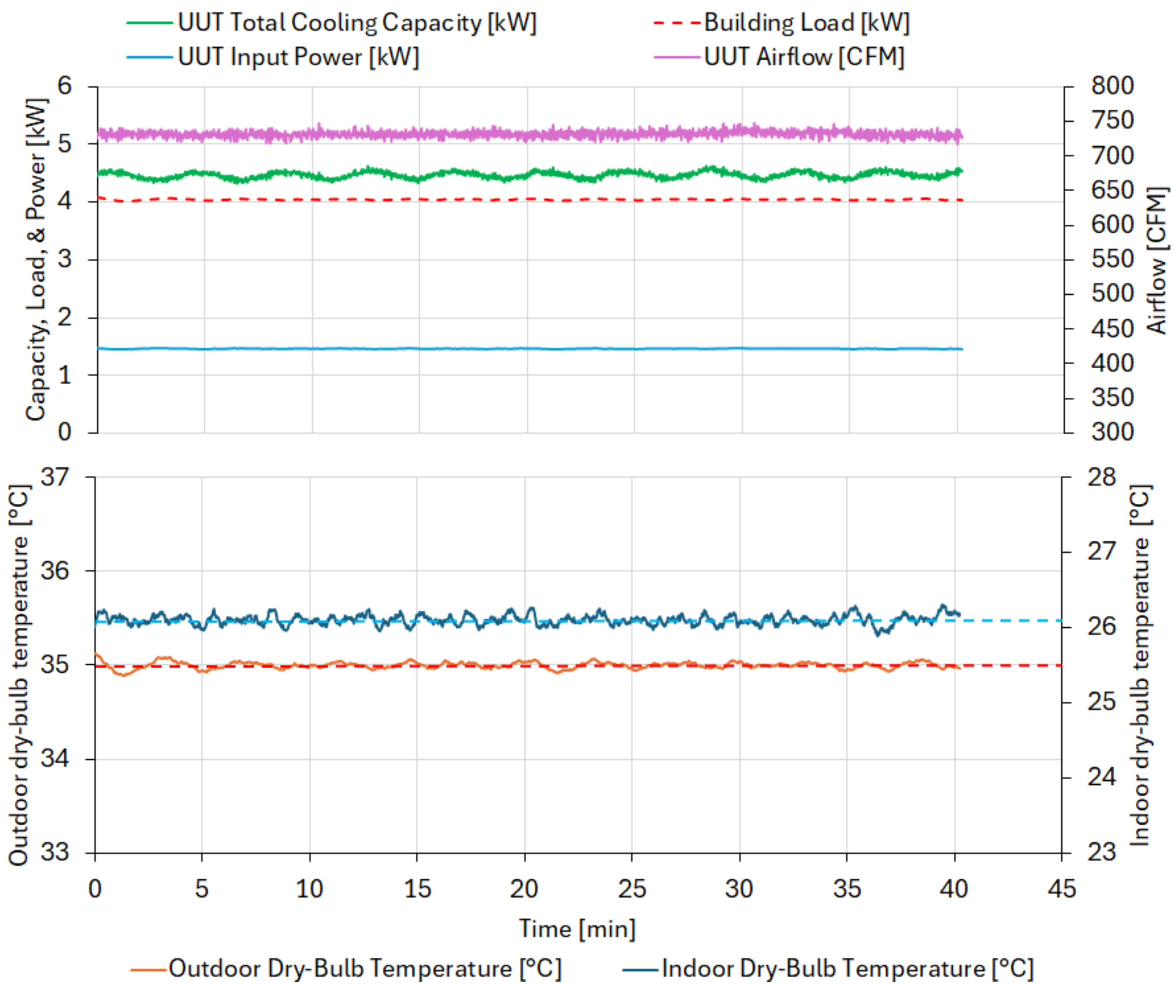


Figure 68. Unit C test results at NRCan for cooling dry climate test interval CC (35°C)[5]

[5] **Note!** Figure 68 and Figure 69 are split by test facility because they did not converge on the same criterion and therefore have different test durations.

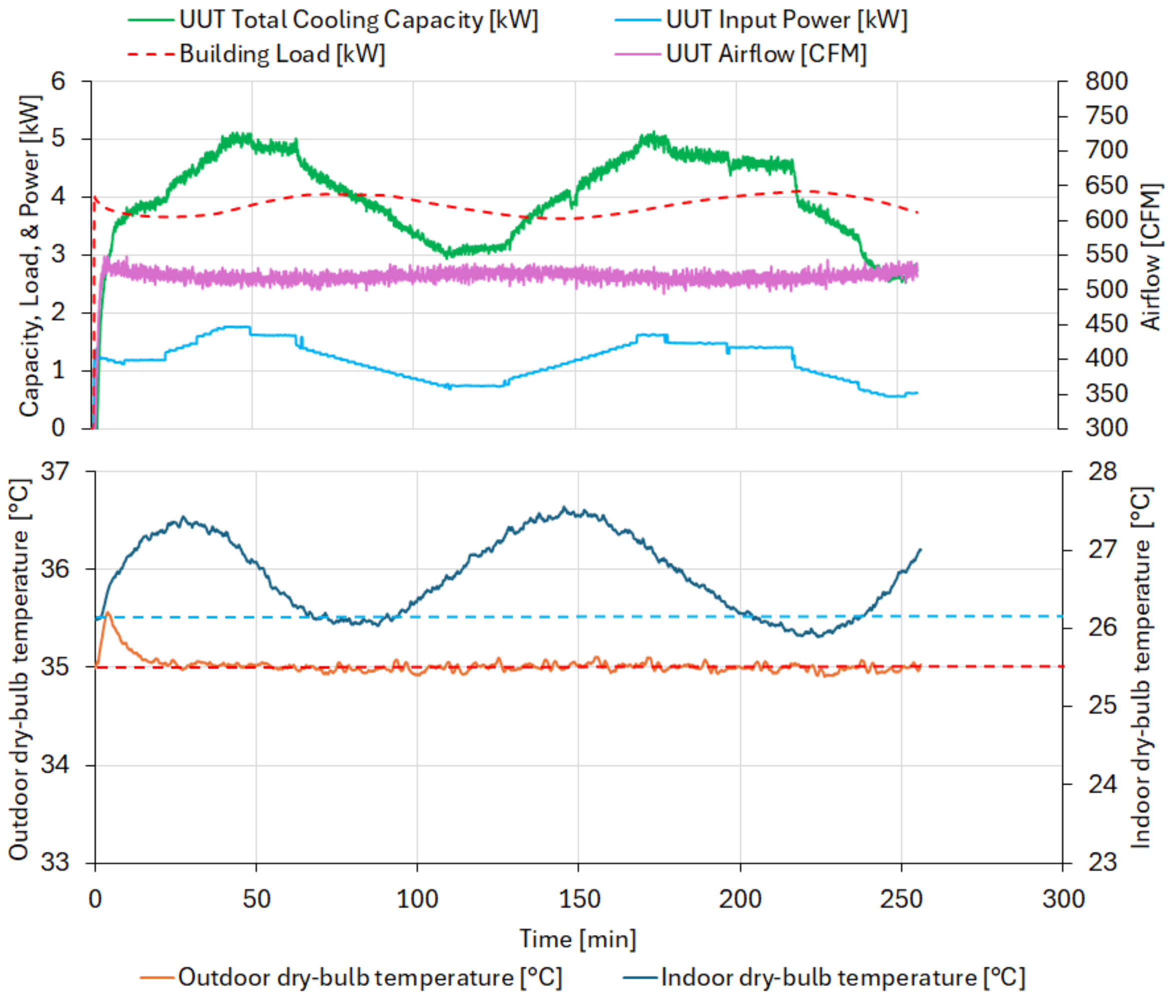


Figure 69. Unit C test results at Lab A for cooling dry climate test interval CC (35°C)

Figure 70 and Figure 71 presents a breakdown of the sensible versus latent cooling capacity at NRCan and Lab A, respectively. Because the sensible virtual building load is calculated using the UUT's sensible capacity, we can see how NRCan was not able to maintain the indoor thermostat setpoint with a sensible cooling capacity slightly higher than the virtual building load with 4.2 kW at full capacity, whereas Lab A was able to achieve a sensible cooling capacity peaking at 4.7 kW, which was enough to cool the indoor chamber. The latent virtual building loads were similar at both facilities and the UUT achieved an integrated humidity ratio of 0.0086 at NRCan and 0.0083 at Lab A. This test demonstrates a case where the UUT was on the verge of its full-load capacity and resulted in two different test conditions at different facilities.

The issue may lie in the thermostat setpoint offset between facilities. Lab A used a setpoint of 80°F (26.7°C), whereas NRCan used a setpoint of 76°F (24.4°C) when first starting its rating test. This 4°F (2.2°C) difference in setpoint may have triggered a different capacity response by the thermostat. Given more than 90 minutes, the equipment may have adjusted its desired target capacity at NRCan and met its thermostat setpoint. However, the test was automated and was not able to run for more than 90 minutes in load-following mode given SPE-07's requirement of maintaining an indoor room temperature within 2 °F (1.1°C) of the target indoor room temperature. Test CB humid (40°C) below reinforces this hypothesis as the UUT was able to output higher cooling capacities at higher load conditions, meaning the limitations observed at NRCan in this test interval may have been thermostat-induced.

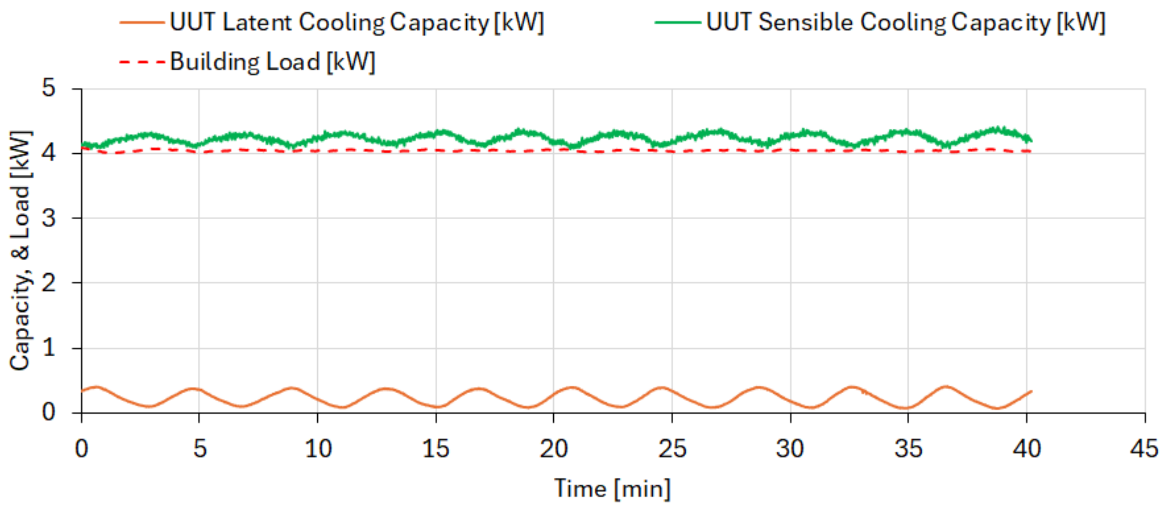


Figure 70. Unit C latent and sensible capacity versus building load during cooling dry climate test interval CC (35°C) at NRCan

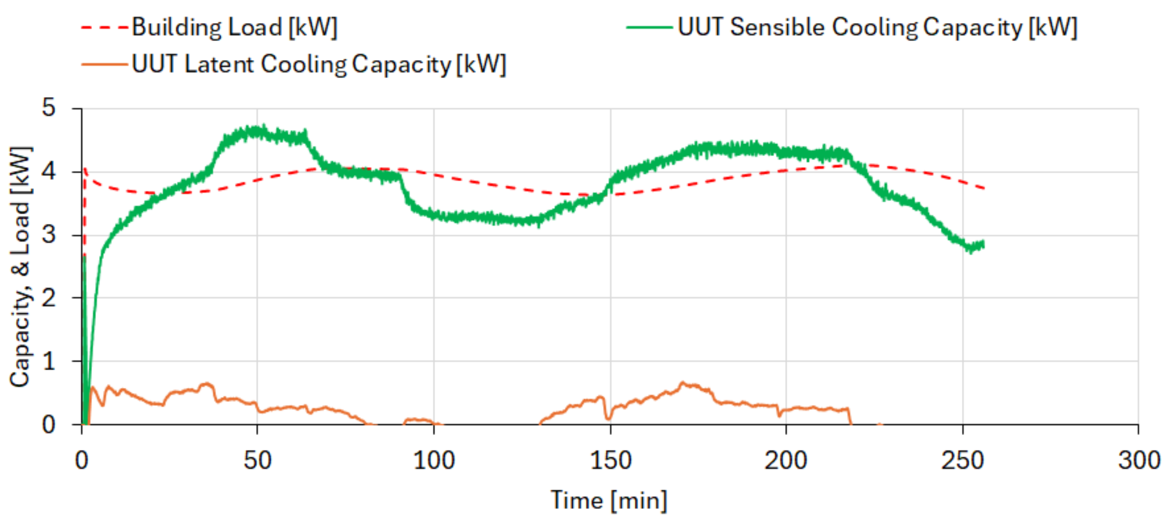


Figure 71. Unit C latent and sensible capacity versus building load during cooling dry climate test interval CC (35°C) at Lab A

3.2.1.4. Humid Cooling Test Interval CB (40°C)

Figure 72 shows the behaviour of rating test interval CB humid (40°C) for UUT capacity, building load, input power, and airflow at NRCan and Lab A. This test was a full-load interval at both facilities and achieved convergence in 40 minutes with two 20-minute periods of steady-state operation. The total cooling capacity was 1.35 kW higher at Lab A. The input power was also higher at Lab A by approximately 1 kW. This resulted in a 10.7% higher COP at NRCan. The difference in UUT response may be due to the thermostat temperature offset observed earlier in testing between facilities. A higher setpoint difference versus sensed temperature may have induced a higher load condition response by the equipment. Unfortunately, no thermostat temperature offset function existed on Unit C's thermostat, and the setting used was the lowest available on the device.

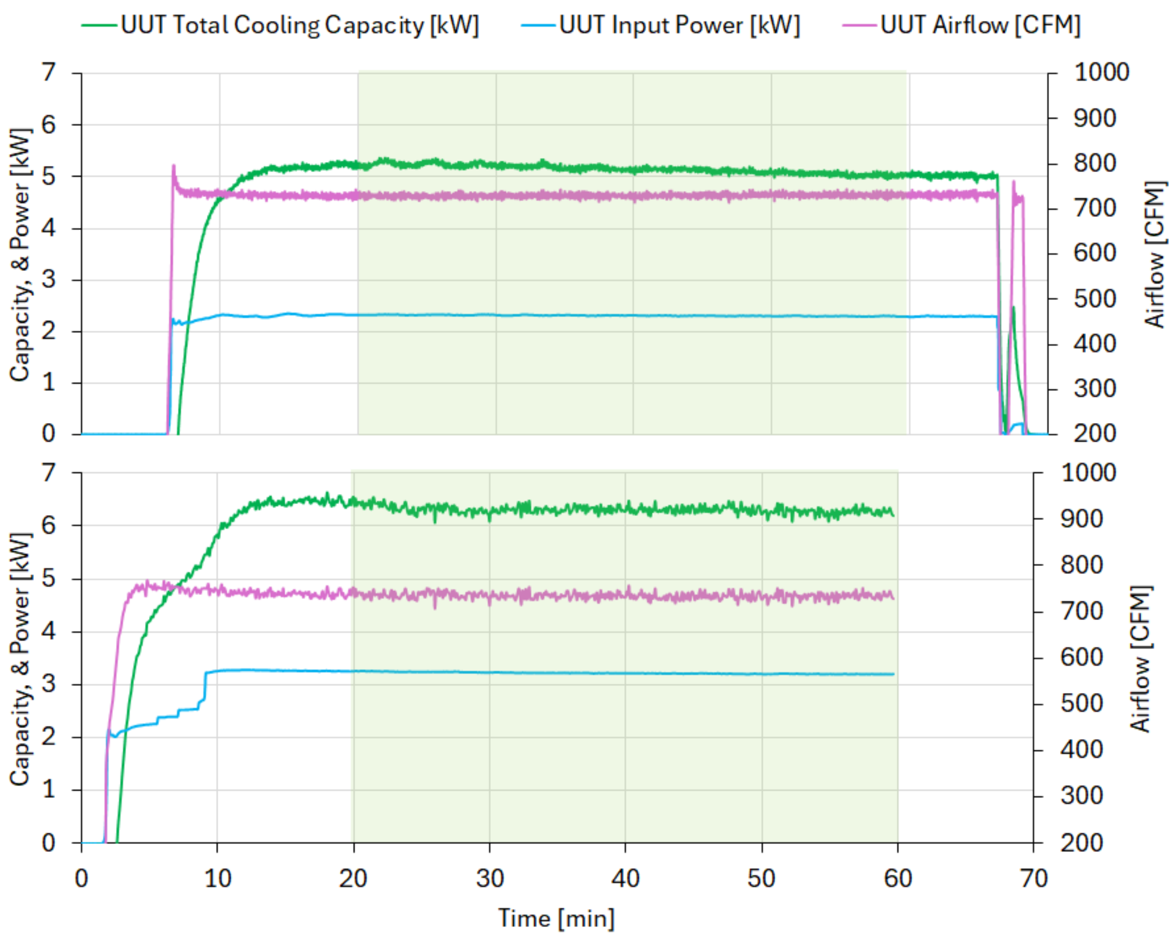


Figure 72. Unit C capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CB (40°C)

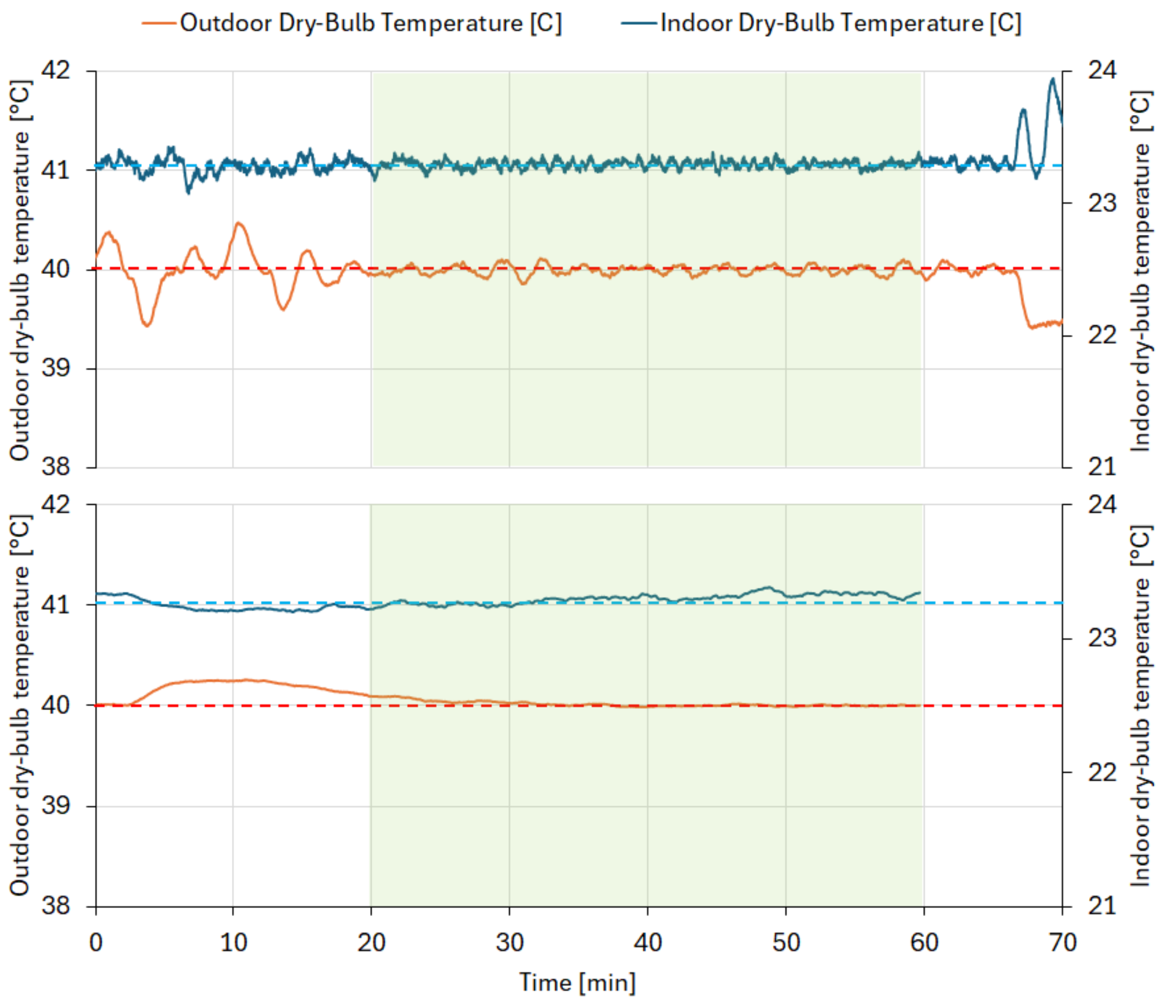


Figure 73. Unit C indoor and outdoor temperature variations at NRCan (top) and Lab A (bottom) for cooling humid climate test interval CB (40°C)

4.0. Discussion

4.1. Unit A

With a mean difference of 8.8% in heating SCOP and 8.6% in cooling SCOP, Unit A achieved “fair” reproducibility. In heating mode, Unit A’s capacity was fairly reproduced in both labs, with some lower-load intervals being outliers. However, these same lower-load intervals such as HF (12.2°C) and HE (8.3°C) were also characterized by higher power consumption which offset the higher capacities and achieved good reproducibility results in terms of COP. On the other hand, the higher-load heating tests achieved good reproducibility in terms of heating capacity, but consumed more power while doing so, resulting in lower COPs at NRCan. In fact, NRCan observed higher input power for all heating rating tests. The disaggregation of input power between indoor and outdoor units did not reveal where this difference may have originated, with a proportional split between both units in each test interval.

Test intervals where defrost operations occurred, such as HC (-8.3°C) resulted in issues with input power. As was presented in section 3.1.2.3, Unit A’s indoor blower stayed on during defrost operations; something which is not common in today's CC-ASHPs. Lab A did not capture the cooling load being introduced by virtue of the indoor unit blower staying on during defrost operations and thus did not properly account for this effect in its determination of the virtual building load, which would have affected the actual load experienced by the UUT and its associated response. NRCan, on the other hand, observed the capacity dropping to -3.5 kW, adding a significant load for the UUT to recover. The UUT response therefore diverged between facilities at this point, leading to poor reproducibility for this test condition. Even so, equipment behaviour seemed to have been consistent between facilities during these test intervals, giving credence to the assumption that the COPs would have been closer had this error not been present in Lab A’s analysis.

In cooling mode, Unit A’s capacity was well reproduced between Lab A and NRCan, with test interval CE dry (25°C) being an outlier. However, as was observed in heating mode tests, the input power was higher at NRCan than Lab A, resulting in overall lower COPs at NRCan. Test interval CE humid had the largest gap in cooling COP between facilities. A deep dive into the behaviour of the UUT during this interval revealed that longer and more intense cycles were observed at NRCan. This behaviour resulted in more energy being used by the UUT to maintain the same target indoor dry-bulb temperature. This trend was again observed during the CC dry (35°C) test interval and resulted in a 10.2% lower COP at NRCan.

No issues were observed in the usage of the thermostat environment emulator during Unit A’s testing at NRCan. However, some potential issues were raised with respect to the UUT thermostat placed inside the TEE. NRCan received the thermostat following Lab A’s testing with the fan airflow mode set to “High”. This was not consistent with the expected mode of operation for SPE-07 testing, which typically requires a “default” thermostat setup with some exceptions. Following consultation with Lab A, NRCan set the thermostat to “Auto” indoor fan airflow mode and conducted the cooling learning and rating tests using this mode. Initial heating mode tests, however, revealed indoor unit behaviour that was not consistent with what had been observed at Lab A, and a change back to “High” indoor airflow mode was made. This introduced uncertainty surrounding the thermostat’s setup between facilities that was never resolved. In some rating test intervals such as HC, NRCan observed lower airflows which could be explained by a different thermostat operating mode. This uncertainty surrounding the thermostat setup could be resolved by creating a list, or notes, of settings in the SPE-07 reporting template that could be used to confirm the modes of operation, if settings other than default were used.

4.2. Unit C

Unit C testing resulted in a mean heating SCOP difference of 9.8% and a mean cooling SCOP difference of 15.8% between Lab A and NRCan, resulting in relatively poor reproducibility. Unit C's heating capacity was consistent between facilities in load-following test intervals. The full-load test intervals, on the other hand, revealed a deeper disconnect in equipment performance between labs. Test HC (-8.3°C), presented in section 3.2.1.2, highlighted how the full-load capacities were different between test facilities. During repeatability testing conducted at Lab A, one of these heating tests conducted during interval HC, resulted in higher integrated capacity and input power. This result more closely aligned with what was observed at NRCan. This discrepancy was again present in test interval HB, where NRCan observed a higher capacity response, resulting in a lower COP. One of the conclusions drawn from heating full-load testing was that the convergence criterion in this case should be based on three consecutive periods of 20-minute steady-state operation, rather than two. This extra time to achieve convergence would almost certainly have resulted in capturing a defrost operation on Unit C at NRCan and would have meant that convergence would probably have been based on regular defrost cycling or a full 6-hour test interval. This change would also allow the convergence criterion to be more aligned with convergence criterion 1 of SPE-07 for part-load, variable speed, operation.

Unit C's cooling capacity showed no consistent trend towards reproducibility with test intervals CE (25°C), CB (40°C) dry, and CC (35°C) with CB (40°C) humid having particularly significant deviations from Lab A results. Likewise, input power was not consistent between facilities and was not useful in revealing any trend that would help in assessing reproducibility. The thermostat was the largest source of error for the cooling rating tests with significant thermostat setpoint deviations determined in the cooling learning series tests observed at NRCan and shown in Table 11. With no thermostat temperature offset available onboard Unit C's thermostat and with no available replacement thermostat, the tests were conducted with the setpoints as determined in the learning series tests. The humid cooling tests were most impacted by this decision as they were all conducted at the lowest thermostat setpoint available of 67°F, regardless of test interval target indoor dry-bulb temperatures. The resulting cooling COPs are therefore not useful in determining the reproducibility of SPE-07.

It is important to note that Unit C's equipment response may have been impacted by how far apart the room temperature and thermostat setpoint were relative to each other. In practical terms, this means that the equipment could have responded differently had the setpoint been closer to the indoor room temperature in full-load test intervals. In this situation, solutions such as incorporating an offset function on the thermostat environment emulator, rather than the UUT thermostat, could be useful in ensuring a consistent response between facilities. This would remove any issues related to thermostat onboard temperature offsets, or lack thereof.

4.3. Factors Affecting Reproducibility

4.3.1. Instrumentation Lag

The virtual building load being calculated using the UUT capacity means it is dependent on the instrumentation to obtain the latter. There are three main measurements to be taken to calculate the capacity: the indoor unit return and supply temperatures, and the pressures through the airflow measuring station. SPE-07 allows for the indoor unit return and supply temperatures to be taken by either an RTD in an aspirated psychrometer, or a grid of temperature sensors (usually thermocouples) in the air stream, as stated in clauses B.1.5.1.2. and B.1.5.1.3. In principle, these two measurement methods as well as the instrumentation used, will result in some differences. Figure 74 demonstrates the difference between a ¼-inch Class A RTD measurement of dry-bulb temperature on the indoor unit supply-side, and a grid of calibrated thermocouples in the plenum used at NRCan. The delayed response, or "lag", present in the RTD measurement in the aspirated psychrometer would change the RAT as calculated by the virtual load emulator. In turn, this could trigger the UUT to respond, and therefore perform, differently in labs that use these types of sensors. The resulting difference is most significant in a load-based test approach where large fluctuations in return / supply temperatures are expected (as opposed to steady-state tests with relatively constant return / supply temperatures once steady-state conditions are reached). The magnitude of the resulting performance difference is not known and could be investigated in future work.

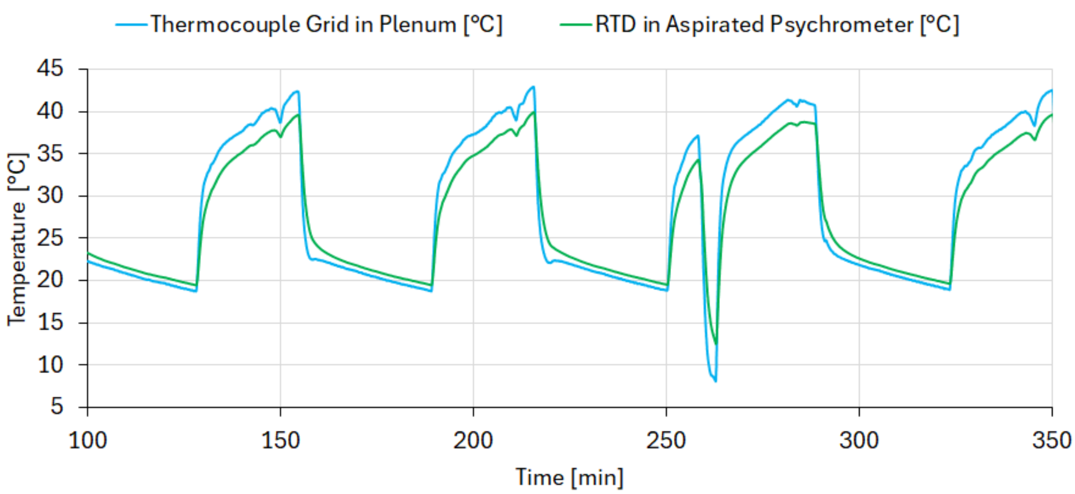


Figure 74. Difference in supply air measurement between an RTD in an aspirated psychrometer and a grid of thermocouples in the air stream at NRCan

4.3.2. Indoor Chamber Response Lag

As mentioned in section 4.3.1, the virtual building load, and the RAT by extension, is dependent on measurements that may have intrinsic lag. Another source of lag is the indoor chamber conditioning equipment response speed. As the virtual building load emulator calculates the next RAT and sends the signal to the indoor room conditioning equipment to respond accordingly, the UUT may momentarily experience a different return temperature than intended, thereby changing the load it is attempting to overcome. This lag from sensor feedback to conditioning equipment response can be in the order of minutes and can significantly alter the reproducibility of results as discovered by Waseda University. [12] Figure 75 presents the difference between an indoor chamber with virtually no delay between UUT return air temperature and RAT, and a chamber with approximately 3-5 minutes of delay. Although both test facilities maintained the test operating and condition tolerances, the potential delay in UUT response may lead to differences in overall test performance.

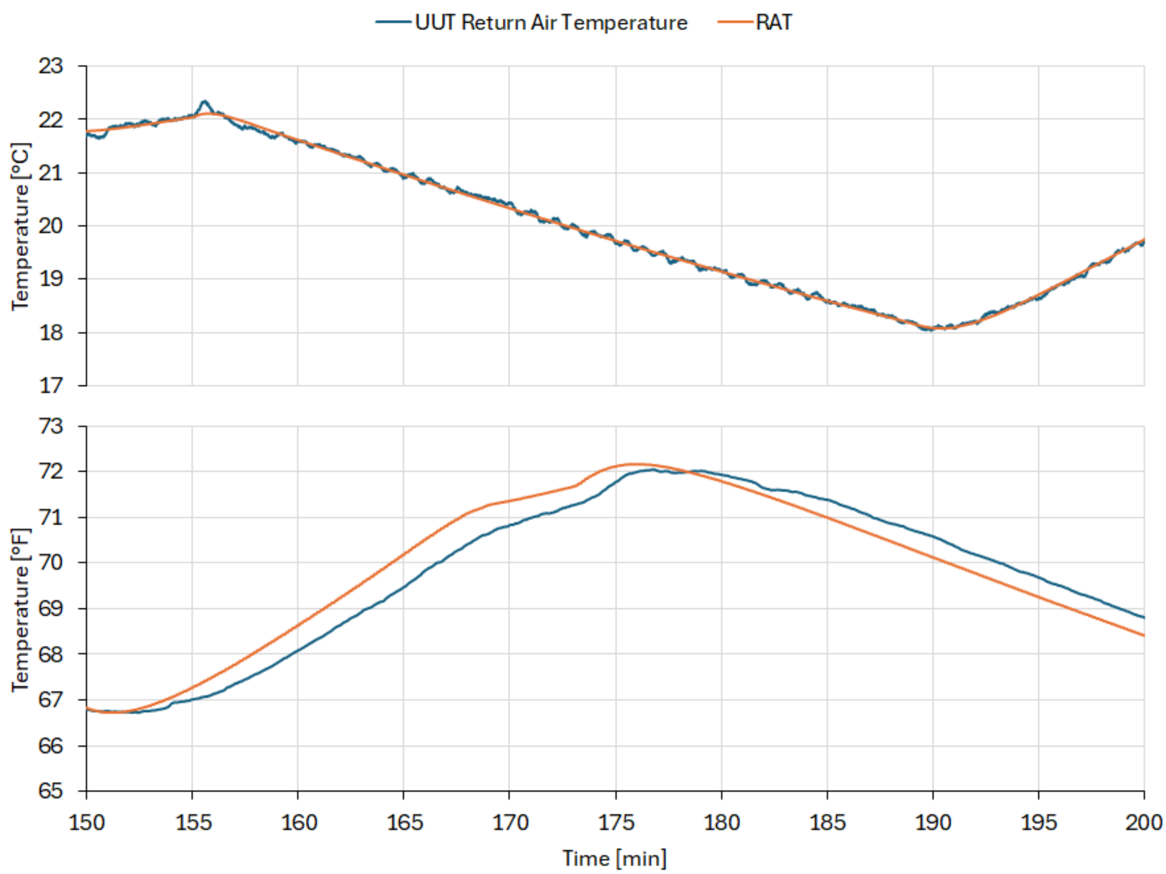


Figure 75. UUT return air temperature versus RAT at NRCan (top) and Lab A (bottom)

5.0. Conclusions

This report examined the reproducibility of the CSA SPE-07:23 procedure using two centrally ducted cold-climate air-source heat pumps evaluated at two independent testing facilities. Both heat pumps were 2-ton variable speed systems with communicating thermostats that had previously been field and lab tested. The objective was to produce heating and cooling seasonal COPs from a complete series of ratings tests using the same building load line and equipment installation from previously conducted field testing.

Unit A achieved an overall reproducibility of "fair" with 8.6% mean difference in $SCOP_C$, and 8.8% mean difference in $SCOP_H$ between facilities. Factors that affected reproducibility included:

a) Instrumentation lag differences between testing facilities,
b) Potential indoor chamber response lag in Lab A,
c) Issues in properly measuring UUT capacity during defrost operations in Lab A,
d) Installation differences between facilities such as the orientation of the indoor unit,
e) Equipment and thermostat dynamic response differences, and,
f) An unverifiable thermostat setup in Lab A.

Unit A was consistent in reproducing individual test interval load-following and full-load behaviour, which reflects favourably on the procedure's ability to induce a consistent equipment response between testing facilities.

While determination of reproducibility of testing for Unit C was difficult to ascertain (for the reasons described below), the unit achieved "poor" overall reproducibility in cooling rating tests, and "fair" overall reproducibility in heating rating tests. Note that the larger mean difference between facilities in cooling may be overstated. Unit C testing experienced issues with thermostat temperature sensor drift that substantially impacted reproducibility between testing facilities. Other factors that impacted reproducibility included:

a) Instrumentation lag differences between testing facilities,
b) Potential indoor chamber response lag in Lab A,
c) Installation differences between facilities, and
d) Equipment and thermostat dynamic response.

Furthermore, Unit C experienced cut-offs at the lowest catalogued temperature heating test interval (-25°C), resulting in no useable data for this bin. For the reasons described above, it was difficult to draw conclusions on the reproducibility of SPE-07 from Unit C testing.

Reproducibility between laboratories for a load-based procedure is inherently more difficult to achieve than steady-state rating tests, and an informal target of ≤ ±5% is often used to determine whether a procedure is reproducible. However, observations made during testing of Unit A would suggest that although 5% may be achievable by aligning the two test facilities more closely in terms of equipment installation and instrumentation used for UUT capacity and room conditions, the 5% target may be overly stringent, and a higher percentage would seem warranted to be used as a goalpost for reproducibility.

5.1. Proposed Amendments to SPE-07 to Enhance Reproducibility

Proposed amendments to the procedure for enhanced reproducibility, and other non-reproducibility related items, were submitted to the CSA C700 technical committee in preparation for the release of the accredited standard as a result of observations made during this project. Some of the proposed changes have been discussed in this report and are summarized here.

a) Instrumentation for UUT capacity measurement

Due to the lag present in some temperature sensor types, it is recommended that the grid of calibrated thermocouples (and thermopile in cases where it is used for a measurement of temperature difference) in both the return and supply air streams be used to calculate UUT capacity. The grid of thermocouples in the indoor unit return air stream should also be used as the indoor room temperature which the room reconditioning apparatus will target to achieve the RAT. This will eliminate the lag present in some RTD sensors as well as the lag inherent to aspirated measurements.

b) Tune chamber to minimize RAT lag (FFC)

As discussed in section 4.3.2, the indoor conditioning equipment response can lag behind the virtual building load emulator, thereby creating the potential for a mismatch in UUT response between test facilities with differing amounts of lag. It is therefore recommended that indoor chamber "tuning" methods be specified by SPE-07 to minimize the lag to a degree that is deemed acceptable. This could include mitigation strategies such as increased room air mixing, changes to the room conditioning equipment PID controls, and changes to the specified instrumentation used for control system feedback. As previously stated, Waseda University has done extensive testing on this very issue and is proposing a solution based on a "feed-forward compensation" methodology. [12] Such a change in control strategy could be specified in SPE-07 as a requirement for testing if deemed appropriate.

c) TEE temperature offset for UUT thermostat

Issues with Unit C's thermostat onboard temperature sensor led to unusable data in the cooling humid rating tests at NRCan. Since this thermostat did not have software temperature offsets available, it is recommended that temperature offsets be available on the TEE-side. In practical terms, this should not change the UUT's response since the indoor unit would still be experiencing the correct virtual building load-calculated RAT, while the UUT thermostat would be experiencing the same RAT trend, only offset by the TEE to account for any onboard temperature sensor "drift".

d) Full-load heating test convergence criterion

NRCan testing revealed that some full-load heating tests converged following two 20-minute periods of steady-state operation with a COP difference of ≤ 1.0%. When comparing results with Lab A, however, there were differences in convergence criterion used that resulted in performance discrepancies. In some instances, Lab A observed the UUT initiate a defrost operation before two 20-minute steady-state periods were achieved, thereby making the convergence based on three defrost cycles, or a full 6 or 7-hour period based on irregular operation. To avoid situations in which the UUT initiates a defrost at different intervals at different labs, and to harmonize the convergence criteria between part-load and full-load testing, it is recommended to increase the number of steady-state periods required for convergence in full-load heating tests to three 20-minute periods.

e) Thermostat setup tracking in SPE-07 reporting template

Unit A's thermostat setup was not clear when NRCan received it following Lab A's testing. Unfortunately, no field or lab notes were taken to specify what settings were used if not default. This led to confusion where NRCan observed some airflow differences in some rating tests when compared with Lab A. Although both test facilities worked to solve this issue, no definitive conclusion could be reached regarding the setup used in the field and at Lab A. It is therefore recommended that the SPE-07 reporting template be amended to include a section for thermostat setup notes if default settings are not used for any reason.

f) Specify stabilized conditions

SPE-07 clause B.2.2.2.3.3. subpart b) specifies that the test operator initiate the virtual building load process "*When [the] indoor and outdoor room conditions have stabilized (...)*". Although the 15-minute period required before the start of measurement integration, and after this "stabilization" condition has been met ensures the room conditions are indeed stable, more specific language, which could include a prescribed duration of stabilization time, could eliminate the variability between test facilities. In equipment with a variety of onboard sensor types, with intrinsic and software-created delays to account for environmental variabilities such as solar gain, this stabilization time could be crucial to ensure reproducible behaviour between test facilities.

g) Manufacturer specified discharge orientation

SPE-07 requires that a manufacturer's "compliance certification report" be used to determine the discharge orientation in multi-position centrally ducted heat pumps as stated in clause B.1.2.1.2. However, no such reports were found for Units A and C. To avoid using different indoor unit orientations and to standardize the procedure, it is recommended that a discharge orientation be specified in SPE-07 for all multi-position centrally ducted air handling units.

6.0. References

[1] P. Dhillon, X. Wang, W. T. Horton, and J. E. Braun, "Repeatability and Reproducibility Assessment of Residential Heat Pump Performance Evaluation Methodologies based on CSA EXP07 and AHRI 210/240," 2022, [Online]. Available: http://docs.lib.purdue.edu/iracc/2477/

[2] P. Dhillon, D. Welch, B. Butler, W. T. Horton, and J. E. Braun, "Validation of a Load-Based Testing Method for Characterizing Residential Air-Conditioner Performance," 2021, [Online]. Available: https://docs.lib.purdue.edu/iracc/2257

[3] B. Harley, J. Butler, C. Dymond, G. Hamer, and D. P. Yuill, "Field tests of variable speed heat pumps to compare load-based and fixed-speed test and rating methods," 2023, [Online]. Available: https://www.hpc2023.org/wp-content/uploads/gravity_forms/3-7075ba8a16c5f78b321724d090fb2a34/2023/05/0438_HPC2023_Full_Paper_Yuill_v02.pdf

[4] Northeast Energy Efficiency Partnerships, "NEEP Residential Heat Pump Efficiency Rating Representativeness Project Phase 1," Northeast Energy Efficiency Partnerships, Lexington, MA, Oct. 2023. [Online]. Available: https://neep.org/sites/default/files/media-files/neep_heatpump_rating_representativeness_study_phase1_final_small.pdf

[5] Northeast Energy Efficiency Partnerships, "NEEP Residential Heat Pump Efficiency Rating Representativeness Project Phase 2," Northeast Energy Efficiency Partnerships, Lexington, MA, Oct. 2024. [Online]. Available: https://neep.org/sites/default/files/media-files/neep_heatpump_rating_representativeness_study-phase2_final.pdf

[6] H. F. Crowther *et al.*, "ANSI/ASHRAE Standard 37-2009 Methods of Testing for Rating Electrically Driven Unitary Air-Conditioning and Heat Pump Equipment," 2009.

[7] D. Kim, P. Dhillon, W. T. Horton, and J. E. Braun, "An Improved Thermostat Environment Emulator (TEE) for Load-based Testing," 2024, [Online]. Available: https://docs.lib.purdue.edu/iracc/2680/

[8] B. Harley, "Heat Pump and Air Conditioner Efficiency Ratings - Why Metrics Matter," Northeast Energy Efficiency Partnerships, Apr. 2022. [Online]. Available: https://neea.org/resources/heat-pump-and-air-conditioner-efficiency-ratings-why-metrics-matter

[9] B. Harley, M. Alatorre, C. Dymond, and G. Hamer, "CSA EXP07: Ongoing Progress, Lessons Learned, And Future Work in Load-based Testing of Residential Heat Pumps," 2022, [Online]. Available: https://neea.org/resources/csa-exp07-ongoing-progress-lessons-learned-and-future-work-in-load-based-testing-of-residential-heat-pumps

[10] B. Harley, "EXP07:19 Load-based and Climate-Specific Testing and Rating Procedures for Heat Pumps and Air Conditioners - Interim Lab Testing and Rating Results," NEEA, Jul. 2020. [Online]. Available: https://neea.org/resources/exp0719-load-based-and-climate-specific-testing-and-rating-procedures-for-heat-pumps-and-air-conditioners

[11] P. Dhillon, D. T. W. Horton, and D. J. E. Braun, "AHRI 8026 - Repeatability and Reproducibility Assessment of CSA EXP07:19 and AHRI 210-240:2023," Dec. 2022, [Online]. Available: https://www.ahrinet.org/system/files/2023-06/AHRI%208026%20Final%20Report%20with%20CP.pdf

[12] N. Giannetti *et al.*, “Feed-forward compensation for emulator-type testing facilities,” *Int. J. Refrig.*, vol. 167, pp. 257–268, Nov. 2024, doi: 10.1016/j.ijrefrig.2024.07.023.

7.0. Appendices

7.1. Appendix A – Detailed Rating Test Results

7.1.1. Unit A

	NRCan - Unit A									
	Q_C	E_C	COP_C	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
CE_D	5676.2	363.6	4.58	74.5	0.0083	76.8	N/A	N/A	14.63	75
CD_D	9934.0	762.3	3.82	74.8	0.0077	86.1	N/A	N/A	14.58	75
CC_D	12604.2	1161.3	3.18	76.5	0.0084	95.1	N/A	N/A	14.62	75
CB_D	19275.2	2058.1	2.74	75.0	0.0085	104.6	N/A	YES	14.80	74
CE_H	8819.1	583.3	4.43	70.0	0.0083	76.8	N/A	N/A	14.61	70
CD_H	14181.9	1082.6	3.84	70.1	0.0081	86.0	N/A	N/A	14.53	70
CC_H	19572.5	1790.0	3.20	70.0	0.0086	95.1	N/A	YES	14.69	68
CB_H	19348.8	2064.8	2.75	70.0	0.0092	104.6	N/A	YES	14.67	Min (52)
	Q_H	E_H	COP_H	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
HF_C	2114.5	312.8	1.98	70.1	0.0048	53.6	0.0047	N/A	14.61	69
HE_C	4699.0	513.5	2.68	68.6	0.0047	46.7	0.0042	N/A	14.55	68
HD_C	9442.8	1096.1	2.52	68.3	0.0051	34.0	0.0031	N/A	14.60	68
HC_C	15588.0	2508.5	1.82	69.4	0.0048	17.1	0.0013	N/A	14.65	70
HB_C	13299.7	3128.4	1.25	69.0	0.0062	5.0	0.0011	YES	14.48	Max (88)
HL_C	9979.2	2760.8	1.06	69.0	0.0046	-3.0	0.0005	YES	14.41	Max (88)
HD_M	9290.1	1087.2	2.50	68.2	0.0046	34.0	0.0035	N/A	14.44	68

	Lab A - Unit A									
	Q_C	E_C	COP_C	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
CE_D	4813.2	313.6	4.50	75.1	0.0099	77.0	N/A	N/A	14.36	75
CD_D	9353.7	660.3	4.15	75.4	0.0082	86.0	N/A	N/A	14.36	75
CC_D	13768.9	1137.0	3.55	76.1	0.0082	95.3	N/A	N/A	14.34	75
CB_D	18557.2	1852.1	2.94	74.7	0.0079	104.0	N/A	YES	14.38	74
CE_H	8540.7	496.0	5.05	70.9	0.0080	77.0	N/A	N/A	14.37	70
CD_H	14146.6	998.3	4.15	69.9	0.0077	86.0	N/A	N/A	14.42	69
CC_H	19340.7	1651.3	3.43	70.0	0.0082	95.0	N/A	YES	14.38	69
CB_H	19741.3	1977.5	2.93	69.9	0.0093	104.0	N/A	YES	14.38	N/A
	Q_H	E_H	COP_H	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
HF_C	1772.0	271.9	1.91	70.8	0.0030	54.0	0.0045	N/A	14.43	69
HE_C	4186.0	438.7	2.80	69.5	0.0029	47.0	0.0042	N/A	14.38	68
HD_C	9502.7	939.5	2.96	69.2	0.0033	34.0	0.0031	N/A	14.42	68
HC_C	14715.0	1908.7	2.26	67.7	0.0034	17.0	0.0013	N/A	14.40	68
HB_C	14851.8	2960.6	1.47	69.0	0.0027	5.0	0.0010	N/A	14.37	N/A
HL_C	9726.8	2511.0	1.14	69.0	0.0025	-3.0	0.0005	YES	14.36	N/A
HD_M	9615.3	952.2	2.96	69.2	0.0029	34.1	0.0035	N/A	14.35	68

7.1.2. Unit C

	Unit C - NRCan									
	Q_C	E_C	COP_C	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
CE_D	3501.4	390.4	2.63	78.6	0.0100	77.0	N/A	N/A	14.68	78
CD_D	9068.2	561.9	4.73	79.0	0.0076	86.0	N/A	YES	14.76	76
CC_D	15261.8	1457.1	3.07	79.0	0.0086	95.0	N/A	YES	14.66	67 (min)
CB_D	18318.9	2321.9	2.31	79.0	0.0080	104.0	N/A	YES	14.63	67 (min)
CE_H	7619.5	742.5	3.01	74.3	0.0092	77.0	N/A	N/A	14.62	67 (min)
CD_H	15187.6	1317.8	3.38	73.9	0.0087	86.0	N/A	YES	14.65	67 (min)
CC_H	14710.9	1458.6	2.96	73.9	0.0091	95.0	N/A	YES	14.70	67 (min)
CB_H	17042.0	2290.3	2.18	73.9	0.0081	104.0	N/A	YES	14.70	67 (min)
	Q_H	E_H	COP_H	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
HF_C	2517.6	234.1	3.15	74.3	0.0072	54.0	0.0045	N/A	14.68	69
HE_C	4713.9	497.8	2.78	74.1	0.0074	47.0	0.0042	YES	14.73	69
HD_C	8337.8	1016.0	2.41	73.7	0.0072	34.1	0.0031	N/A	14.76	70
HC_C	16937.2	2965.3	1.67	73.9	0.0065	17.1	0.0013	YES	14.57	83 (max)
HB_C	13268.8	2518.2	1.54	73.9	0.0058	5.5	0.0008	YES	14.59	83 (max)
HL_C										
HD_M	8561.6	1080.5	2.32	73.1	0.0064	34.3	0.0035	N/A	14.73	70

	Unit C - Lab A									
	Q_C	E_C	COP_C	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
CE_D	4312.2	369.8	3.42	78.5	0.0095	77.0	N/A	N/A	14.55	80
CD_D	8967.5	486.3	5.40	78.3	0.0096	86.0	N/A	YES	14.56	80
CC_D	13937.7	1194.1	3.42	80.0	0.0083	95.0	N/A	N/A	14.60	80
CB_D	22630.7	3252.1	2.04	78.9	0.0089	104.0	N/A	YES	14.61	N/A
CE_H	7814.1	410.3	5.58	73.9	0.0100	77.0	N/A	N/A	14.58	74
CD_H	14565.1	1242.7	3.43	74.8	0.0084	86.0	N/A	N/A	14.58	72
CC_H	20399.1	2291.3	2.61	74.8	0.0074	95.0	N/A	YES	14.56	72
CB_H	21669.3	3224.4	1.97	73.9	0.0089	104.1	N/A	YES	14.57	N/A
	Q_H	E_H	COP_H	ID DB	ID HR	OD DB	OD HR	CONV	BP	TSTAT
HF_C	2708.2	224.4	3.54	75.2	0.0044	54.0	0.0045	N/A	14.51	72
HE_C	5226.8	492.8	3.11	75.3	0.0044	47.0	0.0042	N/A	14.49	72
HD_C	8342.3	981.0	2.49	73.8	0.0043	34.1	0.0031	N/A	14.41	72
HC_C	11170.8	1570.8	2.08	74.1	0.0057	17.0	0.0013	YES	14.26	N/A
HB_C	10900.7	1837.7	1.74	74.0	0.0040	5.0	0.0008	N/A	14.40	N/A
HL_C	6995.3	1715.8	1.19	74.0	0.0040	-13.0	0.0002	N/A	14.52	N/A
HD_M	8517.4	1004.3	2.49	74.2	0.0044	34.0	0.0035	N/A	14.41	72

7.2. Appendix B – Changes to the SPE-07:23 Building Load Line

7.2.1. Unit A

Table 13. Changes to the SPE-07 standard load line values for Unit A versus the load line values as measured during field-testing and used in the reproducibility testing in cooling humid conditions

	SPE-07 Values	Modified Field-Test Values
Q_C(95)	23,600 BTU/hr	21,977 BTU/hr
T_{ZL}	67 °F	62 °F
T_{ID}	74 °F	70 °F
$SHR_{Building}$	0.80	0.83

Table 14. Changes to the SPE-07 standard load line values for Unit A versus the load line values as measured during field-testing and used in the reproducibility testing in cooling dry conditions

	SPE-07 Values	Modified Field-Test Values
Q_C(95)	23,600 BTU/hr	21,977 BTU/hr
T_{ZL}	72 °F	67 °F
T_{ID}	79 °F	75 °F

Table 15. Changes to the SPE-07 standard load line values for Unit A versus the load line values as measured during field-testing and used in the reproducibility testing in heating conditions

	SPE-07 Values	Modified Field-Test Values
Q_C(95)	23,600 BTU/hr	17,318 BTU/hr
T_{ZL}	60 °F	59 °F
T_{ID}	70 °F	69 °F

7.2.2. Unit C

Table 16. Changes to the SPE-07 standard load line values for Unit C versus the load line values as measured during field-testing and used in the reproducibility testing in cooling humid conditions

	SPE-07 Values	Modified Field-Test Values
$Q_C(95)$	18,000 BTU/hr	24,699 BTU/hr
T_{ZL}	67 °F	66 °F
$SHR_{Building}$	0.80	0.86

Table 17. Changes to the SPE-07 standard load line values for Unit C versus the load line values as measured during field-testing and used in the reproducibility testing in cooling dry conditions

	SPE-07 Values	Modified Field-Test Values
$Q_C(95)$	18,000 BTU/hr	24,699 BTU/hr
T_{ZL}	71 °F	72 °F

Table 18. Changes to the SPE-07 standard load line values for Unit C versus the load line values as measured during field-testing and used in the reproducibility testing in heating conditions

	SPE-07 Values	Modified Field-Test Values
$Q_C(95)$	18,000 BTU/hr	15,222 BTU/hr
T_{ZL}	60 °F	62 °F
T_{ID}	70 °F	74 °F

Contact

Jean-Philippe B. Poirier, P.Eng.
Research Engineer
Buildings and Renewables Group
Natural Resources Canada, CanmetENERGY-Ottawa
jean-philippe.poirier@nrcan-rncan.gc.ca

About CanmetENERGY

Natural Resources Canada's CanmetENERGY is the Canadian leader in clean energy research and technology development. Our experts work in the fields of clean energy supply from fossil fuel and renewable sources, energy management and distribution systems, and advanced end-use technologies and processes. Ensuring that Canada is at the leading edge of clean energy technologies, we are improving the quality of life of Canadians by creating a sustainable resource advantage.

Head Office	Devon, Alberta	Ottawa, Ontario	Varennes, Québec
580 Booth Street	1 Oil Patch Drive	1 Haanel Drive	1615 Lionel-Boulet Boulevard
Ottawa, ON	Devon, AB	Ottawa, ON	Varennes, QC
Canada	Canada	Canada	Canada
K1A 0E4	T9G 1A8	K1A 1M1	J3X 1S6