1. State of the art of testing methodologies and performance rating standards for evaluating the actual energy efficiency of heat pumps and air conditioners

Statement of purpose

The following objectives summarise the main purpose of the present subtask (The term “heat pump” includes both air-to-air and hydronic heat pumps.):

- to review presently adopted testing methodologies and performance rating standards for air conditioners and heat pumps,
- to review newly proposed testing methodologies and performance rating standards for air conditioners and heat pumps,
- to define the requirements and tolerance for the testing equipment, instrumentation, and auxiliaries, the system operation and setpoints during the test, as well as performance indices for evaluating the actual system efficiency,
- to consider possible improvements of existing and new testing methodologies for assessing the performance of heat pumps and air conditioners when operated under the same control as operated in buildings,
- to consider methods of utilisation of the test results for performance rating, performance mapping, and energy calculation methods (Subtask C), provide evidence for efficient equipment sizing (Subtasks C and D), system design and control (Subtask D), as well as support for the development of performance monitoring techniques (Subtask B2).

1.1 Categories of testing standards and their backgrounds

Using the inverter technology for variable speed systems has enhanced the system´s adaptability in managing thermal loads with potentially high efficiency in a broad range of operating conditions. During field operation heat pumps and air conditioners respond to specific building loads and indoor temperature variations with dynamic modulations of the compressor speed and expansion valve opening defined by their built-in (or “on-board”) control system, which is referred to as “native control” and is opposed to proprietary modes used only during current rating tests. The control mode adopted during tests is at the essence of the definition of the following two categories of testing methodologies and rating standards. In fact, current testing procedures are experimental tests at fixed compressor speed conducted while deactivating the native control system of the unit.

Therefore, rating procedures based on such a simplified testing method yield results that may deviate significantly from the actual operating performance. This underlying gap between the actual and rated performances has been recognised as a major challenge for effectively driving energy conservation of heat pump installations, besides improving system design and quality of installation.

In the attempt to guide the development of new standards that can cover such performance gap and represent the system field efficiency more realistically, this subtask reviews two categories of testing methodologies and corresponding performance rating standards:

- (Category A) Current testing methodologies and performance rating standards conducted at fixed compressor speed (fixed capacity ratio) while deactivating the native control system.
- (Category B) Newly proposed testing methodologies for evaluating the performance of heat pumps

and air conditioners under the same control as operated in buildings (active native control).

Table 1.1-1 compares Category A standards and Category B standards for heat pumps. The operation mode of the unit defined by the testing methods differentiates the two categories.
Performance rating standards are intended as product-level policies, which provide values representing products' energy efficiency and are referred by users to compare different products of the same kind. However, it should be noted that new testing methods for evaluating heat pumps operated under their native control can also apply to building-level evaluation, extract performance maps, and to support energy calculations.

Table 1.1-1. Comparison of Category A standards and Category B standards.

	Category A standards: Development in HP industry	Category B standards: Being developed by independent research entities
(1) Operation mode during tests for energy efficiency	Native control is overridden by proprietary control as required for testing. Generally, the tested unit is forced in steady state condition by fixing the compressor speed (with proprietary controls). Generally, provides reliable hardware testing, but excludes evaluation of performance in all modes of operation across the operating temperature range. This way of testing is often considered indispensable to maintain a high accuracy and reproducibility, but this comes at the expense of comprehensive performance evaluation of the unit under test. Note that native control testing, cycling mode and defrost mode operation may be tested at discrete conditions, but are not tested across the full operating range of the equipment.	HP is operated under the same control as operated in the building (native controls). Tests are conducted with generally equivalent equipment and instrumentation as in Category A standards. Repeatability, reproducibility and representativeness studies are ongoing. However, evidence of the level of repeatability and reproducibility similar to Category A standards have been presented in recent literature. Recent studies of Representativeness suggest an improvement over Category A standards (i.e., that seasonal performance results using Category B standards are more representative of in-field test results than those produced by Category A standards) Note that Category B standards testing may require specialized test apparatus and/or capabilities (while instrumentation remains generally equivalent).
(2) Seasonal or annual average efficiencies	Necessary for regulating the energy efficiency level of each product category. The choice of test conditions required to extract these seasonal performance indexes is based on: -Assumption of the relationship between heat needs imposed on HP and the maximum capacity of the HP: fixed ratios are applied, such as 1.0 for cooling in JIS C 9612. -Assumption on the relationship between the heat needs and outdoor temperature: a linear relationship is assumed. *(for Category B standards) Assumption of reference building thermal and moisture characteristics are required and affect the time dependent system response captured.	

1.2 Current testing methodologies and performance rating standards for heat pump systems (Category A Standards)

Current performance rating standards are reviewed for the following aspects:

1) Targeted heat pump systems and the scope,
2) Test methods,
3) Temperature conditions,
4) Control of test specimens during tests,
5) Performance indices and requirements for part load tests,
6) Tolerance of measurement uncertainty,
7) Other issues.

The current standards reviewed are listed in Table 1.2-1.

Table 1.2-1. List of the current testing and rating standards reviewed

No.	Title of standard	Year
1	ISO 5151. Non-ducted air conditioners and heat pumps – Testing and rating for performance	2017
2	ISO 13253. Ducted air-conditioners and air-to-air heat pumps – Testing and rating for performance	2017
3	ISO 15042. Multiple split-system air-conditioners and air-to-air heat pumps – Testing and rating for performance	2017
4	ISO 16358. Air-cooled air conditioners and air-to-air heat pumps – Testing and calculating methods for seasonal performance factors – Part 1: Cooling seasonal performance factor, Part 2: Heating seasonal performance factor, Part 3: Annual performance factor	2013
5	EN 14511-1, 2, 3. Air conditioners, liquid chilling packages and heat pumps for space heating and cooling and process chillers, with electrically driven compressors	2022
6	EN 14825. Air conditioners, liquid chilling packages and heat pumps, with electrically driven compressors, for space heating and cooling - Testing and rating at part load conditions and calculation of seasonal performance	2022
7	AHRI 210/240. Performance Rating of Unitary Air-conditioning & Air-source Heat Pump Equipment	2020
8	AHRI 340/360. Performance Rating of Commercial and Industrial Unitary Air-conditioning and Heat Pump Equipment	2022
9	AHRI 310/380. CSA-C744-17. Packaged Terminal Air-conditioners and Heat Pumps	2017
10	AHRI 550/590. Performance Rating of Water-chilling and Heat Pump Water-heating Packages Using the Vapor Compression Cycle	2023
11	AHRI 1230. Performance Rating of Variable Refrigerant Flow (VRF) Multi-Split Air-conditioning and Heat Pump Equipment	2023
12	ANSI/ASHRAE Standard 37-2009 (RA 2019). Methods of testing for rating electrically driven unitary air-conditioning and heat-pump equipment	2019
13	ANSI/ASHRAE 206-2013 (R2017). Method of Testing for Rating of Multipurpose Heat Pumps for Residential Space Conditioning and Water Heating	2017
14	JIS B 8616. Package Air Conditioners	2015
15	JIS B 8627. Gas Engine Driven Heat Pump Air Conditioners	2015

1.2.1 Targeted heat pump systems and scope

Targeted heat pump systems can be categorised according to a) heat source (air or water) and secondary medium for heating and cooling supply to emitters, b) configuration of the heat pump systems, e.g., ‘packaged’, ‘unitary’, ‘multi-split’, ‘ducted’, ‘non-ducted’, c) drive of the compressor (e.g., electrically-driven, gas engine driven), and d) capacity (e.g., 19 kW or greater).

1.2.2 Test methods

This aspect is well categorised in ANSI/ASHRAE Standard 37(ANSI/ASHRAE, 2019) and ISO 5151 (ISO, 2017a). According to the former standard, there are five test methods:

a. indoor air enthalpy method,
b. outdoor air-enthalpy method,
c. compressor calibration method,
d. refrigerant enthalpy method,
e. outdoor liquid coil method.

For the indoor and outdoor enthalpy methods, only the nozzle airflow measuring apparatus, which needs a tunnel (duct) to allow rectifying the airflow before the nozzle and to compensate the pressure loss due to the nozzle and other parts of the tunnel by using a fan, is specified. The tunnel is also used for measuring dry-bulb and wet-bulb temperatures of well-mixed air from the unit.

Besides the five test methods above mentioned, ISO 5151 prescribes calorimeter test methods. According to ISO 5151, capacity tests shall be conducted using either the calorimeter test method or the indoor air enthalpy test method.

1.2.3 Control of test specimens during tests

Most current standards only deal with stable conditions for test specimens. The necessity of manufacture instructions to achieve the stable condition is clearly prescribed by standards. It is well recognised that the intermittent operation of test specimens reduces their energy efficiency compared with continuous and stable operation, and the difference is represented in plural standards by the degradation coefficient (C_D). For the test to quantify the C_D, the cycle test is prescribed by some standards, besides the test under stable conditions. In most standards specifying seasonal average efficiencies, a default value of the C_D is specified, such as 0.25.

In JIS B 8616 (JIS, 2015a), JIS B 8627 (JIS, 2015b), and AHRI 1230 (AHRI,2010), a control verification procedure has been added to verify that the minimum compressor speed for the part-load test can occur without overriding control settings. However, in current testing standards for heat pump systems, including those three standards, overriding control of the specimen during the tests is officially permitted.

1.2.4 Performance indices and requirements for part load tests

For the energy performance rating, EER (Energy Efficiency Ratio) and COP (Coefficient of Performance) are the common basic indices. The unit of the indices varies, but the meaning of the indices does not change, namely the ratio of the capacity to the input energy.

In various standards, integrated indices are provided, of which roles represent seasonal average energy efficiencies. For that purpose, the measurement of energy efficiencies under part load conditions by the tests or the calculation of the energy efficiencies under the part load conditions by using measured values such as for full load and conversion factors is specified in those standards. Instructions and support by manufacturers are necessary to achieve stable operation in part load conditions. As for the part load conditions, in some standards, tests for 75%, 50%, and 25% of full capacity are required, while estimating energy efficiency under lower part load conditions is done using the C_D.

1.2.5 Tolerance of measurement uncertainty

Because of the accuracy limit of measurement devices, standards specify tolerances. Among parameters, measuring airflow rate and temperature may be the most difficult, partly because of their spatial distribution, even if the specimens are in steady operations. If the measurement is conducted under possibly unsteady operation, such as in load-based tests, measurement uncertainty probably becomes larger, and the requirement for the tolerance of measurement uncertainty should be an important issue for their standardization.

1.2.6 Other issues

To express capacities of the specimen under tests, various terminologies are used. They include ‘full capacity’, ‘rated capacity’, ‘nominal capacity’, ‘extended capacity’, etc. Sometimes, the full capacity is the same as the rated capacity. In the definition of part-load ratio, the ratio´s denominator can be different, and the definition resultantly becomes unclear. When analysing energy performance under the part load condition, defining the part load ratio by using the maximum capacity at a certain temperature condition is best. However, there is still an open question on the issue.

1.3 New type testing standards for heat pump systems (Category B Standards)

1.3.1 Overview of load-based testing methodologies

The actual operation of variable-speed heat pumps and air conditioning units may respond with variable or cyclic modulations of the compressor speed and expansion valve opening even to constant thermal loads according to their native control system. Therefore, the corresponding field efficiency of the system may be significantly affected by the control strategy developed and implemented in operating units. Contrary to steady-state operation, the dynamic operation of the system involves a time-dependent thermal interaction between the building thermal characteristics and the capacity supplied by the unit, whereby cooling/heating capacity and the building load are not necessarily and continuously balanced. The magnitude of the unbalance drives a variation of the room temperature and, for a unbalance, the rate of change is related to the equivalent heat capacity of the room. Similar observations apply to the moisture balance, which defines the response of the room condition to a given latent load scenario. Therefore, when dynamic operation is accounted for, the test conditions and the building structural features affect the room thermal response and, in turn, the air conditioner/heat pump performance (Mehrfeld, 2022).

Conventional lab tests for residential heat pumps and air conditioners (such as AHRI 210/240 (AHRI, 2023), JIS B 8615 (JIS, 2015c), or EN 14511 (BSI, 2018)) use fixed compressor speed and expansion valve opening conditions (and hereinafter will be referred to as “fixed condition” tests). In load-based tests, the tested unit is installed following the manufacturer’s instructions as it would be done by a qualified field technician, and during the test the system meets heating and cooling loads that are typical for residential applications, using its own thermostat and internal control logic to respond to changes in the room temperature, in case of air-to-air units, or the water inlet temperature in the case of hydronic heat pumps. In this way, the lab environment during the test process emulates a real-life installation, while allowing for consistent control and measurement so that each test can be consistent in its results and provide fair performance comparisons between different models. In the following paragraphs, the principle of a load-based test is explained using the example of air-to-air units. However, the same concept applies to hydronic heat pumps using air, brine or water on the source side and water on the sink side.

In the load-based test, as in a fixed-condition rating test, the process is conducted using two psychrometric chambers, with one of these chambers called the “outdoor room” where the outdoor unit is placed, with carefully controlled temperature and humidity that represent the various outdoor conditions at which the unit is tested. The laboratory setup of the outdoor room uses reconditioning equipment controlled by computer software to maintain those conditions for the duration of a test condition, before moving on to the next condition.

In both load-based and fixed-compressor speed testing, the second psychrometric chamber, called the “indoor room” is where the indoor unit is installed. Understanding the different control strategies for the indoor room is the key to understanding the load-based test. In a fixed-condition lab test, the tested unit will run in a steady-state mode that is defined by the particular test condition and typically uses a proprietary “test mode” that overrides the unit’s normal control sequences. The indoor room reconditioning system maintains the indoor room temperature and humidity in a steady-state manner for the duration of the test. The computer software controlling the test measures how much heat the tested unit is producing (in heating mode) or removing (in cooling mode), as well as the energy input and other key parameters (such as air flow).

By comparison, in a load-based test, the indoor room condition mimics (or “emulates”) the condition of a room or space that would be heated (and cooled) by the tested unit in response to a heating or cooling load. The loads are carefully chosen to represent a typical house or indoor space, based on the rated heating or cooling capacity (the size) of the heat pump. The lab software controlling the reconditioning equipment is programmed with the indoor room “load” to be imposed, and it continuously senses the amount of heat the tested unit delivers (or removes) from the indoor test room. Based on these values, it updates the actual indoor room temperature every few seconds to simulate an actual load. The tested unit then responds to changes in the indoor room temperature by turning on or off, or changing its output to match the load, according to its own internal logic (using the same control logic it would use in a typical field installation). This is best understood graphically, as follows:

- Figure 1.3.1-1 shows a simplified example of what would happen in the indoor room during a virtual heating load if the tested unit was not running. The test control software senses the output of the unit, and it causes the room to cool off. In this theoretical example, it loses 50 °F over an hour’s time.
- In Figure 1.3.1-2, imagine that the tested unit is continuously producing half of the needed heat. The room temperature drops at half the rate of that in Figure 1.3.1-1, losing only 25 °F in an hour. (In reality, the temperature drop is not a straight line, but it is simplified here for illustrative purposes).
- Figure 1.3.1-3 shows the temperature of the indoor room if the tested unit continuously generates exactly the amount of heat needed to keep up with the simulated heating load: the temperature stays constant throughout. The controls of a variable-speed heat pump should operate this way, if the virtual load is within the range at which that the unit can operate (i.e., between its maximum and minimum capacity), at the outdoor temperature condition in the outdoor room; but a small amount of variation in indoor temperature will always occur in order for the unit to respond accordingly.

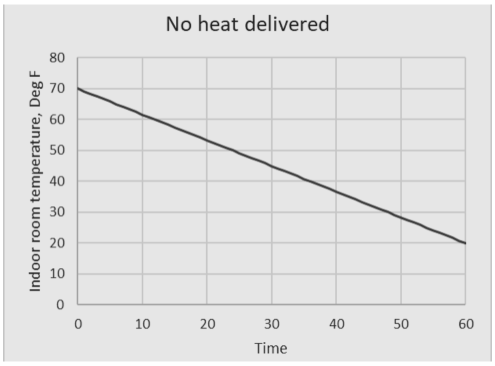


Figure 1.3.1-1. No heat added (Time in minutes)

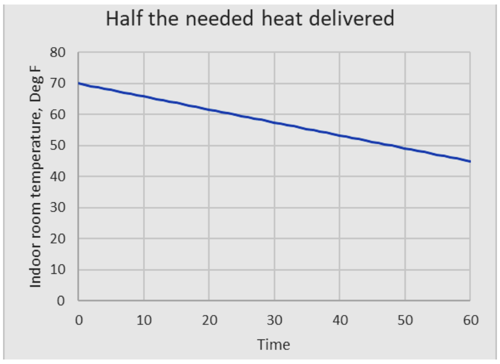


Figure 1.3.1-2. Half the needed heat, temperature drops more slowly (Time in minutes)

Figure 1.3.1-4 is more typical of a real modulation of a tested unit. Imagine at time = 0, the thermostat is turned on at 70 °F, just as the indoor temperature begins at 70 °F, simulating a heating condition in cold weather. As the unit comes on and produces more heat than is needed, the room temperature will increase based on the test control programming. Then, at some point, the internal controls of the tested unit sense that the room is too warm, at which point it will reduce its output (in this example, at minute 8). The lab control software senses the unit having reduced output, and causes the room temperature to drop again, as it would under a real heating load. At some point (in this case, below 69 °F) the internal controls of the tested unit turn the unit back to a higher heating output (in this example, at minute 10), and the cycle continues.

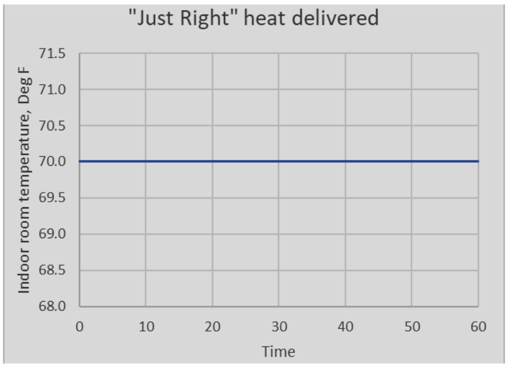


Figure 1.3.1-3. Heat added is correct, stable temperature (Time in minutes)

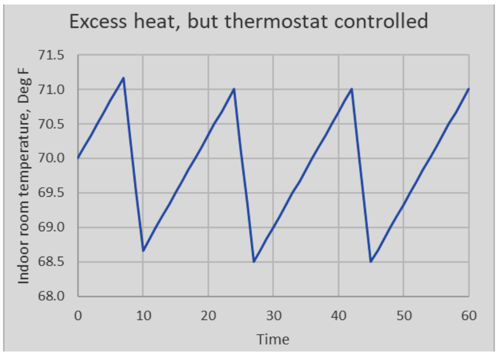


Figure 1.3.1-4. Heat modulates, temperature controlled by thermostat of tested unit (Time in minutes)

Thus, the tested unit is responding to an indoor condition that simulates a heating (or cooling) load that would be found in a house or room that is exposed to the same conditions as the outdoor room where the outdoor unit is located. Although the lab control software and reconditioning equipment is literally controlling the indoor room temperature, the temperature is based on the response of the tested unit just as if it was in a space that was heating up or cooling off in response to a real load.

In an actual test, the behaviour of the lab and the system being tested is, of course, more complicated than what is shown in Figures 1.3.1-4. In some cases, variable speed systems can match the heating or cooling requirement closely (such as in Figure 1.3.1-3) with natural variations based on the unit’s internal controls, as it responds to the virtual load.

However, variable speed systems cannot ramp “down” continuously, all the way to “off”; they always have a lower limit of heating or cooling output. When the load is smaller than that minimum, the unit will have to cycle on and off, which affects the operating efficiency. In the highest load conditions (when outdoor temperatures are also the most extreme), it is expected that tested units will typically lack the heating or cooling output needed to maintain the steady state indoor temperature target. For those test conditions, the unit is set instead to run at full capacity (but still under its normal controls), and the rest of the test is completed while the reconditioning equipment keeps the indoor room under steady state conditions.

In each test condition, the lab software collects data to verify both the heating or cooling output of the tested unit in real time, and to measure the electricity input (power). Sometimes this “steady state” operation over time occurs naturally (as typified by Figure 1.3.1-3). However, during the load-based test, the indoor room conditions vary over time, causing the heating or cooling output to also vary in more complex ways. This may be due to the need to cycle off during low-load conditions because the variable-speed controls are “searching” for the right output to best match the load; the need for defrost cycles in some heating conditions; or for other reasons dictated by the internal control logic of the tested unit.

These general considerations exemplify how the control system and its interaction with the building features and thermal loads define a broad spectrum of possible operating performance.

Load-based tests respond to the necessity of capturing the main characteristics of actual operating performance during laboratory tests while minimising additional effort and cost when compared to current standards.

Newly proposed testing methodologies (Category B) aim to reflect the following aspects:

- Unit performance when operated under its native control and using its own thermostat.
- Characterise efficiency losses or gains of variable speed units (inefficiency of cycling operation and assess the efficiency of the control method).
- Integrate all cycles within a test bin such that defrost cycles, on/off cycles, etc. are directly measured within each temperature bin.
- Capture the interaction of the system operation with the actual load scenario and the thermal features of a representative building.
- Prevent the manufacturer from artificially inflating the unit efficiency during performance rating tests.

The characterisation of these aspects should drive positive developments in the design of efficient control strategies for variable speed units and maximise efficiency during the field operation of heat pump installations.

The proposals developed by 4 independent institutes are reviewed with reference to the following aspects:
- Scope of the test. Including target equipment type and capacity.
- Test conditions.
- Building-side thermal emulation method.
- Analysis of repeatability, reproducibility, and representativeness (3Rs).

Table 1.3-1 provides a first summary of the testing methodologies.

Table 1.3-1. Summary of the reviewed test procedures for the development of Category B Standards

Test method (institution)	Test scope	Heating conditions	Cooling conditions	Building thermal inertia	3Rs analysis
Waseda University	Emulator-type load-based test for air-to-air units	2 tests defined consistently with JIS B 8515 for heating operation *partial-load at 25% of max capacity **(tentative)	3 tests defined consistently with JIS B 8515 for heating operation *partial-load at 25% of max capacity **(tentative)	Defined within the lumped parameter emulator by the values of thermal and moisture inertia	Repeatability (completed) Reproducibility (Cooling tests completed, Heating tests ongoing) Representativeness (ongoing)
CSA	SPE-07:23 load-based and climate-specific test for air-to-air units (using emulator)	5 temperatures (-15 to 12.2C) plus one additional test for marine climate zone as well as optional test at lowest operating temp	4 temperatures (25 to 40C) plus one additional test for hot, dry climate zone	Simulated thermal capacitance (sensible and latent) of building interior included in load calculation	Repeatability (completed) Reproducibility (ongoing) Representativeness (completed)
BRI / Better Living	Load-based test for VRF air-to-air units	OC: 7C (DBT) 6C (WBT) IC: 20C (DBT) 15C (WBT)	OC: 35C (DBT) 24C (WBT) IC: 27C (DBT) 19C (WBT)	Artificial thermal capacitance (sensible and latent)	Repeatability (ongoing) Reproducibility (ongoing) Representativeness (ongoing)
BAM and RWTH	Load-based test for hydronic heat pumps	5 or 6 outdoor temperatures in accordance with EN 14825:2022	Not applied yet (ongoing)	Defined within a simplified building model	Repeatability (completed) Reproducibility (ongoing) Representativeness (ongoing)
RWTH	Hardware in the Loop (HiL) for building energy systems with hydronic heat pumps	Outdoor conditions defined by weather data. Use reference days (~4 days) representing a whole year for a specific geographical location	See heating conditions. Depending on location, some days have cooling demand	Simulated by detailed Modelica model of a specific building and transfer system to be studied	Repeatability (completed) Reproducibility (completed) Representativeness (ongoing)

1.3.2 Emulator-type load-based testing method for air conditioners by Waseda University

This section describes a testing method for residential and commercial air conditioners that can reproducibly assess the energy efficiency of variable speed units and characterise their controllability when operated according to their native control under load scenarios representative of in-field installations. The proposed method essentially relies on a standard air-enthalpy testing facility used for more conventional testing and does not require additional instrumentation and testing time requirements, but only the bidirectional interconnection of a simple simulation software, which acquire the real-time measurement of the supplied capacity from the instrumentation of the tested unit, and also controls the reconditioning unit of the indoor psychrometric room.

1.3.2.1 Overview

The research efforts of Waseda University in the development and evaluation of optimal control strategies for air-to-air vapor compression systems led to a first national project conducted between 2014-2016 for the development of a new testing method able to reproducibly capture the control response of variable speed drive units and correspondingly assess their performance.

This pioneering project resulted in the design, construction, and operation of a first prototype (Ban et al., 2016, 2017). These preliminary results were critical for recognising the main challenges related to the hardware and instrumentation of current air-enthalpy testing facilities in the real-time measurement and dynamic control of the reconditioning unit during dynamic system operation. The necessity of high-accuracy instrumentation and appropriate controllability of the reconditioning unit was recognised and dealt with for developing the testing method reported in Giannetti et al. (Gianetti et al., 2022a,b).

1.3.2.2 Conceptual description of the testing method

As reviewed in Section 1.2, current rating standards rely on forcibly achieved steady-state tests where the native control of the system is deactivated. To assess the energy efficiency of variable speed units when operated according to their native control under load scenarios representative of in-field installations, the emulator-type load-based testing method combines numerical software (room emulator) with the hardware of a conventional testing facility used for category A standards. The software and hardware are interfaced through the reconditioning unit (or “condition generator”), which recreates the modulations of the room conditions as calculated by the emulator, and the “measuring chamber”, which feeds real-time measurements of the cooling capacity supplied by the indoor unit as a digital signal to the emulator, as described in Clause 1.3.1. A schematic representation of this testing method is illustrated in Figure 1.3.2-1.

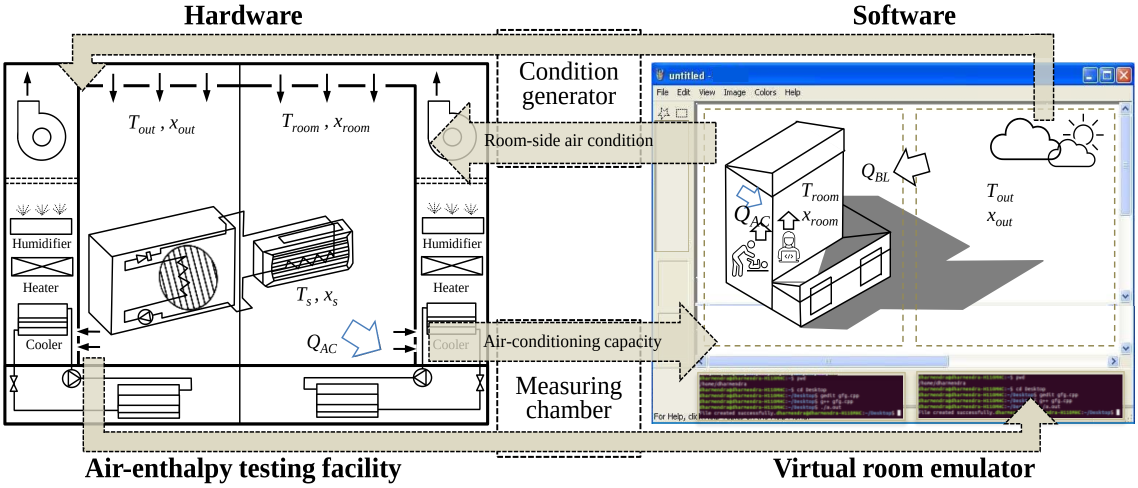


Figure 1.3.2-1. representation of the emulator-type load-based testing methodology (Giannetti et al, 2024)

1.3.2.3 How the tested system is operated

As explained in Section 1.3.1, load-based emulator-type tests are conducted while installing indoor and outdoor units in two separate psychrometric chambers and allowing the system to operate in accordance with its native control. The emulator software calculates the modulations of the indoor air condition while accounting for the dynamic response of the tested system. Complementarily, the reconditioning unit of the psychrometric chamber is controlled to replicate such numerical results in terms of temperature and humidity of the return air to the indoor unit. The system attempts achieving the indoor set temperature for the simulated load scenario and may experience indoor temperature and humidity modulations of the return air to the indoor unit due to variable-speed or on/off cycling operation.

In practice, the use of the emulator software can dynamically generate reproducible testing conditions by controlling the reconditioning unit to make the test independent of the specific thermal features of the testing facility. Meanwhile, temperature and humidity conditions in the outdoor psychrometric room are held constant.

Given the dynamic characteristics of emulator-type load-based tests, a preliminary investigation of the factors affecting measurement error and delay, such as the computational time delay of the emulator, trackability of temperature and humidity in the condition generator and in the measuring chamber, and time delay of the sensors, was carried out (Giannetti et al., 2022b) and represents the basis for reliably defining the level of reproducibility of such tests. Additionally, to minimise the loss of information between software and hardware sections, a tuneable feed-forward compensation module (FFC) was developed using a transfer function system identification approach (Figure 1.3.2-2). This software module may be used to restrain the delay in the reconditioning within the allowable range for enhancing the reproducibility of the test results across different testing facilities (Giannetti et al., 2024).

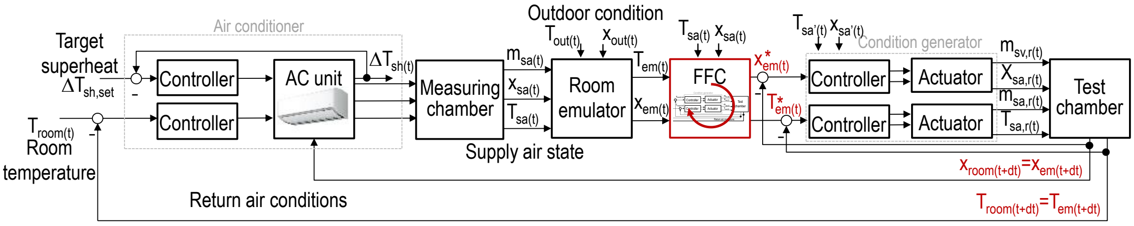
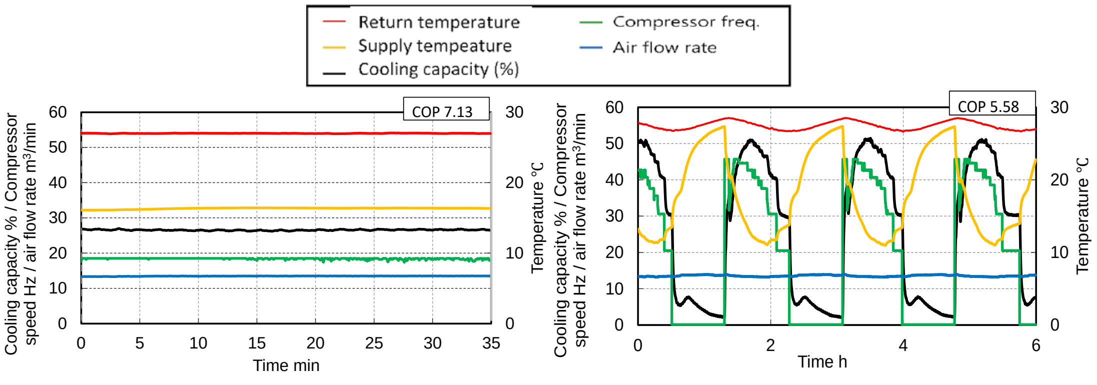


Figure 1.3.2-2. Schematic block-diagram of the emulator-type load-based testing methodology (Giannetti et al, 2024)

1.3.2.4 Illustrative test results

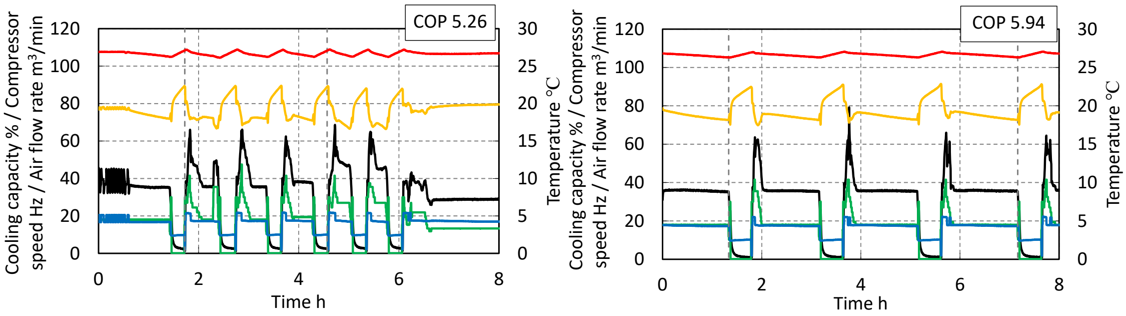
Tests were conducted at corresponding ambient and partial load conditions with the same unit operating according to the current JIS standard [JIS B 8616 (2015a)] and with the emulator-type load-based testing method to characterise the gap between actual system performance and performance recorded with current testing standards. Fig. 1.3.2.3 exemplifies the results obtained at a partial load ratio of 25%. In this case, the air conditioner functions in a cyclic on-off operation when operated with its own native control and exhibits a COP of 5.58, while the fixed-compressor-speed test indicates a COP of 7.13.

Consequently, when testing the system with the emulator-type load-based testing method and setting the building load above 50% of the rated system capacity, the native control could achieve steady-state operation for a virtual room size of 147 m^3. However, minor dynamic modulations of the compressor speed were observed because of oil recirculation manoeuvres (Miyaoka et al., 2023). Conversely, under lower building load conditions, the system responded with on-off cyclic and variable-speed operations. Figures 1.3.2-4 (i)-(ii) illustrate the operation encountered when the building load was set to 30%. The lumped heat capacity of the virtual room was changed according to the size of the room (Togasi&Tanabe, 2009), and the on-off cycling operation of the air conditioner showed different cycling intervals. These results provided evidence for the significance of the building thermal inertia on the system controllability and corresponding performance, and for the necessity of a virtual room emulator for fairly assessing the dynamic performance of air conditioners while equivalently reproducing the “room-side air condition” in different testing facilities. Smaller size rooms correspond to faster cycling and larger efficiency losses, while a larger room thermal inertia allows the control system to operate with longer cycling and reach pseudo-steady operating intervals with lower cycling losses.



(i) (ii)

Fig. 1.3.2-3. 25% Partial load performance (i) JIS standard test (ii) emulator-type load-based test (Giannetti et al, 2022b)



(i) 30%, 147 m^3 (ii) 30%, 300 m^3

Figure 1.3.2-4. 30% Partial load tests with (i) 147 m^3 (ii) 300 m^3 virtual room size (Giannetti et al, 2022b)

1.3.2.5 Selection of set of test conditions

Full load and part load operating conditions presently refer to JIS B 8615 (JIS, 2015c); including 3 operating points for cooling and 2 for heating operation. Adjustments of the selected tested conditions are presently under consideration to capture cycling operation (part-load condition at 25% of maximum capacity) and minimise extrapolation in energy calculation procedures, effectively capture control characteristics, and harmonise test requirements along with test condition for maximising comparability and minimising testing burden. Additionally, pre-defined continuous load patterns are under consideration for test automatisation.

1.3.2.6 Assessment of repeatability, reproducibility, and representativeness of the test results

Evidence for repeatability and reproducibility properties of the emulator-type load-based tests are essential for defining new standards for performance ratings. Dedicated investigations with multiple tests repeated within the same testing facility (Miyaoka et al., 2023) and expanded to four different testing facilities (Dondini et al, 2024), demonstrated results repeatability within 1.5% and reproducibility within 3% standard deviation, respectively.

Table 1.3.2-1. Test conditions of round robin tests from Dondini et al. (Dondini et al, 2024).

Conditions	Indoor dry-bulb temp. (°C)	Outdoor dry- bulb temp. (°C)	Outdoor wet- bulb temp. (°C)	Load Ratio (%)	Simulated room size (m^3)
Low load virtual room 1	27	29	19	25	147
Low load virtual room 2	27	29	19	25	75
Mid load virtual room 1	27	29	19	50	147

Specifically, the performance and control response (such as those illustrated in Figure 1.3.2-5) of a 10-kW R32 ceiling-type unit, operated in cooling mode within the four facilities at the test conditions reported in Table 1.3-1, were analysed to provide a first assessment of the level of reproducibility of the proposed testing method and suggest challenges and possible improvements. The results from all testing facilities demonstrated consistent performance and control responses (as summarised in Table 1.3.2-2).

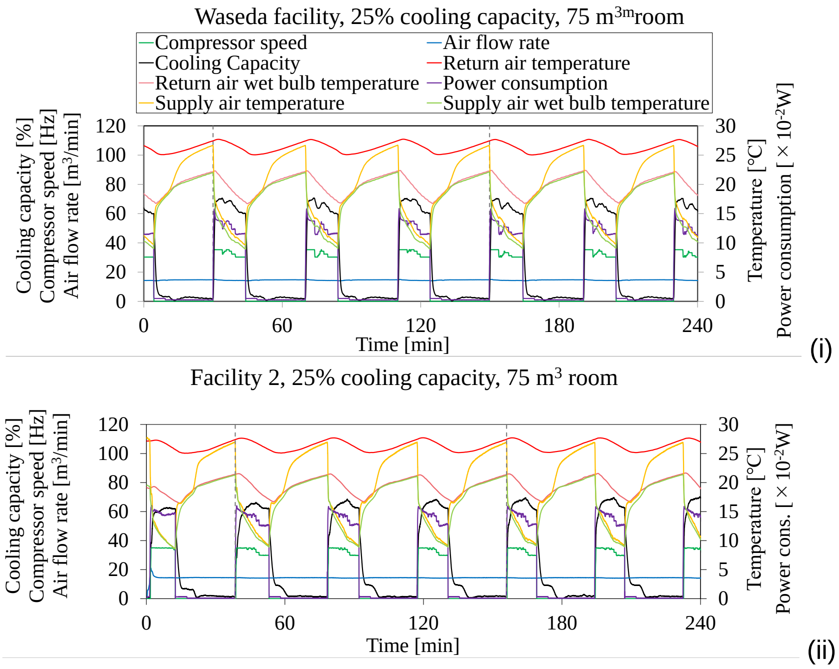


Figure 1.3.2-5. Test results for “Low load virtual room 2” at 25% load, 27°C indoor set temperature, and 75 m^3 for: (i) Waseda, (ii) Facility 2 (Dondni et al, 2024)

Table 1.3.2-2. Summary of Round robin test results (Dondini et al, 2024)

Conditions	COP Waseda	COP Facility 2	COP Facility 3	COP Facility 4	Deviation from average
Low load virtual room 1	5.34	5.57	5.39	5.33	3.01 %
Low load virtual room 2	5.37	5.22	5.23	5.30	1.70 %
Mid load virtual room 1	6.24	6.10	6.04	6.03	2.25 %

1.3.2.7 Definition of seasonal or annual performance indices, and system performance metrics

Seasonal efficiency calculation presently refers to JIS C 9612 (JIS, 2013), which combines the hourly distribution of ambient temperature, regional heating and cooling loads to calculate the APF index. As emulator-type load-based tests characterise the system performance and controllability when operated according to their native control, such seasonal index may provide closer representations of the actual field performance of air conditioners and may drive virtuous developments of efficient control, as well as a method to verify control strategy improvements. Additionally, the performance characterisation extracted through this testing method is being used to construct performance curves and maps for performing seasonal energy calculations.

1.3.3 CSA SPE-07:23 Load-based and climate-specific testing and rating procedures for heat pumps and air conditioners

1.3.3.1 Background

In 2015, the Canadian Standards Association (CSA) began work on a test and rating procedure that would better represent installed performance of variable capacity heat pumps (VCHPs) in a range of climates. In 2019, the Canadian Standards Association (CSA) published a technical review version called EXP-07:19*, Load-based and climate-specific testing and rating procedures for heat pumps and air conditioners*, (referred to as EXP07).[9] After conducting numerous additional lab tests using EXP07 and soliciting public comments, a final revision was made and published as SPE-07:23 (CSA, 2023a) using the same title (hereafter SPE07). SPE07 uses load-based tests at a range of conditions of both heating and cooling operation in order to create a performance profile, which is then used to calculate a set of Seasonal Coefficient of Performance (SCOP) values. These SCOPs are reported separately for heating and cooling for seven different North American climates and represent an estimate of net seasonal efficiency of heat pumps in typical residential applications for each of those climates.

The scope of SPE07 applies to residential, single-zone air-to-air heat pumps and air conditioners less than 65k Btu/h (19 kW) in capacity.

This section provides an overview of how SPE07 works and generally explains the concepts behind the SPE07 rating procedure. It is adapted from the *EXP07 Plain Language Guide* (CSA, 2023b).

1.3.3.2 Load-based Testing

Most fundamental to SPE07 is its approach to testing using a "virtual" or simulated building load, managed by test room system software (sometimes referred to as an "emulator"). As described in Section 1.3.1, SPE07 uses a dynamically-controlled, load-based approach that measures heat pump performance across a wide range of outdoor temperatures, while the system meets heating and cooling loads that are typical for residential applications, using its own thermostat and internal control logic to respond to changes in the room temperature.

The approach taken in SPE07 is very similar to that outlined in Clauses 1.3.2 and 1.3.4, and to a lesser extent 1.3.5 (although all four methods have the same intent to emulate operation of the heat pump under its native control system rather than operation in a special test mode). However, SPE07 differs from the others because it is a published test method that includes both test procedures and performance rating calculations.

To account for the natural variation in the unit operation, the test procedure includes detailed instructions so that the lab can determine at what point during a particular test condition the test may be considered “complete”. This process is defined by a set of rules that require monitoring the heating or cooling output and electric input over time, searching for a during which these measurements are steady, or repeating over time in such a way that additional measurements will not likely change the result in a meaningful way. This is referred to as “convergence”, and once convergence (or a test period time limit) is reached, the test condition is considered complete and the test procedure moves to the next condition, until all tests are completed.

1.3.3.3 The Load Line

During SPE-07 lab testing, two series of heating tests (Continental and Marine), and two series of cooling tests (Humid and Dry) are conducted (as described below). Each climate-based rating is derived from those test results using the appropriate set of heating and cooling tests, mapped into that climate data. For heating, SPE-07 uses a single, linear relationship between outdoor temperature and load, based on the rated capacity of the tested equipment (referred to as a “load line”). The load line is a typical generalised building load profile, and the concept is common to other heat pump rating systems such as AHRI 210/240 (AHRI, 2023) and CSA C656 (CSA, 2014). A single load line is used for the heating load in SPE-07 (CSA, 2023a); this is based on the rated cooling capacity at 95F. The load increases with decreasing outdoor temperature, and the no-load point (intersection with the x-axis) occurs at 60F by definition. The Marine climate zone heating test conditions vary only by the outdoor unit humidity that is used. For cooling, separate load lines are defined for dry and humid conditions, which are then used to generate the cooling SCOP ratings.

Each chosen load line implies an assumed relationship between the size of the equipment and the magnitude of the building load – that is, an implied “equipment sizing”. Defining the “right” load line is a challenge because home efficiency levels vary dramatically, and the relative sizes of a home’s heating and cooling design loads can vary significantly. Even within a given building, loads can vary significantly off of the “average” load line due to transient events, such as changes in solar gain. The SPE-07 (CSA, 2023a)heating load line is used for all the heating climate seasonal ratings. Even though heating design temperatures vary significantly from mild to cold climates, the chosen load line is a compromise that reduces the number of required lab tests while remaining broadly relevant across a range of climates. It generally results in the testing of heat pumps under the full range of operating modes, including cycling, modulating, and full-load, which is an objective of the test procedure.

An analysis of alternative load lines to that used in SPE-07 (CSA, 2023a) was conducted and it concluded that the SPE-07 load line remained robust under a variety of circumstances, except for the extreme case of sizing a heat pump for full-load heating in the Subarctic and Very Cold climate zones. In this case, it was suggested that some additional metric such as cold-ambient capacity maintenance would also be required – especially if the objective is to reduce reliance on auxiliary heat sources (and such a metric has indeed been used in incentive and manufacturer challenge programs in Canada and the United States[2].

Learning Test Cycle

Before proceeding with the cooling and heating rating test series, a learning test series is conducted. The learning test series allows the equipment to run under its own controls and acts as a “break-in” period.

1.3.3.4 Test Conditions

The SPE07 test procedure uses 6 heating conditions and 10 cooling conditions. The tests are run at each condition until the system achieves convergence, as outlined above. At each outdoor temperature, the amount of heating or cooling load that is dynamically simulated in the indoor room (see “the load line” above) is appropriate for the outdoor temperature at which the equipment is tested and is also scaled to the capacity of the tested unit, so that each unit is tested based on its rated capacity.

The heating conditions are divided into two general climate areas, Continental and Marine, each with its own sequence of outdoor temperatures and corresponding loads. The cooling test conditions are divided into humid and dry climate areas, each with its own sequence. In addition, in the humid cooling tests, a dynamic moisture load is applied by monitoring the removal of humidity by the equipment under test, and then updating the indoor humidity in the test room programming. This works in very much the same way that the dynamic heating and cooling loads are applied to indoor temperature for all the tests, and it allows the test to measure how well the units remove moisture in the humid cooling tests. (By contrast, in a conventional test the reconditioning equipment maintains a constant humidity level in the indoor room). Table 1.3.3-1 summarises the four test sequences:

---

[2] In Canada, the Canada Greener Homes Grant Initiative, as a condition of eligibility for it’s cold climate heat pump grant amount, required an equipment capacity maintenance (Max -15 ° C (5 ° F)/Rated 8.3 ° C (47 ° F)) ≥ 70% (with a COP ≥ 1.8). In the US Department of Energy Cold Climate Heat Pump Challenge, the performance requirement at 5F (-15C) was to maintain 100% of the nominal capacity of the system as tested at the AHRI 210/240 Appendix M1 A2 test point for heating/cooling heat pumps (with a COP ≥ 2.4 for equipment up to 4 tons and ≥ 2.1 for equipment > 4 tons.

Table 1.3.3-1. Summary of the four test sequences.

Heating	**Outdoor Conditions**	**Indoor Conditions**
Continental	5 temperatures from **5 to 54 °F (-15 to 12.2 °C),** plus optional test at lowest operating temp, per manufacturer	**70 °F (21.1 °C)** **56% RH max**
Marine	One additional at **34°F (1.1°C)**	
Cooling	**Outdoor Conditions**	**Indoor Conditions**
Dry	5 from **77 to 113 °F (25 to 45 °C)**	**79 °F (26.1 °C) 21% RH max**
Humid	4 from **77 to 104 °F (25 to 40 °C)**	**74 °F (23.3 °C) 55% RH max**

Wherever possible, test procedures, such as measurement techniques, are harmonised with AHRI 210/240. Although the indoor unit air flows during SPE07 tests may vary based on the internal controls of the tested unit, the initial setup to define and measure full-load air flows, and to establish static pressures for ducted systems, are harmonised with conventional test methods.

1.3.3.5 Efficiency Metrics

Once the test results have been measured and recorded, seasonal efficiency values are calculated. The result is a heating and a cooling Seasonal Coefficient of Performance (SCOP) for each climate zone – SCOPh and SCOPc. (Except that there is no cooling SCOP for the Subarctic zone.) The basic method to calculate seasonal efficiencies is called a bin model, consistent with other rating and common HVAC engineering analyses. For each climate, the analysis uses a specific number of hours that represent the number of heating or cooling hours at each temperature “bin” throughout the heating and cooling seasons. The temperature "bins" are divided into increments of 5 °F (2.8 °C), and the unit's heating or cooling efficiency, as determined in the lab, is applied to each bin based on the number of hours within that bin. The size of the heating and cooling loads used for the rating calculation are the same as those used during the tests.

For heating, at any outdoor temperatures for which the tested unit does not have enough heating capacity to meet the full heating load, it is assumed that the difference is made up with electric resistance supplemental heaters with a COP of 1.

For each climate, the total delivered output for the season is divided by the total electrical input to determine the Seasonal Coefficient of Performance (SCOP) for that unit in that climate. The SCOP is a simple ratio, so a COP of 1.0 represents 100% efficiency (such as electric resistance heat). Heating SCOPs are generally higher in warmer climates and lower in colder climates, and cooling SCOPs are generally lower in the hottest climates and increase as summer climates get cooler. The eight representative climate zones are shown in Figure 1.3.3-1.

There is a provision that the lab tests the energy input during “standby” modes of operation (when the unit is not heating or cooling), as a separate procedure. The results are used in the analysis for seasonal COP, which may be reported separately with and without the standby power. The standby power is added for hours (based on each climate), during heating or cooling seasons, for temperatures at which there is no heating or cooling requirement but when the HVAC system unit thermostat is likely to remain in “heat” or “cool” mode. Also, standby power is applied to shoulder periods when there is no heating or cooling demand, and the unit controls are likely to be turned “off,” but the system is still powered on at the circuit panel. Standby energy makes a more significant impact on annual efficiency ratings in climates with long shoulder periods that require no heating or cooling, and of course, for equipment that has higher standby electric energy input.

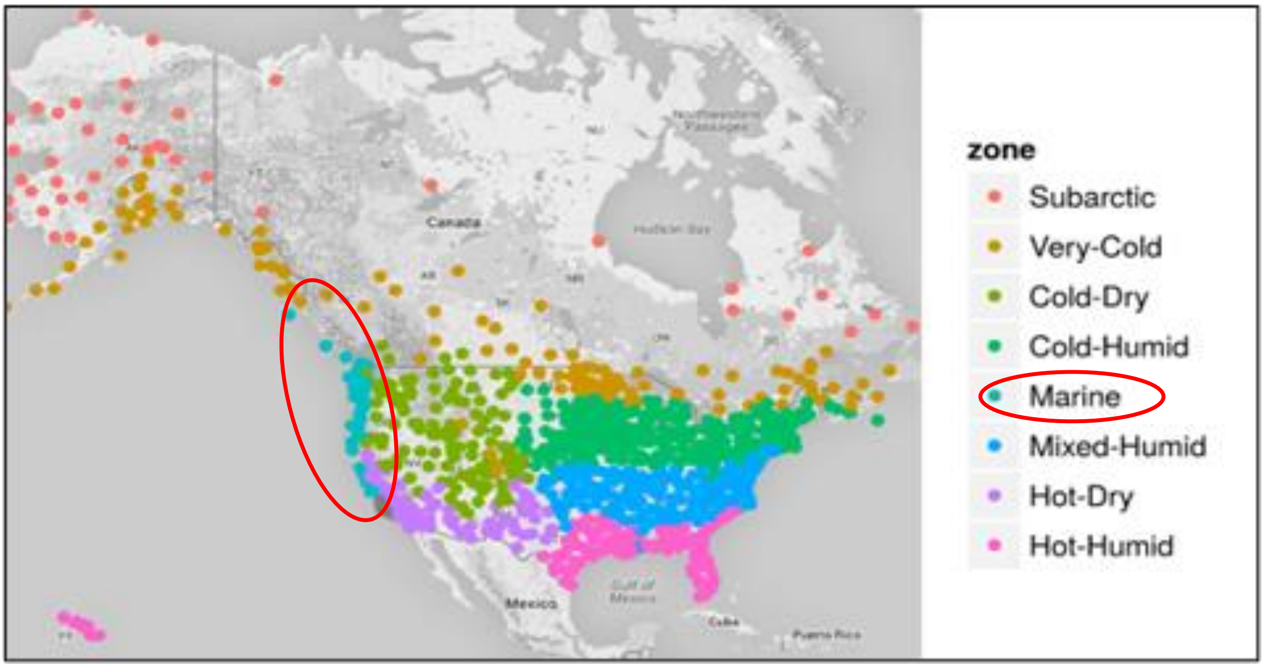


Figure 1.3.3-1. SPE07 Representative Climates [3]

1.3.3.6 Application Ratings

Besides the standard climates and heating and cooling load conditions, Annex F of SPE07 provides alternative rating calculation methods called “Application ratings” so that users can vary the conditions used in the model in a predictable, standardised manner. This allows a designer or analyst to use a specific climate rather than one of the eight prototype climates. It also allows for equipment loads (heating and/or cooling) that vary from the ones used in the test, and for specification of auxiliary heat sources that have a fixed heating output, whether electric or some other fuel. For an application rating, details are provided on how such a result needs to be reported so that the application-specific conditions are properly disclosed.

1.3.3.7 System Performance Metrics

Improved climate-specific metrics such as SCOP provide a mechanism for energy efficiency incentive programs to estimate savings for specific heat pump models appropriate for various climates. Better predictions of performance, using SCOP values based on tests conducted with native controls, will allow programs to more accurately attribute value for incentives and other support, to better match targets for savings to the systems with the highest efficiencies. Load-based test procedures and ratings such as SPE07 should also improve understanding by designers and consumers about the value of various products.

1.3.3.8 Repeatability, Representativeness and Reproducibility

In 2022-2023, a research project organised by the Northeast Energy Efficiency Partnerships (NEEP) and sponsored by many US and Canadian organisations has measured performance of six heat pumps in the field and also in the lab, using both AHRI 210/240 (category A) and SPE07 (Category B) methods, with the purpose of assessing the representativeness of “real” field operation of each laboratory test method. The six heat pumps were installed in three identical (unoccupied), calibrated manufactured homes in Nebraska (US). They were monitored in cooling and heating operation from Aug 2022 to March 2023. The details of the field phase are published in (NEEP, 2023a) and summarised in Harley et al. (Harley et al., 2023)
After the field data collection, the six units were tested by a lab that has much experience using SPE07, to compare the field performance with the reported efficiency metrics from the two test methods. The conclusion is that SPE07 is more representative, although there was more low bias in the cooling rating using

---

[3] The Marine climate is circled for clarity to differentiate it from climates shown in similar colors.

SPE07 than expected. A summary of the results is shown in Table 1.3.3-2 and Figure 1.3.3-2. In all cases, the field data and the M1 (which is related to the conditions of AHRI210/240 results) are both normalised to the same climate used in SPE07 to ensure they are comparable. This is explained in the cited papers.

Table 1.3.3-2. The root mean squared errors (RMSE) and mean absolute percent errors (MAPE) for SPE-07 and M1, using field SCOP as a reference.

	Cooling RMSE		**Heating RMSE**		**Cooling MAPE**		**Heating MAPE**	
	SPE07	M1	SPE07	M1	SPE07	M1	SPE07	M1
Ducted	0.74	0.45	0.26	0.40	13%	9%	11%	17%
Ductless	0.92	2.14	0.20	1.39	13%	43%	10%	64%
Combined	**0.82**	**1.40**	**0.24**	**0.93**	**13%**	**22%**	**10%**	**36%**

In the end, there were only five units with valid data for the comparison, three ducted and two ductless. In all cases for the entire group, the SPE07 errors are smaller, although when looking at the ducted and ductless subgroups, the errors were larger for SPE07 in the ducted group for cooling. The sample size is tiny, however, to generalise the results to ducted and ductless units.

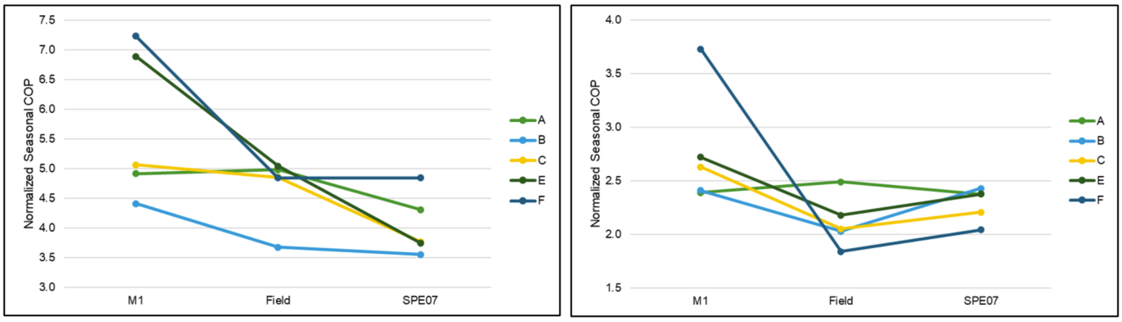


Figure 1.3.3-2. Normalized seasonal COP for M1, field, and SPE07 for cooling (left) and heating (right)

Figure 1.3.3-2 shows a visual representation of the normalised results. Here, the low bias of SPE07 in cooling is apparent (4 of the 5 units), and the more extreme over-statement of efficiency of M1 (AHRI 210/240) for some units in both cooling and heating can be seen. Further details of this study are awaiting publication but should be found in NEEP (NEEP, 2023b) and Harley et al.(Harely et al., 2024) .

In addition, during the lab tests of this study, two of the heat pumps have been re-tested to assess repeatability. Along with a previous study on EXP07 by AHRI and Purdue University (Dhilon et al., 2022), this small sample suggests repeatability is within ±3% at a 95% confidence interval. Although in the AHRI assessment, reproducibility was not very good for EXP07, it is expected that it will be improved for SPE07, and a second lab will begin testing two of the units from the NEEP representativeness study shortly after this writing.

1.3.4 Load-based test to obtain relationships between partial load ratio and energy efficiency of VRF systems by Better Living

The purpose of this proposed test protocol is to improve the testing and evaluation of variable refrigerant flow (VRF) systems.

The testing and evaluation of the multi-split system air conditioner and air-to-air heat pump, according to ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) requires fixing the compressor speed, opening the electronic expansion valves (EEV), and adjusting the unit's set-point to the lowest temperature during cooling or the highest temperature during heating. Currently available products automatically control the compressor speed and electronic expansion valve to maintain a comfortable temperature. This proposed test protocol evaluates the VRF system by analysing automatic control of compressor speed and electronic expansion valve at different thermal loading.

This proposed test protocol shall be used in conjunction with existing testing and evaluation standards, such as ISO 15042 (ISO, 2017b), JIS B 8615-5 (JIS, 2015c), and BS EN 14511-3:2018 (BSI, 2018), to enhance the realism of testing and evaluation. ISO 15042 (ISO, 2017b), for instance, outlines specific conditions in Section 12.2 of this standard that, when followed, can result in different ratings. Annex F of ISO 15042 (ISO, 2017b) displays the part-load capacity test, and Annex G describes the individual indoor unit capacity tests. To apply this proposed test protocol as additional tests for the VRF system, it is necessary to follow the standards mentioned in (ISO, 2017b), (JIS, 2015c), (BSI, 2018). These standards cover the preparation of the VRF system, the arrangement of the testing facility, the selection and installation of sensors and measuring instruments, and the choice of methods used to measure the parameters needed for data analysis.

1.3.4.1 Terms and Definitions

This proposed test protocol uses the terms and definitions provided in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c). Additional terms and definitions are introduced based on testing and measurement evaluations conducted by Building Research Institute (BRI) and National Institute for Land and Infrastructure Management (NILIM) (Enteria et al., 2015, 2016, 2017).

- ***Thermal capacity*** of the indoor unit(s) is measured using the air enthalpy method. The difference lies in the total enthalpy of the supply and return air, which is then multiplied by the mass airflow rate.
- ***Balanced thermal capacity ratio*** means that each indoor unit has the same thermal capacity.
- ***Unbalanced thermal capacity ratio*** means that each indoor unit does not have the same thermal capacity.
- ***Rated thermal capacity*** test measures the heating or cooling capacity of indoor units based on ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c).
- ***Real operational control*** refers to the control logic of both marketed and installed VRF air-conditioning and heat pump systems.
- ***Real thermal capacity test*** measures the heating or cooling capacity of indoor units based on the system's operational control.
- ***Partial thermal capacity test*** measures the capacity at a lower value than the real thermal capacity test.
- ***Cyclic operation*** happens when the compressor turns on and off, especially when the thermal loading for cooling and heating modes is low. In the case of multi-compressors, one compressor may operate while the other(s) is/are off.
- ***Heating-defrosting operation*** is melting ice accumulation from the outdoor unit's heat exchanger in a heating-and-defrosting cycle.

1.3.4.2 Responsible Persons

Engineers who test and evaluate VRF systems using the methods described in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall also use the protocol proposed in this document to test and evaluate these systems. By incorporating the evaluation method mentioned in this test protocol, test engineers shall have less difficulty in testing products than they currently have when using ISO 15042 (ISO, 2017b), JIS B 8615-3 (JIS, 2015c) and/or BS EN 14511-3:2018 (BSI, 2018) This will also ensure that the person testing the VRF system, based on the existing standards and the proposed test protocol mentioned in this document, can easily differentiate, determine commonalities or similarities in the results, and make a final report on the product being tested. Hence, this proposed testing method will evaluate the VRF system based on its actual operational control strategy, not on the manipulated control strategy used in existing testing standards.

1.3.4.3 Capabilities and Description

1.3.4.3.1 Description

The VRF system shall be tested and evaluated according to the procedure outlined in this proposed test protocol. The test protocol tests the VRF system based on the actual product - the operating logic of the compressor and electronic expansion valves are the same as those found in the commercial marketplace. Prior to this, it shall be tested under the rating conditions specified in ISO 15042 (ISO,2017b), JIS B 8615-3 (JIS, 2015c), BS EN 14511-3:2018 (BSI, 2018), or other relevant national standards. The preliminary test based on standards aims to confirm that the new VRF system adheres to the agreed-upon rules and regulations for its development, creation, and performance. The results of the ISO 15042 (ISO,2017b) and JIS B 8615-3 (JIS, 2015c) standards mentioned in the text can be compared with the results of the proposed test protocol described in this document.

A manufacturer of VRF systems or a third party shall test and measure the performance of the VRF systems based on the test and measurement method described in this proposed test protocol. The test can be performed after the completion of testing and performance measurements recommended by ISO 15042 (ISO,2017b), JIS B 8615-3 (JIS, 2015c) and/or other standards. By incorporating the test and measurement procedure described in this proposed test protocol, more information about the actual performance of VRF systems can be gathered and evaluated than can be gathered using existing protocols alone.

1.3.4.3.2 Testing Facility

A company that manufactures VRF systems or a third party is expected to have a facility designated for testing and evaluating the VRF systems, as shown in Figure 1.3.4-1 The design, construction, maintenance and operation of the test facility are expected to follow ISO 15042 (ISO,2017b), JIS B 8615-3 (JIS, 2015c) and other standards. In a test facility, a VRF system shall be tested and evaluated based on the procedures mentioned in this proposed test protocol, with the proper installation of the required sensors that measure the actual performance of the system, as shown in Figure 1.3.4-2.

1.3.4.3.3 Data Measurement

To ensure that the compressor operates continuously and predictably, data for analysis shall be collected at least 20 minutes after observing stability in the VRF system's operation, as outlined in ISO 15042 (ISO,2017b) and/or JIS B 8615-3 (JIS, 2015c). To analyse cyclic compressor operation, data is collected for three cycles (when the compressor is on and off for cooling, heating, and defrosting) after monitoring the stability of the VRF system operation, as outlined in ISO 15042 (ISO,2017b) and/or JIS B 8615-3 (JIS, 2015c).

1.3.4.3.4 Test Conditions

To test a VRF system using the proposed test protocol, the same control strategy currently in use shall be employed. Prior to testing with the proposed test protocol, the system shall be tested according to the rating conditions outlined in ISO 15042 (ISO,2017b) and/or JIS B 8615-3 (JIS, 2015c).

1.3.4.3.5 Thermal Capacities

The thermal capacities of the VRF system are determined using standardised testing and performance evaluation methods outlined in ISO 15042 (ISO,2017b) and/or JIS B 8615-3 (JIS, 2015c). This ensures that the thermal capacities are rated accurately and consistently. The results of the proposed test protocol shall be compared against the testing results based on ISO 15042 (ISO,2017b) and/or JIS B 8615-3 (JIS, 2015c) as a reference.

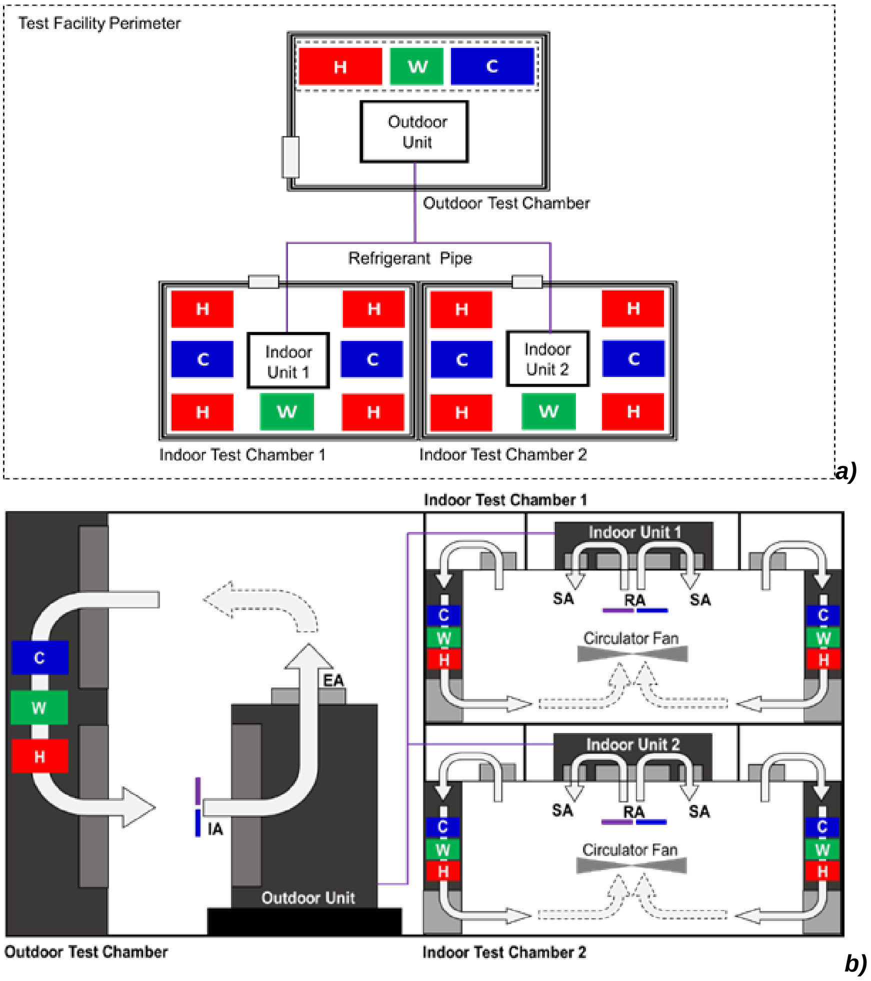


Figure 1.3.4-1. Test facility with one outdoor chamber and two indoor chambers: a) General diagram, and b) Specific diagram. Where H=Heater, C=Cooler, W=Humidifier, SA= Supply air, RA= Return air, IA=Inlet air, EA= Exit air.

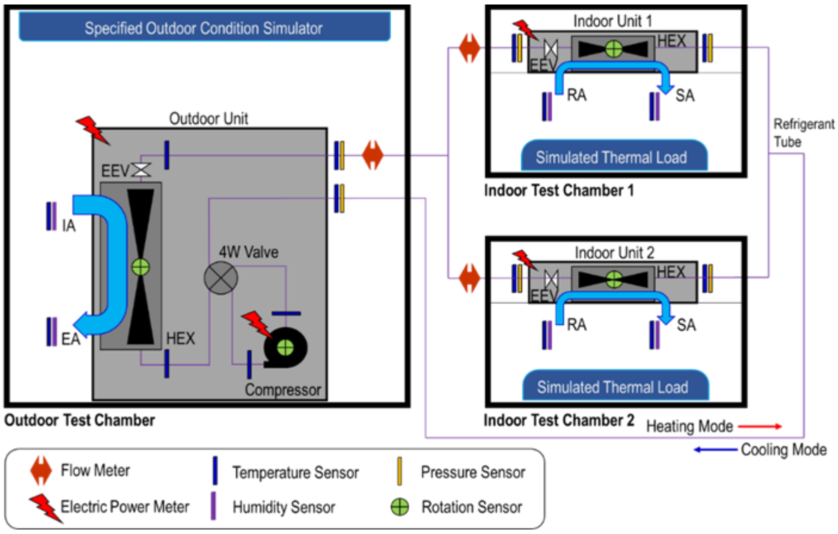


Figure 1.3.4-2. Representational one outdoor unit and two indoor units VRF system with installation and location of important sensors. Where, HEX=Heat exchanger, SA= Supply air, RA= Return air, IA=Inlet air, EA= Exit air.

1.3.4.4 Real Performance Testing

1.3.4.4.1 Outdoor unit

The air flow measurement procedure shall follow the standards set out in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c). Other standards, such as ISO 5167-1 (ISO, 2022) and ISO 5151 (ISO, 2017a) shall also be consulted when making air flow measurements. In addition, as mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c), the instructions given by the manufacturer of the VRF system shall be followed when making air flow measurements.

1.3.4.4.2 Indoor unit

The air flow measurement method used for each indoor unit shall follow the standard method discussed in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c). Other standards, such as ISO 5167-1 (ISO, 2022), ISO 5151 (ISO, 2017a) and ISO 3966 (ISO, 2020) shall also be consulted. In addition, as mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c), the instructions of the manufacturer of the VRF system shall be followed. The air flow rate and noise level suggested by the manufacturer shall be used for testing. This information shall be available in the product catalogue. In addition, the suggested air flow rate shall be an available option for the actual operation of the VRF system. In addition, an actual airflow measurement shall be performed on the installed VRF system to determine the actual airflow of the indoor units of the VRF system (Enteria et al., 2023). The airflow measurement used a device that measures the actual airflow of the installed indoor unit of the VRF system in the actual building. The actual air flow measurement shall be used in the thermal capacity.

1.3.4.5 Cooling Tests Real Performance Testing

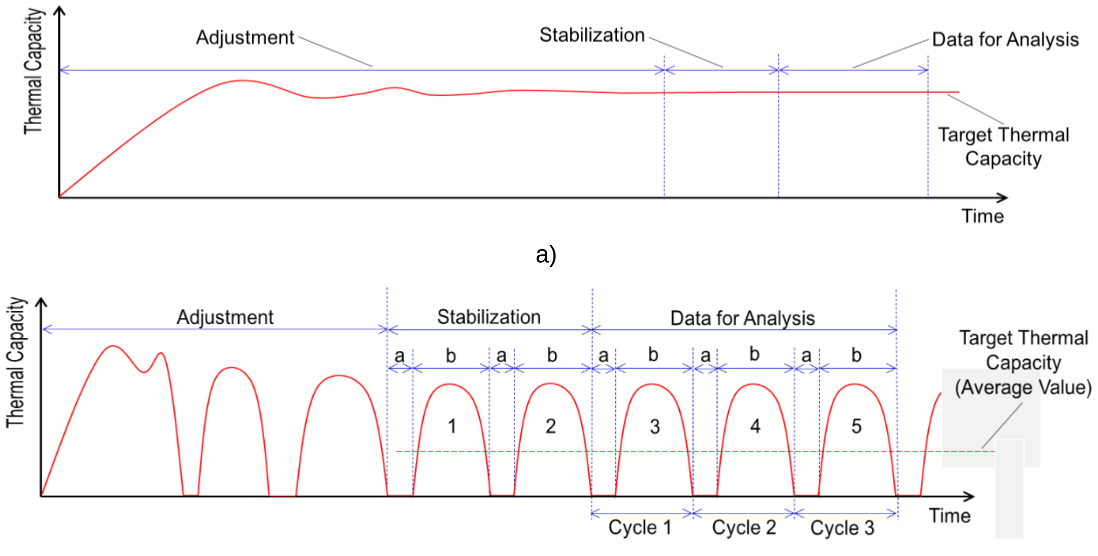
The actual cooling capacity test referred to in this test report shall be performed with all indoor units operating under the air conditions specified in T1 of Table 2 of ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c).

Table 1.3.4-1. Comparison of system performance at cooling mode.

	Cooling mode	
	Cooling capacity, kW	Power consumption, kW
Catalogue value (rated)	22.40	6.61
Measured value (real)	21.41	6.65

The standard capacity test value based on ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be referred to when measuring the actual capacity. The results of the two tests shall be compared and made available as shown in sample Table 1.3.4-1. The measurement of the cooling capacity shall be derived from the air enthalpy method, the calculations of which are specified in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c).

The appropriate conditioning of the air to stabilise the VRF system mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be adopted to ensure the reliability of the data obtained. The data collection and analysis procedures shall follow the standards mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c). In particular, the sample data evaluation and analysis shown in Figure 1.3.4-3 shall be considered. The adjustment period is to evaluate the thermal capacity setting of the chambers. The stabilisation period is to make sure that the thermal capacity reading is already the stable target reading.

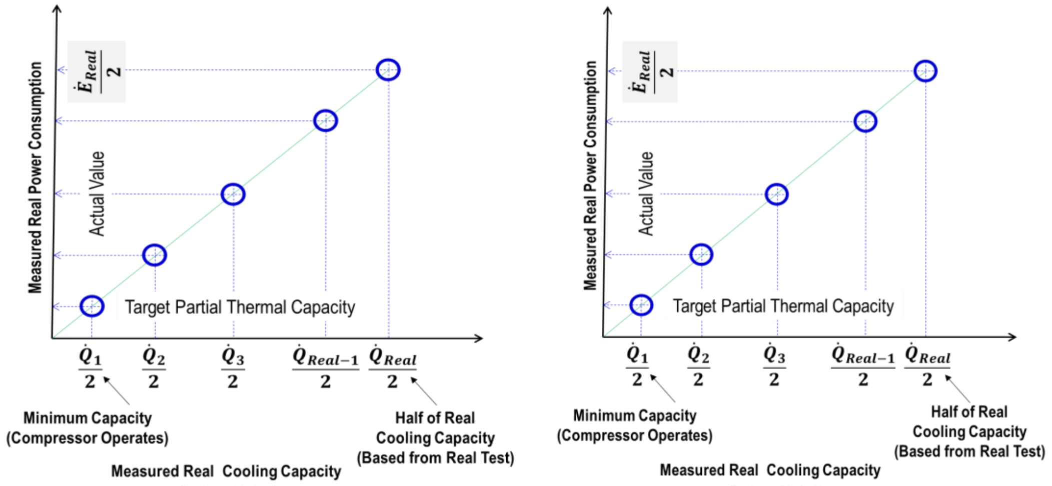


b)

Figure 1.3.4-3. Raw data observation and data for analysis: a) Steady compressor operation, and b) On and off compressor operation.

1.3.4.6 Partial Cooling Capacity Test

As part of the evaluation of a system, a partial cooling capacity test with a balanced thermal capacity ratio shall be conducted. During this test, the total indoor capacity of all units is reduced from the actual cooling capacity to the minimum possible cooling capacity of the VRF system, as shown in Figure 1.3.4-4. In addition, the air conditions specified in T1 of Table 2 in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be used. In addition, the appropriate air conditioning conditions for the stabilisation of the VRF system mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be followed. Data collection and analysis shall also follow the standards mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c). The sample data evaluation and analysis shown in Figure 1.3.4-4 shall be considered, especially during cyclic operation (Figure 1.3.4-4b).



a) b)

Figure 1.3.4-4. Power consumption at partial thermal capacity with balanced cooling capacity: a) indoor unit 1, b) indoor unit 2. Where $\dot{E}_{Real}$=Measured power consumption, $\dot{Q}_i$=Thermal capacity.

1.3.4.7 Heating Tests

1.3.4.7.1 Real Heating Capacity Test

The actual capacity test referred to in this test report shall be performed when all indoor units are operating at the outdoor and indoor air conditions specified in H1 of Table 7 of ISO 15042 (ISO, 2017b) and/or JIS B

8615-3 (JIS, 2015c). The catalogue capacity test based on ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be used as a reference when measuring the actual capacity test and the results shall be compared as shown in Table 1.3.4-2. Heating capacity measurements shall be based on the air enthalpy method with calculations taken from ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c).

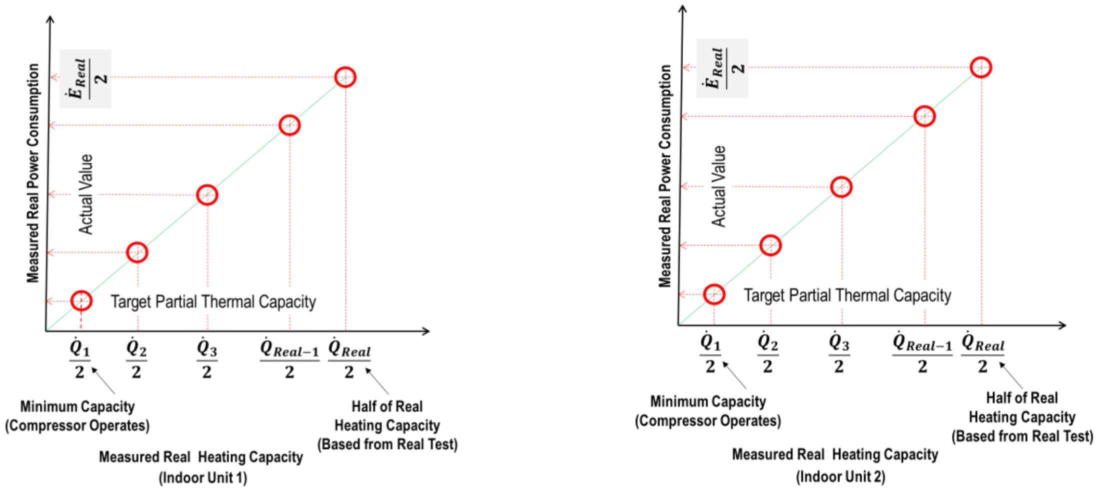
Table 1.3.4-2. Comparison of system performance at heating mode.

	Heating mode	
	Heating capacity, kW	Power consumption, kW
Catalogue value (rated)	25.00	6.43
Measured value (real)	23.88	6.48

The conditioning of the air required to stabilise the VRF system as specified in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be followed. Data collection and analysis shall follow the standards mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c). The sample data evaluation and analysis shown in Figure 1.3.4.3 shall be consulted.

1.3.4.7.2 Partial Heating Capacity Test

The partial heating capacity test shall be performed with a balanced thermal capacity ratio. In this test, the total capacity of all indoor units is reduced from the capacity of the real heating capacity test to the minimum possible heating capacity at which the compressor of the VRF system can operate (Figure 1.3.4.5). The air conditions recommended in H1 of Table 7 ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be used. The air conditioning required to stabilise the VRF system according to ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c) shall be maintained. The data collection and analysis procedures shall adopt the standards mentioned in ISO 15042 (ISO, 2017b) and/or JIS B 8615-3 (JIS, 2015c). The sample data evaluation and analysis shown in Figure 1.3.4.5 shall be considered, especially during cyclic operation (Figure 1.3.4-5b).



a) b)

Figure 1.3.4-5. System power consumption at partial thermal capacity with balanced heating capacity: a) indoor unit 1, b) indoor unit 2. Where $\dot{E}_{Real}$=Measured power consumption, $\dot{Q}_{\iota}$=Thermal capacity.

1.3.5 Load-based testing of hydronic heat pumps - load-based method (by BAM) and hardware-in-the-loop testing (by RWTH)

This section describes load-based test methodologies for hydronic heat pumps that are connected to a water-based heating system on the sink side. The source side can be air, brine, or water. The main section deals with a load-based test which aims to compare products under standardised yet representative test conditions, whereas the final section gives a brief overview of hardware-in-the-loop testing as a holistic evaluating method.

The Federal Institute for Materials Research and Testing (BAM) assessed the current standards EN 14511 (BSI, 2018) and EN 14825 (BSI, 2022) within a research project ("Support for market surveillance – NAPE" (2015-2022). To solve multiple issues arising from fixing the compressor speed and overriding the heat pump controller during the standard test, BAM developed the load-based test to measure units with active control under normal operation mode. This yields representative operation behaviour (on-off cycling under part-load), ensures that different appliances are tested under the same conditions (no individual increase of supply temperature/heating capacity under part-load conditions) and enables testing independent from the manufacturer. In 2019, BAM submitted a proposal in the review process of the EU ecodesign and energy labelling regulations for space heating appliances to revise the current EN-standards summarising the shortcomings of the current standard and the benefits of the load-based method (Simo et al., 2019). To validate the new method, repeatability and reproducibility were investigated in two round-robin tests (Wachau et al., 2023a). It was found that the inertia of the test stand impacts the operating behaviour under part-load conditions (Göbel et al., 2022). Therefore, the method was refined by introducing a simplified building model (emulator), which ensures the same response of different test stands (aligned inertia) and ensures the test stand responds like a real building (increased representativeness). The method is described in a test guideline published by BAM (BAM, 2023a) and an example of the building model is available in the form of a Python script on GitHub (BAM, 2023b). The proof-of-concept was successfully demonstrated in 2023 (Wachau et al., 2023b) and is followed by further round-robin tests. The following sections describe the emulator approach as of January 2024, including the two-mass building model.

1.3.5.1 Conceptual description of the testing methodology

As mentioned before, conventional heat pump test stands can vary significantly in their hardware and control design, resulting in very different response (physical and virtual inertia) under dynamic operation of the tested unit (e.g. on-off cycling or defrosting). Therefore, the emulator approach developed by BAM aligns inertia across different test stands by implementing a virtual building model in the test stand to ensure reproducible operating behaviour. This technology neutral approach allows existing hardware to be used (With minor hardware modifications/extensions any test standard is suitable).

The simplified building model (two-mass model, cf. Figure 1.3.5-1) computes the heat pump's water inlet/return temperature $\vartheta_{\mathrm{R,calc}}$ (and the building temperature $\vartheta_{\mathrm{B,calc}}$, where applicable) which is coming from the virtual building towards the heat pump. The test stand emulates these computed temperatures during the entire test duration. In particular, for each time step Δt_{step}, the model calculates the return temperature $\vartheta_{\mathrm{R,calc}}$ and the building temperature $\vartheta_{\mathrm{B,calc}}$ (output variables) based on the measured heat pump's water outlet temperature/supply temperature ϑ_{S} and the heating capacity $\dot{Q}_{\mathrm{HP}}$ (input variables) of the unit under test. Through the measured heating power, the model considers the supply temperature ϑ_{s}, return temperature $\vartheta_{\mathrm{R,emu}}$, and mass flow rate $\dot{m}_{\mathrm{W}}$. Like the supply temperature, the mass flow is controlled by the heat pump.

Based on a simple energy balance, the test stand dynamically adapts the so-called compensation load to match the calculated return temperature $\vartheta_{\mathrm{R,calc}}$. The heat pump responds in accordance using its heating curve or its indoor temperature control or both as it tries to maintain the required supply temperature ϑ_{s}.

Aligned with testing conditions in EN 14825 (BSI, 2022), the outdoor room shall maintain constant conditions over a temperature range associated with different climate zones for testing air-water heat pumps.

Compared to the conventional testing procedure, the heat pump under the test is operated with its onboard control system (native control) active and not in a fixed-speed mode. The heat pump is, thus, permitted to switch into on/off operation of the compressor.

To ensure representativeness the building model is parametrised according to the test conditions defined in EN 14825 (BSI, 2022) considering the temperature application (e.g. low, medium or high) and the climate zones (cold, average, warm) (The concept can be applied to any temperature application, but representative time constants must be applied). In addition, the P_{design} of the heat pump is considered scaling the size of the virtual building. The equations and a detailed description can be found in the test guideline published on the BAM website (BAM, 2023a).

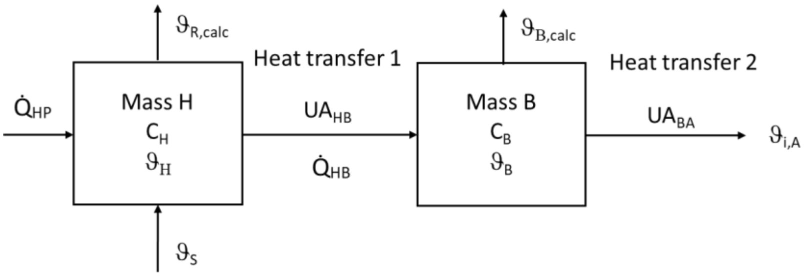


Figure 1.3.5-1. Schematic of the two-mass building model

1.3.5.2 How the tested system is operated

The emulator method allows for the use of conventional test stands with minor modifications to implement the building model. Depending on the type of unit, the source side contains a water loop or a climate chamber which ensures constant brine/water and outdoor conditions, respectively. The sink side comprises a water cycle, which is used to apply the required load. In accordance with EN 14825 (BSI, 2022), a test condition dependent setpoint for the supply temperature (water outlet temperature) and the heating capacity must be reached. For both measurements, the arithmetic mean value (over full cycles for on-off or defrost operation) is used.

The emulator method subjects the unit under test to the load dynamics of a representative building. In contrast to the current standard, the unit under test is operated with its on-board (native) control active. Prior to testing, only slight adjustments of the factory settings must be made on the installer level (single heating circuit, disable domestic hot water, etc.). In addition, the heating curve settings inside the controller are adjusted to match the set point for the supply temperature required by the specific test condition, as an installer would do. The controller modulates the compressor speed to match the load as it would do in the field. Hence, on-off operation is observed for loads below the modulation limit of the compressor. For loads below the bivalence point, two options can be applied: (a) the real or (b) a virtual electrical auxiliary heater is active, the power input of which is considered in the evaluation.

Figure 1.3.5-2 illustrates the difference between the fixed compressor speed test according to EN 14511 (BSI, 2018) and the load-based test on the same heat pump. The test conditions (E, A, B, C and D) defined in EN 14825 (BSI, 2022) are the same in both cases. However, to ensure steady-state operation of the compressor, the heating capacity must be increased below the modulation limit of the compressor. Thus, any deviation from the prescribed load-line is allowed in the standard test for conditions where on/off operation would occur and must be corrected afterwards. Consequently, different heat pumps are not tested under the same test conditions since the adjustment is individual for each heat pump. In contrast, the required load is always met during load-based tests within the permissible deviations, since on/off operation is enabled via active control. Hence, the operation behaviour is much more representative.

Finally, load-based tests can be performed independently from the manufacturer, whereas the standardised fixed frequency test in general, requires intervention from the manufacturer to set different parameters in the test mode.

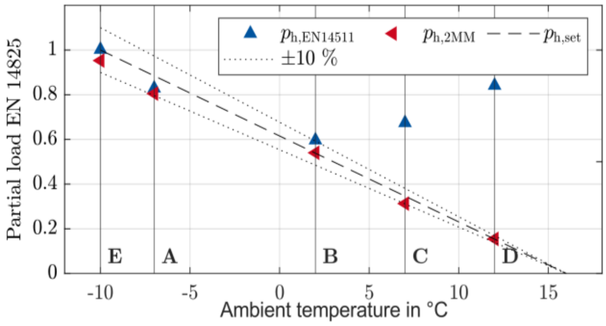


Figure 1.3.5-2. Partial loads measured on the same heat pump according to EN 14511 (BSI, 2018) and the emulator method (2MM) at different test conditions defined in EN 14825 (BSI, 2022). The load curve

(setpoint) and permissible deviations according to EN 14825 (BSI, 2022) are depicted by dashed and dotted lines, respectively

1.3.5.3 Illustrative test results

Figure 1.3.5-3 compares measurements at part load condition C according to EN 14511 (BSI, 2018) and the emulator method. During the standard test, the supply and return temperatures are quasi-constant since the compressor speed is fixed, whereas on-off operation is observed for the load-based test, which is reflected in the periodic increase in supply and return temperatures. The dynamic operation in the latter case is due to the native controller trying to match the load below the modulation limit of the compressor. As previously emphasised, the mean value of the supply temperature is much higher for the standard test, which requires steady-state operation of the compressor, resulting in a too high heating capacity (supply temperature) compared to the test condition defined in EN 14825 (BSI, 2022).

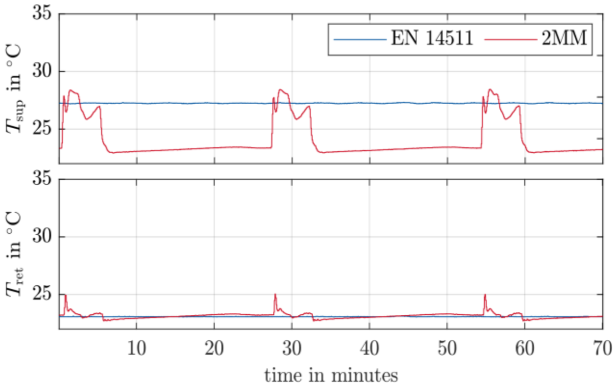


Figure 1.3.5-3. Measured supply and return temperatures according to EN 14511 (BSI, 2018) and the emulator method (2MM) on the same heat pump.

1.3.5.4 Selection of set of test conditions

The load-based test can be applied to any test condition if the source side conditions are constant. So far, test conditions according to EN 14825 (BSI, 2022) have been used since the load-based test was developed to replace the fixed frequency test defined in EN 14511 (BSI, 2018).

1.3.5.5 Assessment of repeatability, reproducibility, and representativeness of the test results

The repeatability and reproducibility of the load-based test has been assessed in two round robin tests (RRT) with an A/W and a W/W heat pump from 2020-2021 (Wachau et al., 2023a). Similar reproducibility was found compared to the current standard. However, very different inertia of the test rigs in the RRT leads to non-uniform operating behaviour. Especially, quick responding test rigs with low inertia failed to reach the setpoint for the heating capacity. In the following, the two-mass building model was introduced to align the test stand response independently from its physical inertia and ensure the same operating behaviour. The concept has been proven on three test stands with three different heat pumps (Wachau et al., 2023b). Starting from September 2023, a new round-robin test is launched by BAM and RWTH to investigate the reproducibility of the emulator (building model) based approach and to refine the test guideline based on the observations.

1.3.5.6 Definition of seasonal performance indices

The calculation of the seasonal coefficient of performance (SCOP) is defined in EN 14825 (BSI, 2022) and can be applied to load-based measurements. Slight adjustments in the calculation procedure are required since the electrical power consumption for the back-up heater is directly recorded during the load-based measurement and included in the measured COP opposed to the standard correcting a lack in heating capacity during the SCOP calculation. The electrical power consumption of the back-up heater can either be measured directly, in case of a real back-up heater, or calculated virtually.

1.3.5.7 Outlook: Holistic Testing of Building Energy Systems

The load-based-testing method described in the following is based on the Hardware-in-the-Loop (HiL) approach that couples hardware and software in real-time. At RWTH Aachen University, we developed a method for testing the holistic building energy system, including further components like the hydraulic transfer system, PV-systems or thermal energy storages (TES) (Mehrfeld, 2022). The device under test can be the heat source (e.g. heat pump), the TES, and the control algorithms. Therefore, the scope of this method goes beyond the load-based approach described above, which aims to compare product performance under standardised conditions.

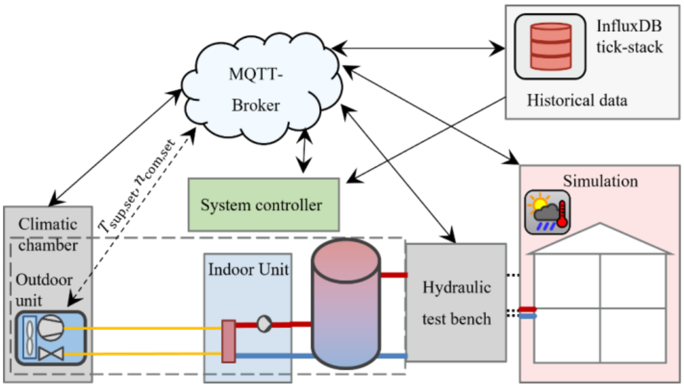


Figure 1.3.5-4. Schematic overview of holistic BES testing.

The developed method creates an experimental-based annual KPI (e.g. SCOP) by performing the following steps:

1. The BES is modelled, including all components.
2. A sensitivity analysis and clustering algorithm delivers typical days for a specific location.
3. The typical days are experimentally investigated with the HiL approach
4. Annual KPIs are calculated from the daily KPIs.

Figure 1.3.5-4 shows the schematic overview of the holistic test approach. A modern model predictive controller (MPC) is investigated in the example. We use a fully controllable heat pump test bench for a deep control interface. To couple the climatic chamber, the hydraulic test bench, the system controller, and the heat pump with the building performance simulation, we transfer data via the MQTT protocol. The building performance simulation is a multi-zone Modelica model realised with the BESMod Modelica library (Wüllhorst et al., 2022).

Figure 1.3.5-5 shows exemplary the test of one typical day for the BES controlled by an MPC. The figure shows the room temperature (red line) and comfort bounds (black line) at the top. The middle figure illustrates the heat pump's relative compressor speed while the bottom figure gives the set supply (black dashed line) and the measured supply temperature (red line). The experiments show the potential for holistic BES testing and support the introduction of complex control algorithms into practice. Further details can be found in (Göbel et al., 2023).

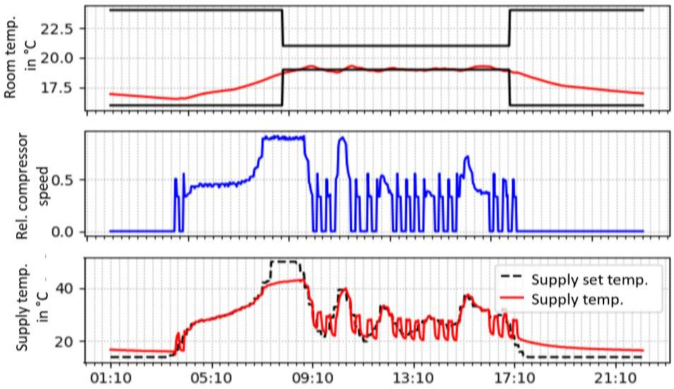


Figure 1.3.5-5. Hardware-in-the-Loop test for a building energy system controlled by an MPC.

1.4 Perspectives of load-based test standards and R&D plans in Annex 88

1.4.1 Perspective of load-based tests and considerations on methods of utilisation of the results

When the building-level policies started in the late 1970s, they dealt only with the thermal performance of the building envelope. After year 2000, many national or regional building-level policies started covering equipment's energy performance, including heat pump systems. Because of the fundamental nature of heat transfer, it is conceptually incorrect to separate the performance of heating and cooling supply technologies from the building envelope and vice versa. However, in practice, the different approaches adopted in product- and building-level policies have resulted in fundamental differentiations, if not incompatibilities, between the two levels of analysis. Insufficient communication between tests conducted for product- and building-level policies results in inconsistent testing conditions and methodologies, making results obtained when testing product efficiency not applicable for building evaluations, and eventually increasing the required testing time and cost to the industry.

The review of new testing methodologies and rating standards conducted in Annex 88 is also intended to recognise potential convergences between the information extracted during product-level performance ratings and building-level energy calculations, simulations, and equipment sizing.

As mentioned in Section 1.1, product-level policies are intended to provide values representing products' energy efficiency and are used to compare different products of the same kind. Product-level evaluation of energy efficiency does not aim to characterise the complete spectrum of possible load scenarios and building characteristics, and simplifying assumptions must be made on the relationship between outdoor temperature and heat needs, for instance.

On the other hand, building-level policies and standards rely on values representing products' energy efficiency, which are used to evaluate overall energy performance (i.e., total energy consumption) of the buildings. The energy calculation results are frequently used to compare energy reductions with different products and technologies. Additionally, designers' decisions related to equipment sizing (e.g., heat pumps) are one of the main targets for evaluating the building´s energy performance. To increase the resolution and reliability of building-level evaluations, energy efficiency of equipment under low partial load conditions has become critical, mainly because the actual partial load ratio for the equipment may substantially deviate from the assumptions made in the product-level standards, and it is not uncommon for designers to oversize building equipment to avoid any shortage in heating/cooling capacity. Therefore, building-level policies increasingly require evaluations with higher resolution and complexity, which imply a continuous methodology improvement in the search for higher resolution of the energy characterisation. Examples of policies are added in the lower part of Table 1.4.1-1, which summarises the essential features of product- and building-rating policies.

Testing methodologies for product-level policies provide the fundamental measurements and material for:

- the development of effective Minimum Energy Performance Standards for meeting the conflicting challenges of increasing demand for heating and cooling with the necessity of energy saving,
- defining the basis for performance rating of units available in the market,
- capturing realistic operation characteristics that may stimulate technology developments, evidence-based policies, and guide consumers to beneficial choices.

The development of testing methodologies for assessing the performance of heat pumps and air conditioners when operated under the same control as in buildings presents both challenges and opportunities. These arise from the dynamic characterisation of system operation and the performance relationships to building and load features, similar to those observed in field installations. The reviewed testing methodologies are intended to: develop new product level standards, support building-level policies by providing data for energy modelling and simulation purposes, provide evidence for efficient equipment sizing and selection, as well as for the development of more efficient design and control.

Table 1.4.1-1. Comparison of product- and building-level standards.

		Product-level standards:	Building-level standards:
(1) Scope		Provide comparable values representing products' energy efficiency to compare different products of the same type. Allow for determination of high and low performing equipment of the same equipment type.	Evaluate overall energy performance of the building and evaluate the suitability of different kinds of systems within the building. *default characteristics for energy efficiency under partial load conditions are presently being utilised along with the rated energy efficiency of the HP.
(2) Seasonal or annual average efficiencies		Necessary for regulating the energy efficiency level of each product category. The assumption of the relationship between heat needs imposed on HP and the maximum capacity of the HP: fixed ratios are applied, such as 1.0 for cooling in JIS C 9612. Assumption on the relationship between the heat needs and outdoor temperature: a linear relationship is assumed.	Not necessary. Instead, whole building energy performance is regulated with calculated energy use by buildings. The relationship between the thermal load imposed on HP and the maximum capacity of the HP is influenced by building/interior space usage and designers' decision on sizing the HP. The thermal load is also influenced by solar radiation and outdoor humidity.
Examples of relevant policies	EU	The Ecodesign Directive prescribes minimum requirements for SEER and SCOP, and only products compliant with the requirements can be sold. The definitions of the SEER and SCOP are prescribed in EN14825 based on EERs and COPs at load ratios of 100%, 75%, 50%, and 25%.	National building energy standards based on EPBD. European standards on the methods for energy calculation are developed as EPB standards. EN 15316-4-2 is one of them, which is for space heating heat pump systems.
	US	AHRI standards prescribe SEER and HSPF (Heating Seasonal Performance Factor) based on test results at full, intermediate, and low compressor stages. DOE implements minimum SEER and HSPF with the authorization of the National Appliance Energy Conservation Act of 1987.	For non-residential buildings, ASHRAE Standard 90.1 prescribes the whole building performance approach using a simulation tool, such as EnergyPlus. The building energy codes are implemented based on the Energy Conservation and Production Act of 1976.
	JP	Top-runner programs for HP systems are implemented in the Energy Conservation Law and its ordinances. Relevant JIS standards with testing methods for full and intermediate capacities define the Annual Performance Factor (APF).	In the Building Energy Conservation Law and its ordinances, the calculation methods for primary energy use of buildings are prescribed with standard primary energy uses. In the methods, rated energy efficiencies and default curves for the relationship between partial load ratio and input power for HPs are used for energy use calculations.

In principle, load based tests rely on the same equipment and instrumentation required by current standards, while revisiting the software elements of the testing facility (though some specialised test apparatus may be required depending on the method undertaken). It can be arguably stated that load-based tests might require more time for test convergence than current standards, but this may be related to the necessary learning curve needed for new procedures and could be quantitatively assessed during subsequent efforts of subtask B1 on testing methodologies. The possibility of testing heat pumps under the same control as operating in field installations provides opportunities for automating tests and provides additional value in terms of representativeness of field operation and transparency.

Finally, a strategic choice of test points for product standards has the potential to narrow the gap between product- and building-level policies, and eventually limit overall testing time and cost when considering overall interests of manufacturers, designers, planners, and installers. One challenge to this process is that the interest of planning and product comparison is to have a relatively simple (and less expensive) method to reasonably demonstrate a standardised metric, mostly for product comparison or differentiation; but this lower cost approach does not provide the complete performance maps that are needed for accurate simulation and design. More work can be done to bridge that gap, which may include: streamlining test procedures and the choice of test conditions to provide a better compromise between these differing needs; the use of load based testing to validate a subset of fixed-speed data (that may be available at a wide range of operating conditions); the creation of better models that allow interpolation of tested, load-based operating conditions to other conditions that facilitate both types of metrics; or other innovations, which also may vary by different heat pump technologies.

1.4.2 Perspective within Subtask B1 of Annex 88 Perspective of load-based tests and considerations on methods of utilization of the results

Efforts to advance load-based and innovative testing methodologies should address the challenges and technical solutions necessary for developing experimental methods that go beyond product performance characterisation. These methods may also serve the purposes of design and control development, modelling, energy calculations, and efficient energy management techniques, particularly in the context of heating and cooling technologies and their complex interaction with the built environment and grid dynamics.

During the working phase of Annex 88, the reviewed proposals for the testing methodologies of Category B standards will be compared in detail to exchange expertise and provide evidence of the required testing time and cost, as well as repeatability, reproducibility, and representativeness of the results. Additionally, the comparison of procedures and results with the corresponding Category A standards shall provide quantitative insights for clarifying the performance gap with heat pumps and air conditioners when operated under the same control as operated in buildings.

Consequently, pathways toward adoption including such opportunities as regulator adoption, program adoption (e.g., as an eligibility for incentives or subsidy programs), and building code reference, should be discussed.

The activity of Annex 88 is also intended to increase result comparability across jurisdictions and harmonise standard performance rating procedures toward convergence to a common proposal: for instance, the final results of Annex 88 will support the proposal for load-based tests currently under review within the ISO TC86 for the drafting of a new standard.

1.5 References for testing methods

AHRI. Standard AHRI 1230 (2010). Performance Rating of Variable Refrigerant Flow (VRF) Multi-Split Air-Conditioning and Heat Pump Equipment.

AHRI. Standard AHRI 210/240 (2023). Performance Rating of Unitary Air-conditioning & Air-source Heat Pump Equipment.

ANSI/ASHRAE. Standard ANSI/ASHRAE 37-2009 (2019). Methods of Testing for Rating Electrically Driven Unitary Air-Conditioning and Heat-Pump Equipment.

BAM -Bundesanstalt für Materialforschung und -prüfung (2023a). Test guideline for a load-based performance testing (heat pumps) 12/2023. Available: https://netzwerke.bam.de/Netzwerke/Content/DE/Downloads/Evpg/Heizen-Kuehlen-Lueften/bam%20test%20guideline%20-%20load-based%20testing%20of%20heat%20pumps%202023.pdf.html.

BAM -Bundesanstalt für Materialforschung und -prüfung (2023b). Python script of two-mass building model for load-based-testing of hydronic heat pumps. Available: https://github.com/BAMresearch/bam-load-based-testing

Ban, T.; Ohno, K.; Yamaguchi, S.; Saito, K.; Yamaguchi, H. (2016). Evaluation Method of Compression type Heat Pump for Actual Driving Performance - Construction of evaluation device and proposition of unsteady evaluation method. Proceedings of 2016 JSRAE Annual Conference, Kobe, Japan. (in Japanese).

Ban, T.; Nakano, D.; Ohno, K.; Yamaguchi, S.; Saito, K.; Yamaguchi, H. (2017). Testing facility for unsteady performance of air conditioning System. Proceedings of 12th IEA Heat Pump Conference 2017, Rotterdam.

BSI – British Standard Institution (2018). Standard BS EN 14511-Part 3: Air Conditioners, Liquid Chilling Packages and Heat Pumps for Space Heating and Cooling and Process Chillers, with Electrically Driven Compressors. Test methods.

BSI – British Standard Institution (2022). BS EN 14825:2022: Air conditioners, liquid chilling packages and heat pumps, with electrically driven compressors, for space heating and cooling, commercial and process cooling. Testing and rating at part load conditions and calculation of seasonal performance.

CSA. CSA C656 (2014). Performance standard for split-system and single-package air conditioners and heat pumps.

CSA. CSA SPE-07:23 (2023a). Load-based and climate-specific testing and rating procedures for heat pumps and air conditioners. Available: https://www.csagroup.org/store/product/CSA%20SPE-07%3A23/ (free of charge at http://community.csagroup.org/).

CSA. EXP07:19 Plain Language Guide (2023b). Available in: https://www.csagroup.org/store/product/CSA%20SPE-07%3A23/.

Dhillon, P; Wang, X; Horton, W.T.; Braun, J. (2022). Repeatability and Reproducibility Assessment of Residential Heat Pump Performance Evaluation Methodologies based on CSA EXP07 and AHRI 210/240. Herrick Refrigeration and Air Conditioning 2022 Conference Proceedings. West Lafayette, IN: Purdue University; 2022. http://docs.lib.purdue.edu/iracc/2477/.

Dondini, D.; Giannetti, N.; Mizuno, A.; Miyaoka, Y.; Saito, K. (2024). Reproducibility assessment of an emulator-type load-based testing methodology, International Journal of Refrigeration, Volume 159, 2024, Pages 39-49.

Enteria, N.; Yamaguchi, H.; Miyata, M.; Sawachi, T.(2015). Performance Evaluation of the Variable Refrigerant Flow (VRF) Air-Conditioning System Subjected to Partial and Unbalanced Thermal Loading. *Journal of Thermal Science and Technology*, Vol. 11.

Enteria, N.; Yamaguchi, H.; Miyata, M.; Sawachi, T.(2016). Performance Evaluation of the Variable Refrigerant Flow (VRF) Air-Conditioning System Subjected to Partial Loadings at Different Outdoor Air Temperature. *Journal of Thermal Science and Technology*, Vol. 11.

Enteria, N.; Yamaguchi, H.; Miyata, M.; Sawachi, T.; Kuwasawa, Y. (2017). Performance Evaluation of the Variable Refrigerant Flow (VRF) Air-Conditioning System during the Heating-Defrosting Cyclic Operation. *Journal of Thermal Science and Technology*, Vol. 12.

Enteria, N.; Yamaguchi, H.; Miyata, M.; Sawachi, T.; Kuwasawa, Y. (2023) An Air Enthalpy Method for Measuring the Thermal Capacity of an Installed Variable Refrigerant Flow System, Chapter 1. In: Enteria N., Sawachi T., Saito K. (Editors) Variable Refrigerant Flow System – Advances and Applications of VRF, Springer, Singapore.

Giannetti N.; Ariyadi, H.; Miyaoka, Y.; Jeong, J.; Saito, K.(2022a). Development of an Emulator-Based Assessment Method for Representative Evaluation of the Dynamic Performance of Air Conditioners. International Refrigeration and Air Conditioning Conference. Paper 2448.

Giannetti, N.; Matsui, S.; Mori, R.; Jeong, J.; Ariyadi, H.; Miyaoka, Y; Togashi, E.; Saito, K. (2022b). Emulator-type load-based tests for dynamic performance characterization of air conditioners. Energy and Buildings 273, 112411.

Giannetti, N.; Sholahudin; Mizuno, A.; Miyaoka, Y.; Sei, Y.; Enoki, K.; Saito, K. (2024). Feed-forward compensation for emulator-type testing facilities, International Journal of Refrigeration, Volume 167, 2024, Pages 257-268, ISSN 0140-7007, https://doi.org/10.1016/j.ijrefrig.2024.07.023.

Göbel, S.; Zottl, A.; Noack, R.; Mock, D.; Wachau, A.; Vering, C.; Müller, D. (2022). How to calibrate heat pump test stands for load-based testing – towards technology-neutral prescriptions. Proceedings of the 14th International Conference on Applied Energy (ICAE2022), Bochum, Germany.

Göbel, S.; Stoffel, P.; Will, F.; Müller, D. (2023). Maximizing operational efficiency of heat pumps with Model Predictive Control: An experimental case study for residential application, Proceedings of the 14th IEA Heat Pump Conference, Chicago, May 2023.

Harley, B. Butler, D., Dymond, C. Hamer, G., McWilliams, J. and Yuill, D. (2023). Field tests of variable speed heat pumps to compare load-based and fixed-speed test and rating methods. Proceedings, International Energy Agency, 14th IEA Heat Pump Conference, Chicago, IL. www.hpc2023.org/wp-content/uploads/gravity_forms/3-7075ba8a16c5f78b321724d090fb2a34/2023/05/0438_HPC2023_Full_Paper_Yuill_v02.pdf .

Harley, B.; Dymond, C.; Yuill, D. Hamer, G; McWilliams, J.; Chen, Y. (2024). Field and Lab Testing of Residential Heat Pumps to Assess Representativeness of Fixed-speed and Load-based Test Methods, Proceedings of Purdue University, 2024 Herrick Refrigeration and Air Conditioning Conference.

ISO - International Organization for Standardization (2017a). Standard ISO 5151:2017: Non-Ducted Air Conditioners and Heat Pumps Testing and Rating for Performance. International Organization for Standardization.

ISO - International Organization for Standardization (2017b). Standard ISO 15042:2017: Multi Split-System Air Conditioners and Air-To-Air Heat Pumps – Testing and Rating for Performance. International Organization for Standardization.

ISO - International Organization for Standardization (2020). ISO 3966:2020. Measurement of Fluid Flow in Closed Conduits Velocity Area Method Using Pitot Static Tubes. International Organization for Standardization.

ISO - International Organization for Standardization (2022). ISO 5167-1:2022: Measurement of Fluid Flow by Means of Pressure Differential Devices Inserted in Circular Cross-Section Conduits Running Full. Part 1: General Principles and Requirements. International Organization for Standardization.

JIS - Japanese Industrial Standards (2013). Standard JIS C 9612:2013: Room air conditioner.

JIS - Japanese Industrial Standards (2015a). Standard JIS B 8616:2015: Packaged air conditioner.

JIS - Japanese Industrial Standards. Standard JIS B 8627:2015 (2015b). Gas heat pump air conditioner.

JIS - Japanese Industrial Standards (2015c). Standard JIS B 8615-Part 3: Multi Split-System Air Conditioners and Air-To-Air Heat Pumps – Testing and Rating for Performance. Japan Industrial Standards.

Miyaoka, Y.; Mori, R.; Giannetti, N.; Jeong, J.; Togashi, E.; Saito, K.(2023). Emulator-type Load-based Testing Methodology for Dynamic Performance Characterization of Air Conditioners. Transactions of the Japan Society of Refrigerating and Air Conditioning Engineers, e.22-17 _OA, February 2023. (in Japanese)

Mehrfeld, P. 2022. Evaluation of heat pump systems under dynamic operating conditions; PhD thesis, RWTH Aachen University, 2022. doi: 10.18154/RWTH-2022-06640.

NEEP. NEEP (2023a). Final Report – Public Version, Residential Heat Pump Efficiency Rating Representativeness Project Phase 1.

NEEP. NEEP (2023b). Working title: Final Report, Residential Heat Pump Efficiency Rating Representativeness Project Phase 2, awaiting publication by NEEP.

Simo, A.; Palkowski, C.; Wachau, A. (2019). *Proposal for the revision of the harmonised test standard EN 14825:2016*. Available: https://netzwerke.bam.de/Netzwerke/Content/DE/Downloads/Evpg/Heizen-Kuehlen-Lueften/proposal_revision_en14825_final.pdf.html

Togashi, E.; Tanabe, S. (2009). (Methodology for developing heat-load calculating class library with Immutable interface), Technical papers of annual meeting, The Society of Heating, Air-Conditioning Sanitary Engineers of Japan, 3 2009 1995-1998. (In Japanese).

Wachau, A.; Mock, D.; Minelli, P.(2023a). *BAM technical report on load-based testing of hydronic heat pumps - compensation method*. Available : https://netzwerke.bam.de/Netzwerke/Content/DE/Downloads/Evpg/Heizen-Kuehlen-Lueften/bam%20report%20load%20based%20testing%20heat%20pumps%20-%20compensation%20method%202023.html

Wachau, A.; Göbel, S.; Vering, C.; Mock, D.; Müller, D. (2023b). *Increasing energy efficiency and acceptance of heat pumps through realistic test procedures – Proof of concept on compensation method*. Available: https://www.dbu.de/OPAC/ab/DBU-Abschlussbericht-AZ-38677_01-Hauptbericht.pdf

Wüllhorst, F.; Maier, L.; Jansen, D.; Kühn, L.; Hering, D.; Müller, D. (2022). BESMod - A Modelica Library providing Building Energy System Modules, American Modelica Conference 2022, Dallas, Oct. 2022.