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

An analysis of domestic heat pumps performance

Jacopo Vivian[a,*], Thibault Péan[b,c], Francesco Benà[a], Glyn Hudson[d], Angelo Zarrella[a]

[a]University of Padua, Dept. of Industrial Engineering, via Venezia 1, Padua, Italy
[b]IREC Catalonia Institute for Energy Research, Jardins de les Dones de Negre 1, 08930 Sant Adrià del Besòs (Barcelona), Spain
[c] UPC Universitat Politècnica de Catalunya, Avgda Diagonal 647, 08028 Barcelona, Spain
[d]OpenEnergyMonitor, Brynrefail, Caernarfon LL55 3NR, Wales, UK

Abstract

The performance of heat pumps is influenced by several factors related to sizing, installation, and actual operating conditions. For instance, oversizing can lead to low partial load factors throughout the year; the type of heat emission systems and the presence of thermal storage systems affect the frequency of start-ups and its performance. Moreover, weather conditions and the manufacturer's control criteria influence the number of defrosting cycles. Due to the interplay of all these factors, the actual performance of heat pumps often deviates from values calculated based on COP figures declared by manufacturers. This paper analyzes the monitored performance of 182 air-to-water and 4 ground-source heat pumps, predominantly installed in residential settings, using data made available by the HeatpumpMonitor project. The results reveal general trends that can be used in preliminary analysis for both retrofit and new installations, providing an overall picture on the expected performance of heat pumps for residential applications. Besides the temperature lift, oversizing seems to have an influence on the variability in the actual seasonal performance, which can deviate significantly from average values.

Keywords: monitored data, data analysis, air-source heat pumps, ground-source heat pumps

1. Introduction

According to the latest European heat pump market report, sales growth across 19 European countries has increased the installed stock from 6.3 million units in 2012 to 25.5 million in 2024 [1]. Today, roughly 12% of European homes use a heat pump: 41% air-to-air, 40% air-to-water, 13% sanitary hot-water, and 4% ground-source systems.

This paper focuses on air-to-water and ground-source heat pumps, examining how these systems actually perform in real operating conditions.

Dongellini et al. [2] simulated three heat-pump types (mono-compressor on/off, multi-compressor, and inverter-driven) in three European climates (Frankfurt, Istanbul, Lisbon), showing that system sizing is a primary driver of annual energy use. Optimal performance was observed when the ratio between nominal heating capacity and building peak load lay between 80% and 120%.

Bruedermüller et al. [3] analyzed more than 5,000 residential heat-pump installations in Switzerland and found that real-world efficiencies (SPF) typically fall below laboratory values. Median SPFs ranged from 2.6-3.4 for air-source and 3.5-4.5 for ground-source systems, with performance largely influenced by flow temperature, sizing, and control strategies. The authors highlight flow-temperature misadjustments and cycling as frequent causes of underperformance, and show that well-tuned controls can bring systems close to their design efficiency.

* Corresponding author. Tel.: +39 049 827 6875
E-mail address: jacopo.vivian@unipd.it

Another field study was recently completed in France [4], surveying 100 heat pump installations in residential buildings. They measured on average a SCOP of 2.9 for air-to-water heat pumps, and 4.3 for water-to-water heat pumps in space heating. They also found out that the monitored SCOP was reduced by 17.3% in comparison with the declared ErP seasonal performance for the average European climate (Strasbourg). The authors highlight potential for improvements in the heating curves configuration, the limiting of high cycling behavior or the more accurate accounting of the off-mode electrical consumption. Heat pumps using R290 as refrigerant, which are gaining market shares following the new f-gas regulation, display 13% higher efficiency. Another study [5] in Spain has corroborated the differences between declared manufacturer efficiency according to standard EN14825, independent laboratory testing in dynamic mode, as well as dynamic simulations in different climates.
Staffell et al. [6] provided one of the earliest comprehensive reviews of domestic heat-pump operation, noting substantial differences between catalog COPs and field values due to climatic and installation-dependent factors. They proposed simplified correlations based on the temperature lift between heat source and sink that remain widely used today. However, given the technological advances of the past 15 years, these correlations may now underestimate the performance of modern systems.
Starting from this hypothesis, the present paper addresses two questions: (i) can Staffell's correlations still be applied in 2026, and if not, how should they be updated to reflect current heat-pump technology? (ii) how strongly does oversizing affect real-world performance?
To answer these questions, we analyze field data from the Heatpumpmonitor database, as described in the following section.

2. **Methodology**

Section 2.1 describes the data used in this paper, Section 2.2 describes the key performance indicators used to analyse the performance of the monitored heat pumps, and Section 2.3 shows how the impact of oversizing on heat pump performance was quantified.

2.1. *Dataset*

HeatpumpMonitor.org is an open-source platform developed by the *OpenEnergyMonitor* community to collect and compare real-world heat pump performance data. As of November 2025, over 600 systems across the UK continuously upload operational data, providing insight into current best practices and achievable efficiencies through proper system design and commissioning. The platform serves both as a benchmarking tool and a learning resource, enabling analysis of factors such as flow temperatures, modulation behavior, and performance variability across models and conditions. Data reveal significant differences among systems, highlighting the impact of control strategies and design choices on efficiency and stability. This open dataset supports installers, researchers, and policymakers in improving heat pump performance and informing design and operational standards. The present paper analyzes a subset of 200 available installations, including 195 air-source heat pumps and 5 ground-source heat pumps during winter 2024/25. The sampling time of the energy meters is 10 s. The metadata includes a lot of details about the installation such as the data about the heat pump size and refrigerant fluid, the building type and annual heat demand, the type of heat emitters, the presence of buffer tanks etc. The data of thermal power at the condenser side, supply water temperature, electrical input and external air temperature have been downloaded using the provided API from 1/10/2024 to 30/4/2025 with 5-minute averaged values and have been further processed to calculate hourly averages. From the initial dataset, only 186 machines have been selected (182 air-source and 4 ground-source) as they presented data holes for less than 10% of the period considered.

2.2. *Performance assessment*

In their review, Staffell et al [6] proposed a very simple model to calculate the COP based on a second-order polynomial of the temperature lift, as shown in Eq. (1). The coefficients were calculated by regression using catalogue data provided by manufacturers.

$$COP = c_0 + c_1\,\Delta T + c_2\,\Delta T^2 \tag{1}$$

Such a simplified method allowed to calculate industry-averaged performance as a function of the temperature lift for two groups of heat pumps: air-source heat pumps ($c_0 = 6.81$, $c_1 = -0.121$, $c_2 = 0.000630$)

and ground-source heat pumps (c_0 = 8.77, c_1 = -0.150, c_2 = 0.000734). The temperature lift ΔT was calculated using the water supply temperature at the condenser outlet for the heat sink. For the heat source, ASHP use drybulb temperature of outdoor air and GSHP use the undisturbed ground temperature, assuming all the considered installations have vertical boreholes. The undisturbed ground temperature has been assumed equal to the annual average of the external air temperature. The latter was retrieved from a *pvlib* library [7] that parses data from PVGIS based on latitude and longitude of the considered installation.

In this paper, we recalculated the coefficients based on the measured data described in the previous section. Therefore, a regression has been carried out for each heat pump installation. This provides a set of coefficients (one for each machine). Then, they have been averaged to have two correlations: one for ASHPs and one for GSHPs. The results Section compares the correlations obtained here to those obtained by Staffell.

For each installation, the seasonal coefficient of performance (SCOP) has also been calculated as the ratio between the heat output on the condenser side and the electrical input. The latter includes the electricity absorbed by the compressor and by auxiliary systems (fans, backup/immersion heaters, circulation pumps etc) as the boundaries of most monitoring systems are classified with label H4.

2.3. *Impact of oversizing*

Since heat pumps capacity decreases with temperature lift (difference between heat sink and heat source temperature), inverter-driven heat pumps work most of the time at partial load. In this sense, they are oversized by design. However, such a problem can be exacerbated in case of overestimation of building heat load or inefficient regulation. A usual way to improve efficiency at partial load consists in reducing supply temperatures and increasing the number of operating hours to reduce intermittent operation.

The impact of oversizing on heat pump performance has been assessed by using two metrics: the number of equivalent full-load hours (FLH) and the oversizing factor (OF). The first indicator, shown in Eq. (2), was calculated by dividing the heating demand of the building (from 1/10/2024 to 30/4/2025) by the nominal heating capacity (rated capacity) of the heat pump. The second indicator, shown in Eq. (3), was calculated by dividing the nominal heating capacity (rated capacity) of the heat pump by the peak heating load of the building. The latter can be taken either from design calculation values based on transmission and ventilation losses (mandatory in the UK for building retrofits) or by regressing values of daily heat load as a function of outdoor air temperature -see Fig. 1. For instance, Figure 1 shows the measured heat loss (each blue dot is a daily value calculated as heat delivered from the heat pumps divided by number of operating hours) as a function of the temperature difference between indoor and outdoor air. Indoor air is an average measured value in case there is a sensor, otherwise is set to a setpoint (20°C). It can be seen that the measured heat loss at design conditions (ΔT = 23 K since design outdoor temperature is -3°C) does not exactly correspond to the calculated value. In the case of this installation the measured heat loss is around 4.4 kW and the calculated heat loss is approximately 5.5 kW. The datasheet of the machine provides a much higher value for the heat capacity at design conditions (7.3 kW) and the badge capacity is rounded to 8 kW. The oversizing factor of this installation would be 1.81 with the measured heat loss (8/4.4) or 1.45 (8/5.5) with the calculated (design) heat load of the building.

$$FLH = \frac{Q_H\ [kWh]}{q_{hp,nom}\ [kW]} \tag{2}$$

$$OF = \frac{q_{hp,nom}\ [kW]}{q_{H,max}\ [kW]} \tag{3}$$

The question is whether these metrics have or not an effect on the seasonal performance of the monitored installations. To this end, the deviation (ε) between measured performance and the performance of the average machine has been calculated, as defined in Eq. (4). A positive (negative) value of ε means that for that particular installation the seasonal performance is better (worst) than the average machine operating under the same boundary conditions (heat source and heat sink temperatures and heating load). The performance of the average machine is calculated with the correlation shown in Eq. (1), where the coefficients have been calculated based on the whole dataset (see Section 3.1 for the exact values).

$$\varepsilon = SCOP - SCOP_{avg} \tag{4}$$

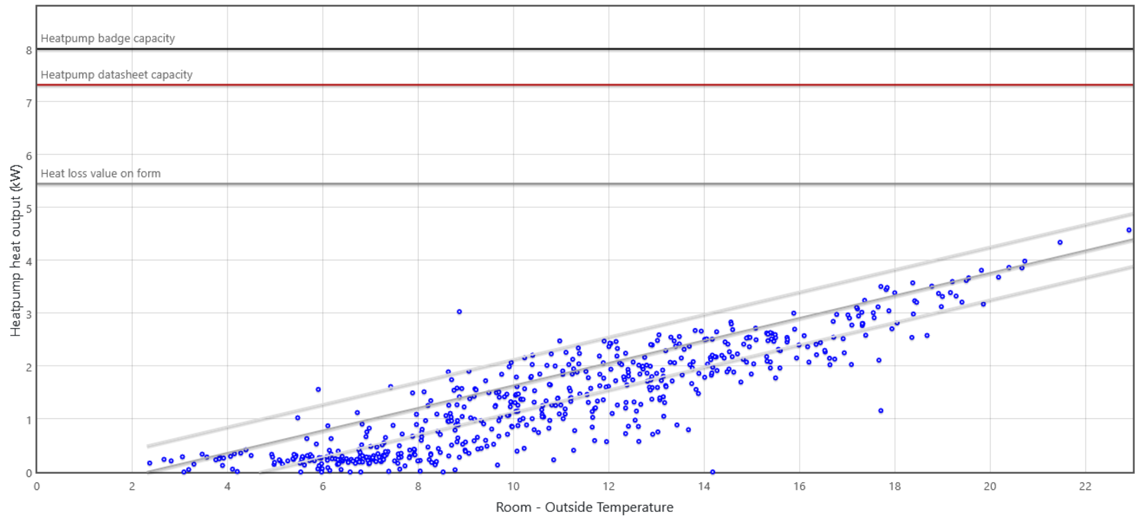

Figure 1. A screenshot from the HeatpumpMonitor website showing the difference between measured and calculated heat loss of the building.

3. Results

The results are divided into two Sections. Section 3.1 analyzes the average performance of air-source and ground-source heat pumps, by comparing measured and rated performance of all the installations considered. Section 3.2 focuses on two factors that drive the difference between actual and expected performance, i.e. the impact of terminal units and oversizing.

3.1. *Performance assessment*

As Staffell correlations depend on the temperature lift, it is worth understanding the two terms that contribute to determine it, i.e. the supply temperature and the source temperature. To this end, Figure 2(a) shows the design supply temperature, which depends on the heat emission system. Four groups of terminal units were detected: old radiators, new radiators, underfloor heating (UFH), and UFH with radiators. Although old radiators have the highest supply temperature as expected, its average value seems to be quite low (42°C). This might be due to oversized radiators in retrofitted buildings, which have high enough heat exchange surface to justify these values. The second category corresponds to situations where new radiators were added or replaced during retrofit, to allow operation at lower supply temperatures. In this case, the average design flow temperature is around 37°C. Underfloor heating is just 2 K lower on average, although the distribution is skewed towards lower values compared to new radiators. The fourth category, as expected, shows an intermediate behavior between UFH-only and new radiators. Figure 2(b) shows the distribution of hours in the different temperature bins for all the HP locations, compared to the climates defined in EN14825 (warmer, average and colder climates). The distribution reveals clearly that, as most installations are UK-based, their climate correspond in general to the average climate, slightly warmer (shifted a few degrees towards the right). A few outliers are present, but none of the location display a warmer or colder climate, as defined by the standard. The expected performance can therefore be reasonably compared with the SCOP rated by the manufacturer, often given for the average climate.

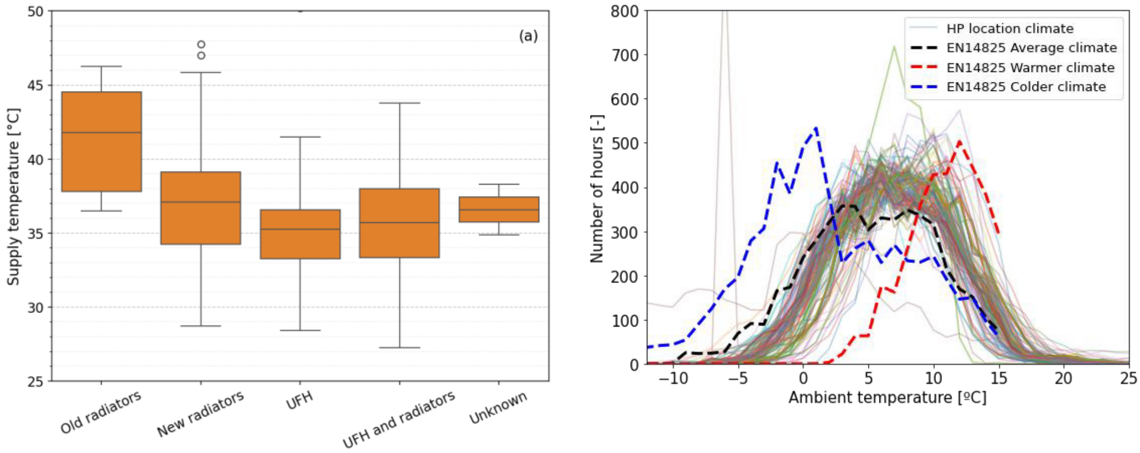

Figure 2. (a) Boxplot with supply temperature according to the terminal unit; (b) distribution of hours according to external air temperature.

Figure 3(a) shows the polynomial curves obtained from fitting 182 air-source heat pump performance data measured by Heatpumpmonitor (transparent lines) and their average curve (in blue). The shadow area represents the variance of the data around the average curve. The dashed line represents Staffell's correlation for air source heat pumps. It can be observed that the monitored heat pumps overperform Staffell's performance curve for ΔT values below 40 K. The distribution of ΔT shown in the histogram in the bottom part of the Figure shows that the most frequently occurring ΔT range is between 25 and 30 K, and that the occurrences with ΔT below 15 K and above 45 K are rather negligible. This means that, on average, the air source heat pumps work with a better performance compared to the one declared by Staffell during 93% of the time (i.e. when ΔT < 40 K), thus confirming our initial hypothesis. For instance, with ΔT = 25 K, Staffell correlation gives a COP of 4.18 while this analysis shows that the average performance of the monitored air-source heat pumps is 4.60. The better performance is probably due to more machines being inverter controlled, thus improving the performance at partial loads.

While the Staffell correlation was deemed representative between 15 and 60 K, the variance bands and the distribution chart shown in Fig. 3 suggest that the validity of the proposed correlation for air-source heat pumps should be shifted between 10 K and 45 K. In this regard, it is worth mentioning that all Heatpumpmonitor installations show flow temperatures below 50°C at design conditions, as shown in Fig. 2. Differently from air-source heat pumps, Figure 3(b) shows that ground-source heat pumps performance from monitored data seems to be overestimated by Staffell's correlation. This might due to the low number of installations considered (only 4), as confirmed by the irregular distribution of ΔT, and highlighting the need for more GSHP data to obtain more meaningful results.

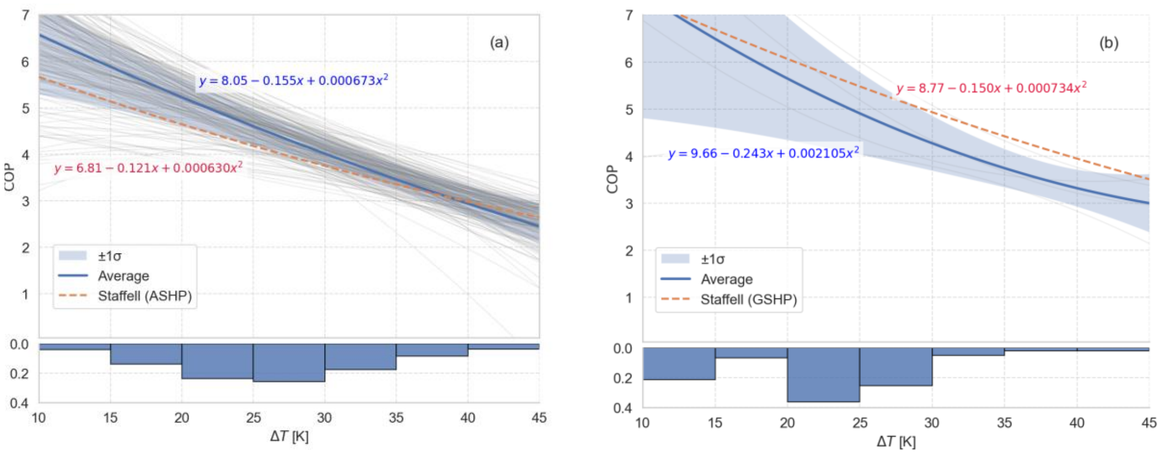

Figure 3. New polynomial correlation vs Staffell's correlation for (a) air-source and (b) ground-source heat pumps.

3.2. *Impact of oversizing*

The distribution of the oversizing factors (OF) is shown in Figure 4, differentiating the OF calculation with respected to the heat loss calculated at design stage and the measured heat load deduced from the data. Especially looking at the OF calculated with measured heat loss, it highlights that most installations are oversized compared to the maximum heat load that they have required during the monitoring period. This can be due to conservative measures, overestimated heat loss calculation, safety margins taken by installers and designers. Most installations are in the range of OF between 1 and 1.7, however some installations are quite importantly overdimensioned, with OF > 2.0. Since the sizing of the heat pump is most probably based on the heat loss calculation at design stage, it is logical that the OF design is more centered around 1, while the measured OF is more spread and really reveals the actual oversizing of the heat pump installations.

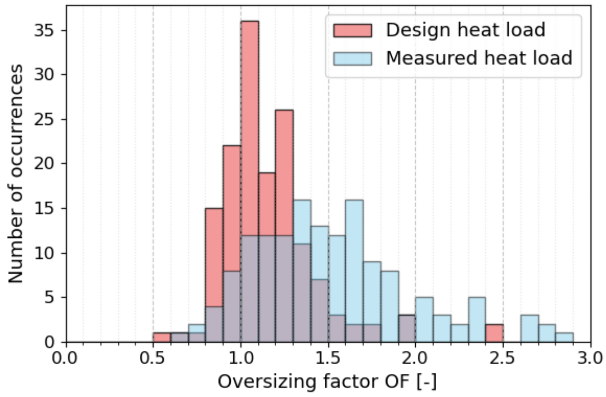

Figure 4. Distribution of oversizing factors calculated with design heat load (red) vs measured heat load (light blue).

Figure 5 shows the deviation of the measured performance compared to the average performance calculated using the correlation of Eq. (1). Therefore, each dot corresponds to an installation. Positive values of ε mean that measured SCOP is higher than the average SCOP, calculated with the average performance correlation. Figure 5(a) shows that, on average, heat pumps with low full load hours lead to lower values of ε. For instance, installations with FLH<1000 tend to have negative ε, while those with FLH>2500 have positive values of ε.

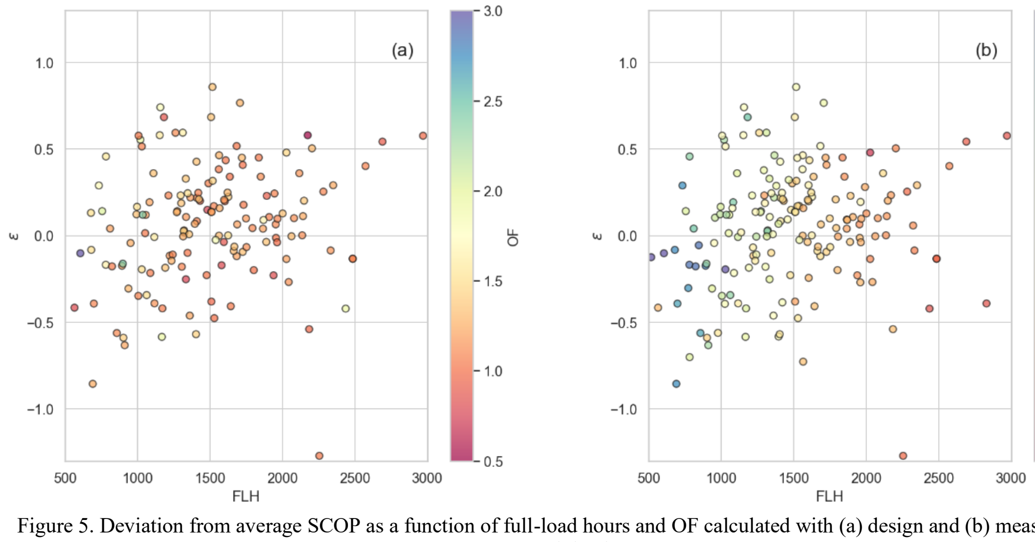

Figure 5. Deviation from average SCOP as a function of full-load hours and OF calculated with (a) design and (b) measured building peak load.

The measured SCOP is represented in Figure 6a in function of the OF. There is not a really evident tendency, as the plot is quite scattered, however the clearly oversized installations display a lower efficiency. To confirm this, the data is split into 3 categories: undersized or normally sized installations with OF below 1.2, slightly oversized installations with OF between 1.2 and 2.0, and oversized installations with OF above 2.0. The performance degradation becomes more evident with a median SCOP of 3.50 for oversized installations compared to 3.92 for undersized or normally sized. This highlights the importance of a proper dimensioning of the heat pumps: if they are too oversized, the important cycling effect and part-load operation leads to SCOP degradation. A slight oversizing (up to OF = 2) does not heavily penalize the unit's performance. This happens because, if the unit is oversized, most of the time il works at low PLR, thus at higher COP whereas, if it is properly sized, the unit operates at design conditions most of the time. Thus, less variability is obtained. Larger OF (>2.0) lead to a drop in performances because the machines work in on-off mode during periods of low heating demand. This result was obtained without considering the effects of thermal storage tank, that could eliminate the inefficiency caused by oversized machines. Moreover, some heat pump models have better internal control algorithms which are able to better handle oversizing. Another important factor that deserves attention is compressor sharing, as sometimes manufacturers share the same compressor across a range of models. As a result, some heat pumps with a low thermal capacity may have a large compressor, thus making these models losing performance at low modulation and increase cycling.

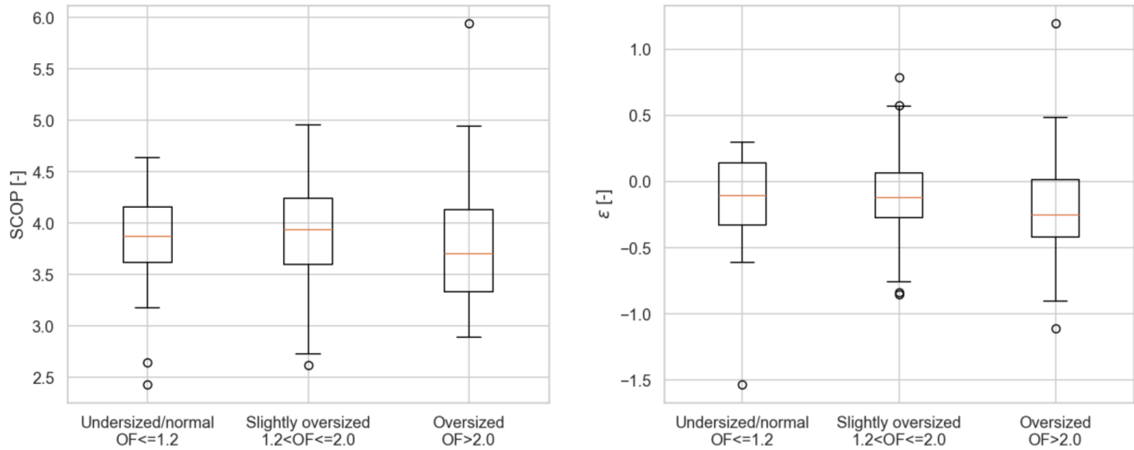

Figure 6. (a) Actual SCOP and (b) deviation from average performance in 3 categories of OF.

4. **Conclusions**

The analysis of 186 monitored heat pump installations provides several insights into the real-world performance of residential systems.

- The measured seasonal coefficients of performance are generally higher than those predicted by Staffell's correlations, reflecting substantial technological progress over the past decade. Improvements in refrigerants, inverter-driven compressors, installation standards and control strategies contribute to better part-load performance and reduced cycling losses. New coefficients have been proposed to reflect such improvements for air-source heat pumps. The limited sample did not allow a reliable regression of ground-source systems.
- Oversizing slightly reduces the average performance and increases its variability: while inverter-driven units mitigate some drawbacks of oversized systems, excessive capacity still leads to unnecessary transients, more frequent cycling, and thus reduced efficiency.

Further work is needed to expand the robustness of the analysis. In particular, the analysis needs to cover ground-source heat pumps and to consider the balance between operational performance and investment costs, which could also be significant in case of oversized machines.

Moreover, current results obtained for oversizing could be affected by the presence of thermal storage effects. Since installations with and without buffer tanks were not considered separately, the impact of oversizing on the measured performance could be in principle underestimated.

The same holds true for user and building load profiles: differences between continuous and intermittent operation driven by occupancy patterns, thermostat strategies, and building thermal inertia appear to influence performance and should be analyzed in greater depth.

Overall, the study confirms both the significant progress made by modern heat pump technologies and is in line with the findings of previous studies dealing with the effect of oversizing.

Acknowledgements

The authors gratefully acknowledge the people behind the OpenEnergyMonitor project for making the heat pump data openly available through the HeatpumpMonitor initiative.

References

[1] G. Uguen and S. Azau, “2025 Heat pump market report,” European Heat Pump Association (EHPA), Brussels, 2025. [Online]. Available: https://www.ehpa.org/wp-content/uploads/2025/07/EHPA-Market-Report-2025-executive-summary.pdf?utm_source=chatgpt.com

[2] M. Dongellini, C. Naldi, and G. L. Morini, “Sizing effects on the energy performance of reversible air-source heat pumps for office buildings,” *Applied Thermal Engineering*, vol. 114, pp. 1073–1081, Mar. 2017, doi: 10.1016/j.applthermaleng.2016.12.010.

[3] T. Brudermueller, M. Kreft, E. Fleisch, and T. Staake, “Large-scale monitoring of residential heat pump cycling using smart meter data,” *Applied Energy*, vol. 350, p. 121734, Nov. 2023, doi: 10.1016/j.apenergy.2023.121734.

[4] Ademe, Enertech, M. Dupret, and U. Bramkamp, “Performance Pompes à chaleur : réalisation d’une campagne de mesure sur des pompes à chaleur en résidentiel individuel.,” ADEME, Sept. 2025. [Online]. Available: https://www.renofute.fr/wp-content/uploads/2025/10/performance_pac_rapport_final_vf2.pdf

[5] T. Péan, J. Romaní, E. Fuentes, and J. Salom, “Seasonal performance and potential of heat pumps in Catalonia,” presented at the 2021 Building Simulation Conference, Sept. 2021. doi: 10.26868/25222708.2021.30523.

[6] I. Staffell, D. Brett, N. Brandon, and A. Hawkes, “A review of domestic heat pumps,” *Energy & Environmental Science*, vol. 5, no. 11, p. 9291, 2012, doi: 10.1039/c2ee22653g.

[7] A. R. Jensen *et al.*, “pvlib iotools—Open-source Python functions for seamless access to solar irradiance data,” *Solar Energy*, vol. 266, p. 112092, Dec. 2023, doi: 10.1016/j.solener.2023.112092.