15TH IEA HEAT PUMP CONFERENCE
MAY 26–29, 2026 | HOFBURG
VIENNA 2026
DECARBONISATION THROUGH INNOVATION

Evaluating the reproducibility of a load-based and climate-specific test procedure using a centrally ducted cold-climate air-source heat pump

Jean-Philippe B. Poirier[a,*], Jeremy Sager[a], Bruce Harley[b]

[a]CanmetENERGY-Ottawa, Natural Resources Canada, Ottawa K1A 1M1, Canada
[b]Bruce Harley Energy Consulting LLC, Stamford 05352, United States

Abstract

The authors evaluated the reproducibility of a load-based rating and test procedure using a centrally ducted cold-climate air-source heat pump as part of the "3R's" testing (Representativeness, Repeatability, Reproducibility). The procedure requires heat pumps to operate in all operating modes with native controls, such that performance results encompass steady-state and part-load cyclical operation while also taking into consideration defrost cycles and standby energy consumption. The unit under test (UUT) was a variable speed 7 kW capacity system with communicating thermostat that had previously been field and laboratory tested for evaluations of representativeness and repeatability. The UUT was installed and tested in a laboratory designed to meet the requirements of the procedure's testing tolerances and virtual building load emulation. The objective was to produce cooling and heating seasonal coefficients of performance ($SCOP_C$ and $SCOP_H$) from a complete series of rating tests using the same building load line and equipment installation as that used during previous testing conducted in another laboratory. The UUT achieved an overall reproducibility of "fair", with 8.6 % mean difference in $SCOP_C$, and 8.8 % mean difference in $SCOP_H$ between facilities. There were several factors affecting reproducibility identified during testing, including (a) differences in instrumentation measurement lag between testing facilities; (b) potential indoor chamber response lag; (c) issues in properly measuring UUT capacity during defrost operations; (d) installation differences between facilities; (e) equipment and thermostat dynamic response differences between labs, and (f) an unverifiable thermostat setup. Overall, the UUT was consistent in reproducing individual part-load and full-load behaviour under matching test conditions, which reflects favourably on the procedure's ability to induce a consistent equipment response between testing facilities.

Keywords: heat pump; cold-climate; load-based testing; reproducibility

1. 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

* Corresponding author. Tel.: +1 (343) 543-7693
E-mail address: jean-philippe.poirier@nrcan-rncan.gc.ca

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 across all test intervals. 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 a real-world application. The product of this work was the Canadian Standards Association's (CSA) EXP-07, published in 2019 (EXP-07:19).

Extensive laboratory testing of EXP-07:19 beginning in 2018 had revealed substantial differences between the fixed-speed, or "steady-state" 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 suggested that more in-depth understanding of the differences between the two test approaches would require an understanding of the reproducibility and repeatability of both testing methods.

For reproducibility assessment, a unit is tested in one lab, then shipped to another lab 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. Audit testing for enforcement purposes are nearly always conducted by a different lab than tests used for rating purposes, so large variations 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, repeatability tests using EXP-07:19 tests [1] showed variations from test to test in the same lab of less than 5 %, similar to the steady-state test repeatability, except in instances where the unit was 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 EXP-07, compared with 1.5 % to 4.5 % variation using steady-state testing. 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 variations in setup and implementation of the load-based test process between test labs. It was therefore recommended to repeat an assessment of reproducibility using the SPE-07 procedure to estimate the extent of improvement, if any.

The reproducibility testing was conducted to address these concerns, with NRCan's CanmetENERGY-Ottawa (CE-O) committed to contributing to testing via use of its Climate Controlled Test Facility (CCTF). The testing results generated at NRCan were compared with those produced at a certification laboratory, hereafter referred to as "Lab A", with the same unit. A second heat pump was also tested but thermostat anomalies led to unreliable test results.

2. Methods/Approach

SPE-07 is a load-based test method that uses a sequence of simulated heating and cooling loads that are imposed in the indoor test chamber. The simulated, or "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 UUT is set up to heat or cool the indoor chamber using its own controls in response to the virtual loads, while its heating or cooling capacity and power input are measured in real time.

At each test interval condition outlined in SPE-07 (shown in Table 1 and Table 2), the outdoor room temperature is held constant, while the indoor room reconditioning equipment is programmed to respond to the current performance of the UUT. 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 UUT's thermostat to sense that condition and increase its capacity in response. The virtual building load includes a term for realistic indoor mass capacitance, which slows the room temperature response, helps stabilize operation during the test, and eliminates test chamber capacitance from the response.

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 this 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 based on limited testing, as is the case in conventional fixed-speed testing. Experience has shown that some units' behaviour and rated efficiency can vary dramatically with load-based testing versus fixed-speed testing [2–4].

In the normal calculation process of SPE-07, the results of the tests are mapped into a set of climate models based on tranches of outdoor temperature ranges, or "bins", to create a set of seasonal coefficients of performance (SCOP) for eight heating and seven cooling climates. The SCOPs presented in this paper were calculated using the reporting template spreadsheet available with SPE-07.

2.1. Laboratory testing overview

The UUT was a 7-kW capacity, variable speed, cold-climate heat pump with a vertically discharging outdoor unit, and a multi-position, ducted indoor unit.

NRCan conducted SPE-07 testing at the CCTF, which is a laboratory designed for testing residential heating and cooling equipment located at the Bells Corners Complex in Ottawa, Canada. It is equipped with outdoor and indoor environment chambers 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 CCTF was designed to use the tunnel air enthalpy method as described in ASHRAE 37 [5] to obtain the air-side measurements of capacity, as shown in Fig. 1.

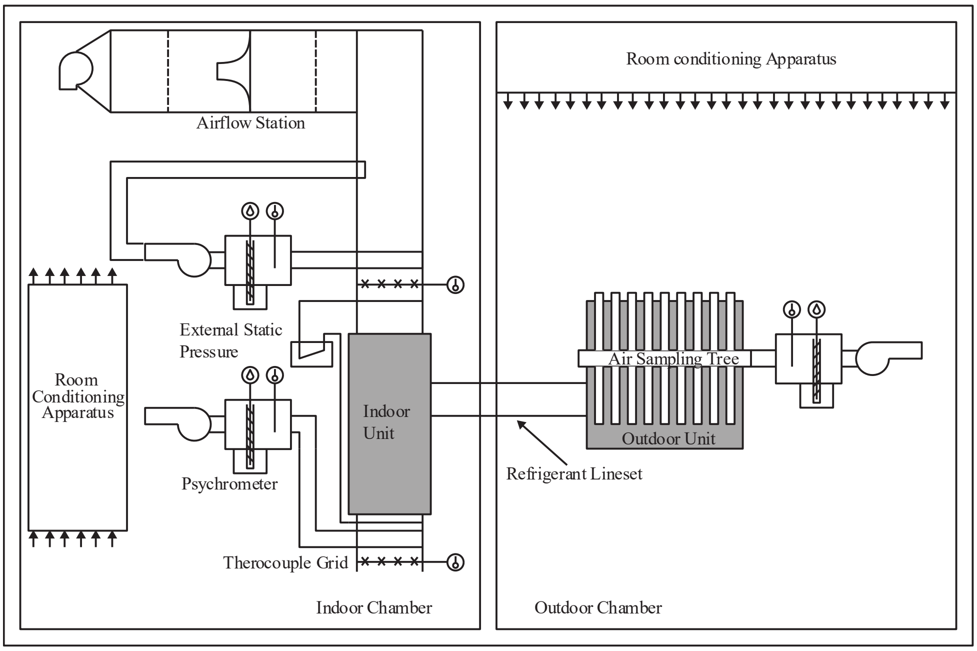

Figure 1. ASHRAE 37 tunnel air enthalpy arrangement at the CCTF

The outdoor chamber is also equipped with a secondary room reconditioning apparatus that is activated when either the primary system, or the UUT, is undergoing a defrost operation, in order to maintain the outdoor temperature setpoint.

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 in-duct heater to precisely control the dry-bulb temperature. Circulation fans are strategically located in the chamber to ensure the air is properly mixed and uniform. A steam humidifier with 1 %-precision relative humidity control, as well as a mechanical dehumidifier are used to control the humidity ratio.

A thermostat environment emulator (TEE) designed and built by Purdue University [6] and shown in Fig. 2, was used to precisely control the temperature and air velocity experienced by the UUT's thermostat and eliminate variability in thermostat response due to placement in the indoor test chamber. The TEE's onboard conditioning equipment and controls ensure that the thermostat housed within it receives the same room air temperature from the virtual building load emulator software that is maintained in the indoor test chamber and experienced by the UUT. The TEE was installed in laboratory space that is maintained at approximately 23 °C year-round.

Figure 2. Thermostat environment emulator (TEE) installed in the main laboratory area

3. Results and Discussion

The test begins with a "learning" series, used to calibrate the thermostat for each test condition to determine the setpoint temperature that most closely induces a UUT response resulting in an indoor temperature maintained at the control target indoor dry-bulb temperature. The thermostat setpoint for test interval HC differed between laboratories, which is likely due to an issue in capturing defrost cycles and the resulting drop in indoor dry-bulb temperature as explained in section 3.3.

3.1. Heating rating test summary

The results of capacity, input power, coefficient of performance (COP), and SCOP, are presented below with Lab A's test results serving as the reference against which NRCan's results are compared.

The heating learning tests determined that test intervals HF through HC, shown in Table 1, were part-load tests, whereas test intervals HB and HL were full-load, or maximum capacity, tests. This behaviour was consistent between Lab A and NRCan as were the thermostat setpoints.

Table 1. UUT thermostat setpoints in heating mode for NRCan and Lab A as determined by the heating learning series tests (C = Continental, M = Marine)

Test interval	Outdoor dry-bulb temperature (°C)	Indoor dry-bulb control target (°C)	Thermostat setpoint at NRCan (°C)	Thermostat setpoint at Lab A (°C)
HF_C	12.2	20.6	20.6	20.6
HE_C	8.3	20.6	20	20
HD_C	1.1	20.6	20	20
HC_C	-8.3	20.6	21.1	20
HB_C	-15	Max	Max (31.1)	Max (31.1)
HL_C	-19.4	Max	Max (31.1)	Max (31.1)
HD_M	1.1	20.6	20.6	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 Fig. 3 reveal that test intervals HF and HE had higher capacities at NRCan's test facility, test HB had lower capacity, and other tests were more consistent. The UUT's power consumption, however, as shown in Fig. 4, was consistently higher at NRCan's test facility for all rating tests, with test interval HC being particularly high at NRCan, with more detail on the likely contributor being presented in section 3.3.

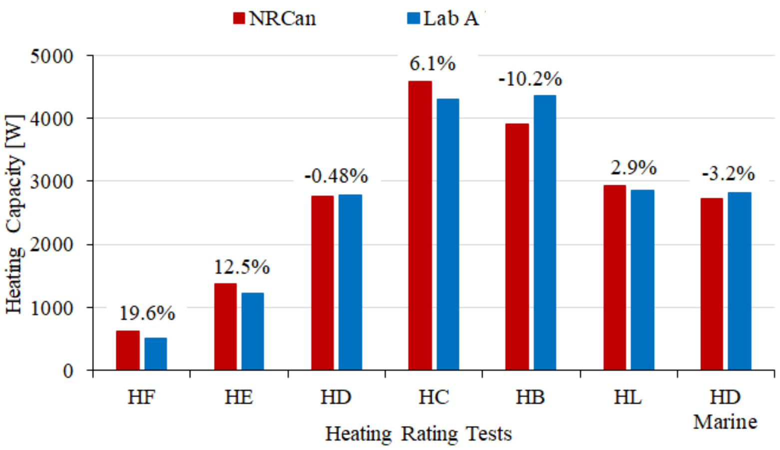

Figure 3. UUT heating capacity in watts, NRCan and Lab A results

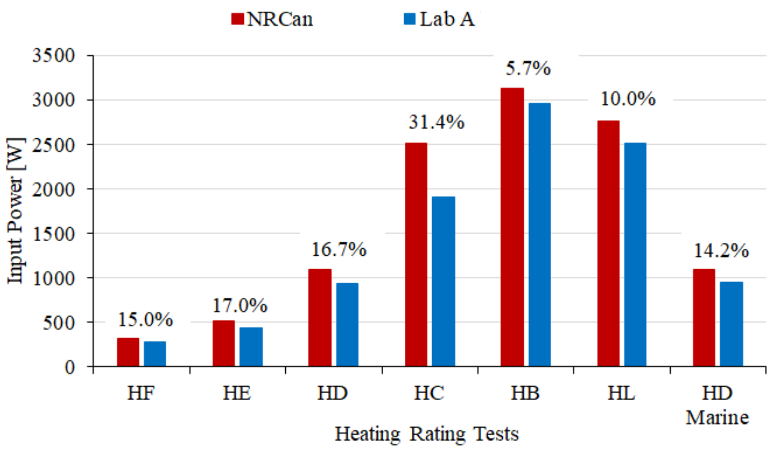

Figure 4. UUT input power in watts, NRCan and Lab A results

Fig. 5 shows that the rating test COPs at intervals HD continental and marine (characterized by different outdoor humidity ratios), HC, and HB diverged most substantially between test facilities.

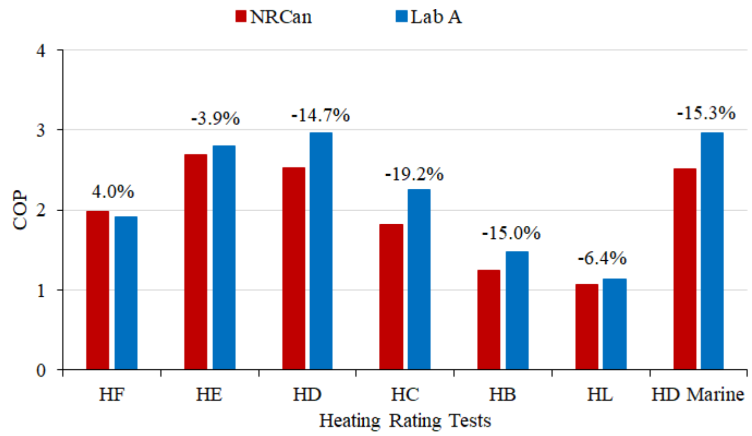

Figure 5. UUT heating COPs, NRCan and Lab A results

Finally, the heating SCOP's ($SCOP_H$) for the climate zones evaluated are presented in Fig. 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 %, which represents "fair" reproducibility overall. The subarctic and very cold climates are considered "good" in reproducibility terms, while the rest of the climates are considered "fair" or "poor".

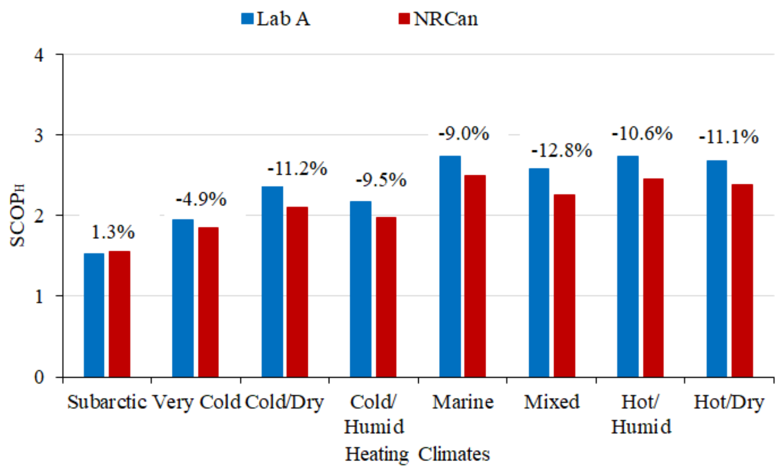

Figure 6. UUT $SCOP_H$, NRCan and Lab A results

Fig. 7 shows the UUT's behaviour during rating test interval HD continental (1.1°C) and demonstrates a consistent equipment response at the individual test interval level. A comparison of the heating capacity, building load and input power at NRCan (top) and Lab A (bottom) is presented. This test interval was part-load and ran for 6 hours as required by SPE-07, with six heating cycles and one defrost cycle being observed at both test facilities. The average heating capacity was 6-7 kW for both tests, with airflows being 0.024-0.035 m^3/s higher on average at Lab A. The input power consumption diverged slightly between tests as well, with NRCan observing approximately 160 W higher peak input power during heating cycles. Green shading is used in the figure to highlight the difference in heating capacity slope between NRCan and Lab A. This was mainly due to different instruments being used to measure the supply air temperature, leading to the observed gradual increase and decrease in heating output at Lab A, for the equivalent duration of compressor run time as represented by the input power in blue.

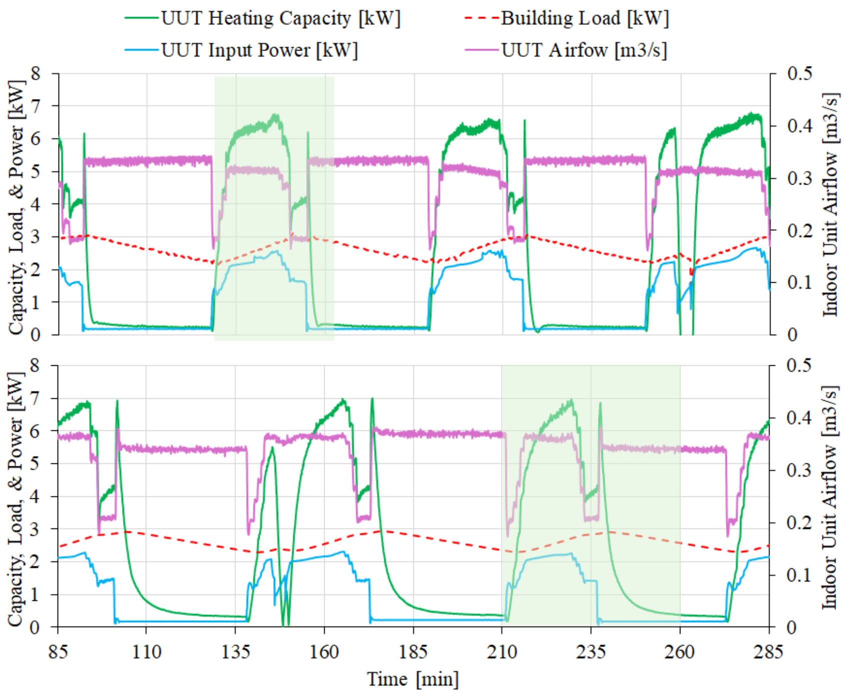

Figure 7. UUT capacity, building load, and input power results at NRCan (top) and Lab A (bottom) for heating continental test interval HD – Continental (1.1°C).

3.2. Cooling rating test summary

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 conditions 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 2. Thermostat setpoints determined in the learning test for conditions CD and CC in humid conditions differed slightly between laboratories, resulting in consistent indoor temperatures during testing, this being the intent of the thermostat calibration step during the learning tests.

Table 2. UUT thermostat setpoints in cooling mode for NRCan and Lab A as determined by the heating learning series tests

Test interval	Outdoor dry-bulb temperature (°C)	Indoor dry-bulb control target (°C)	Thermostat setpoint at NRCan (°C)	Thermostat setpoint at Lab A (°C)
CE_D (dry)	25	23.9	23.9	23.9
CD_D	30	23.9	23.9	23.9
CC_D	35	23.9	23.9	23.9
CB_D	40	23.9	23.3	23.3
CE_H (humid)	25	21.1	21.1	21.1
CD_H	30	21.1	21.1	20.6
CC_H	35	21.1	20.0	20.6
CB_H	40	Min	Min (11.1)	Min (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 combine latent and sensible cooling, are presented in Fig. 8 for dry and humid conditions. The results indicate consistency between facilities, except for test CE in dry conditions.

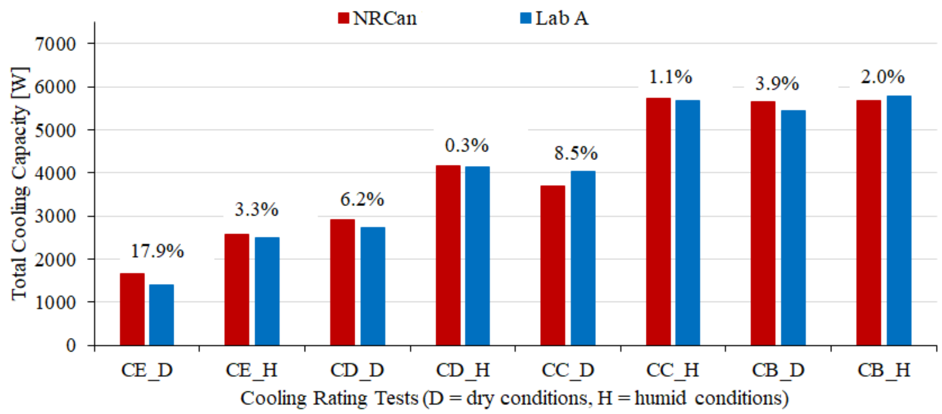

Figure 8. UUT total cooling capacity in watts, in dry and humid conditions, NRCan and Lab A results

The UUT's power consumption by test interval is presented in Fig. 9 for dry and humid conditions. Here, we can observe a consistent trend towards more power consumption at NRCan across all test intervals.

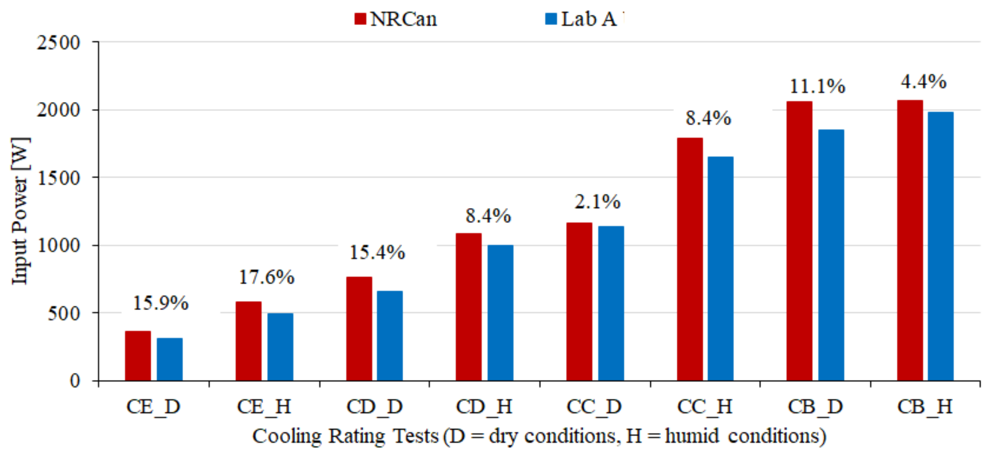

Figure 9. UUT input power in watts, in dry and humid cooling conditions, NRCan and Lab A results

The resulting rating test COPs shown in Fig. 10 reveal that all tests had lower COPs at NRCan due to the higher input power, except for test CE in dry conditions.

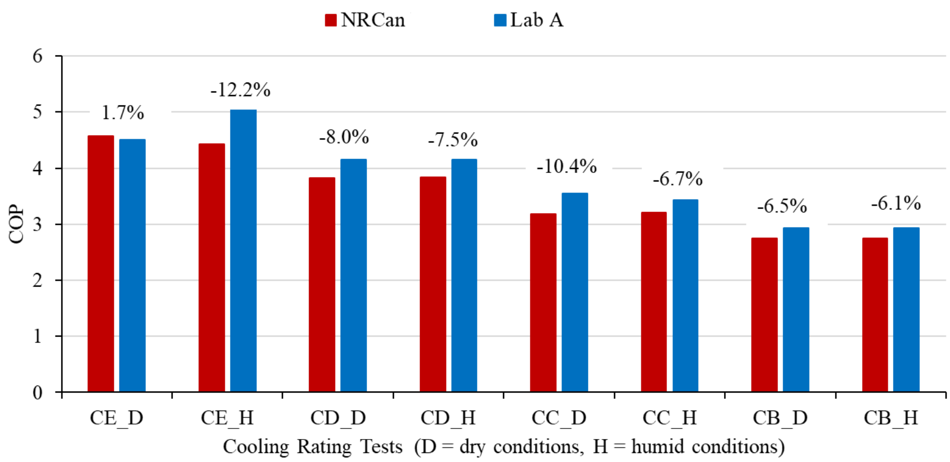

Figure 10. UUT COP, in dry and humid cooling conditions, NRCan and Lab A results

Finally, the cooling SCOP's ($SCOP_C$) for the climate zones evaluated are presented in Fig. 11. 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.

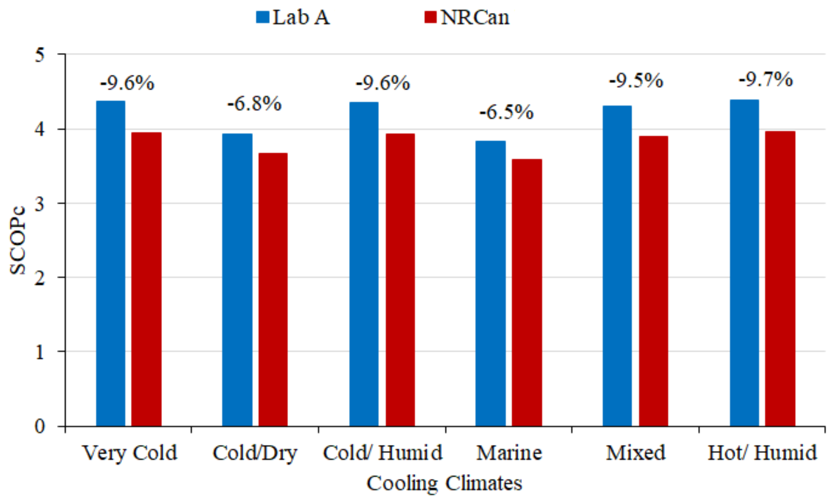

Figure 11. UUT $SCOP_C$, NRCan and Lab A results

3.3. Discussion of results

With a mean difference of 8.8 % in $SCOP_H$ and 8.6 % in $SCOP_C$, the UUT achieved "fair" reproducibility. In heating mode, the UUT's capacity was fairly reproduced in both labs, with some lower-load test intervals being outliers. However, these same lower-load test intervals such as HF (12.2 °C) and HE (8.3 °C) were also characterized by higher power consumption at NRCan, 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 similar differential 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. Testing revealed that the UUT's indoor blower stayed on during defrost operations; something which is not common in today's cold-climate air-source heat pumps. Lab A did not capture the negative heating capacity that was 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 while in defrost for test interval HC (-8.3 °C), adding a significant load for the UUT to overcome. 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. The UUT's response therefore diverged between facilities when defrost cycles occurred, leading to poor reproducibility for these test intervals. It also likely contributed to the difference in indoor thermostat setpoints as determined by the learning test for test interval HC, which in turn increased the discrepancy between the labs. The broader impact was also observed when comparing $SCOP_H$'s in climate regions with bin fractions that are heavily skewed by these test intervals. 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 testing. The error was indeed a misapplication of the procedure itself during heating mode tests, rather than a variation in lab results while following the procedure.

In cooling mode, the UUT's capacity was well reproduced between Lab A and NRCan, with test interval CE (25 °C) in dry conditions 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 in humid conditions 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 power being consumed by the UUT to maintain the same target indoor dry-bulb temperature. This trend was again observed during the CC (35 °C) test interval in dry conditions and resulted in a 10.2 % lower COP at NRCan.

No issues were observed in the usage of the thermostat environment emulator during the UUT's testing at NRCan. However, some potential issues were raised with respect to the UUT's thermostat settings. 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 would typically require airflow set to "Auto". 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 to ensure consistency of airflows between Lab A and NRCan. This introduced uncertainty surrounding the thermostat's setup between facilities that was never resolved. In some rating test intervals such as HC (-8.3 °C), NRCan observed lower airflows which could be explained by a different thermostat operating mode.

3.4. Factors affecting reproducibility

The virtual building load is dependent on instrumentation as it is calculated using the UUT's measured capacity. 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 for a calculation of mass flow rate. SPE-07 allows for the indoor unit return and supply temperatures to be taken by either a resistance temperature detector (RTD) sensor in an aspirated psychrometer, or a grid of temperature sensors, usually thermocouples, in the air stream. In principle, these two measurement methods, as well as the instrumentation used, will result in some differences. Fig. 12 demonstrates the difference between a ¼-inch Class A RTD sensor measurement of dry-bulb temperature through an aspirated psychrometer 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 sensor measurement in the aspirated psychrometer would result in different measured heating capacities as is apparent in Fig. 7, where Lab A's measured heating capacity (in green) was delayed, relative to the heating capacity measured at NRCan. This results in a different virtual building load, which in turn could trigger the UUT to respond, and therefore perform, differently in labs that use these sensors. The resulting difference is significant in a load-based test approach where fluctuations in return and supply temperatures are expected, as compared with steady-state tests with relatively constant temperatures.

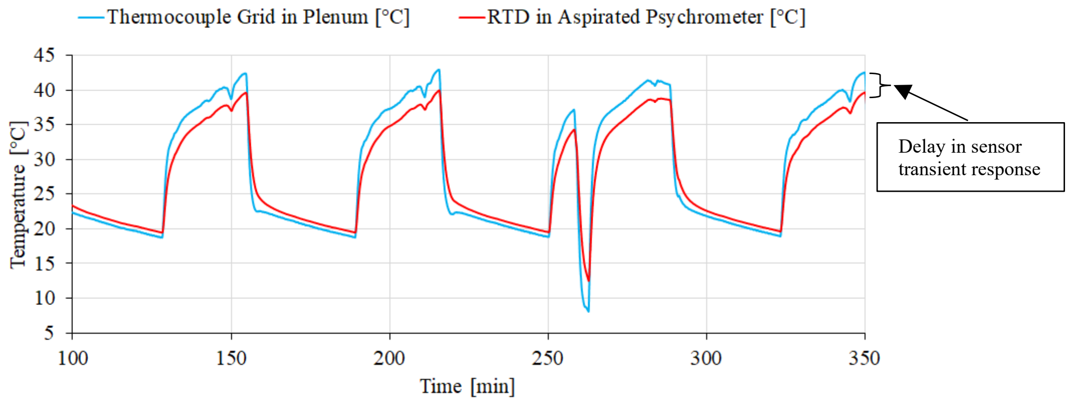

Figure 12. Difference in supply air measurement between an RTD in an aspirated psychrometer and a grid of thermocouples in the air stream at NRCan

As previously mentioned, the virtual building load, and the room air temperature target (RAT) by extension, is dependent on measurements that may have intrinsic delay, or "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 reconditioning equipment to respond, the UUT may momentarily experience a different return temperature than intended, thereby changing the load it must 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 [7]. Fig. 13 presents the difference between an indoor chamber with virtually no delay between UUT return air temperature and RAT, and a chamber with approximately 2.5 minutes of delay. Although both test facilities maintained the test operating and condition tolerances as defined in SPE-07, the potential delay in UUT response may lead to differences in overall test performance.

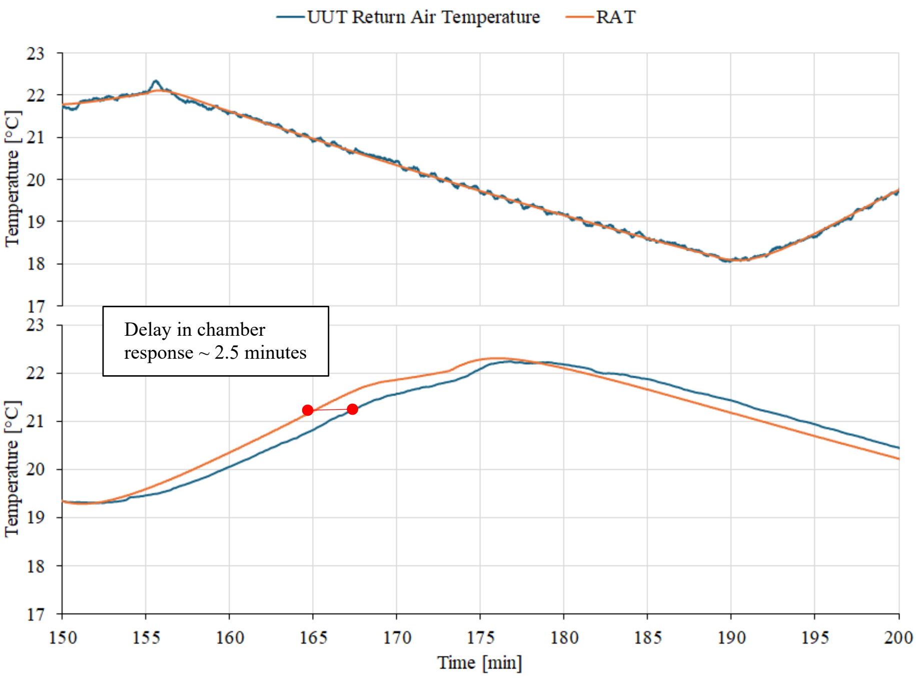

Figure 13. UUT return-air temperature versus RAT at NRCan (top) and Lab A (bottom)

4. Conclusions

This paper examined the reproducibility of the CSA SPE-07:23 procedure using a ducted cold-climate air-source heat pump evaluated at two independent testing facilities. The heat pump was a 7-kW capacity variable speed system with communicating thermostat 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.

The heat pump testing 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:

- Variation in indoor room temperature resulting from different learning test setpoint determinations,
- Instrumentation lag differences between testing facilities,
- Potential indoor chamber response lag in Lab A,
- Issues in properly measuring UUT capacity during defrost operations in Lab A,
- Installation differences between facilities such as the orientation of the indoor unit,
- Equipment and thermostat dynamic response differences, and,
- An unverifiable thermostat setup in Lab A.

The heat pump 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. Some of the variability in heating test results between laboratories appears to be related to differences in how the procedure was applied, particularly in accounting for negative heating capacity during defrost. If the virtual building load had been fully adjusted for these effects, the variations would likely have been smaller.

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 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. Improvements in the procedure could result from better instructions regarding capacity measurements during defrost; implementing standard installation requirements for equipment orientation; clarifying thermostat setup requirements; and implementing methods such as feed-forward compensation to reduce the impact of instrumentation and test chamber lag.

Acknowledgements

The authors would like to acknowledge Liam Jones and Cory Balderson from 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.

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] 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, July 2020. [Online]. Available: https://neea.org/resources/exp0719-load-based-and-climate-specific-testing-and-rating-procedures-for-heat-pumps-and-air-conditioners

[3] 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

[4] 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

[5] H. F. Crowther *et al.*, "ANSI/ASHRAE Standard 37-2024 Methods of Testing for Rating Electrically Driven Unitary Air-Conditioning and Heat Pump Equipment," 2024.

[6] 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/

[7] N. Giannetti *et al.*, "Feed-forward compensation for emulator-type testing facilities," *International Journal of Refrigeration*, vol. 167, pp. 257–268, Nov. 2024, doi: 10.1016/j.ijrefrig.2024.07.023.