

## Secondary Publication



Brudermueller, Tobias; Potthoff, Ugne; Fleisch, Elgar; u. a.

### Estimation of energy efficiency of heat pumps in residential buildings using real operation data

Date of secondary publication: 17.10.2025

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-110865x

#### Primary publication

Brudermueller, Tobias; Potthoff, Ugne; Fleisch, Elgar; u. a. (2025): Estimation of energy efficiency of heat pumps in residential buildings using real operation data, in: Nature Communications, London: Springer Nature, vol. 16, Nr. Article number: 2834, pp. 1–15, doi: 10.1038/s41467-025-58014-y

#### Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>

# Estimation of energy efficiency of heat pumps in residential buildings using real operation data

Received: 31 January 2024

Accepted: 4 March 2025

Published online: 22 March 2025

 Check for updates

Tobias Brudermueller <sup>1</sup>✉, Ugne Potthoff <sup>1</sup>, Elgar Fleisch <sup>1,2</sup>,  
Felix Wortmann <sup>2</sup>✉ & Thorsten Staake <sup>1,3</sup>

As heat pumps become more prevalent in residential buildings, effective performance monitoring is essential. Design flaws, incorrect settings, and faults can escalate energy consumption and costs, leading to discrepancies in user expectations and hindering the widespread adoption of this technology crucial for the heating transition. However, field studies using large data sets to offer insights into real-world performance and methods for identifying low-performing systems in practical, scalable applications are lacking. In the largest field study to date, we analyze sensor data from 1023 heat pumps across Central Europe monitored over two years. Based on existing approaches for controlled laboratory conditions, we derive methods to evaluate and classify real-world performance using operational data. Applying these methods, we find that 17% of air-source and 2% of ground-source heat pumps do not meet existing efficiency standards. Additionally, around 10% of systems are oversized, while approximately 1% are undersized. This underscores the need for standardized post-installation performance evaluation procedures and digital tools to provide actionable feedback for users and installers to enhance operational efficiency and guide future installations.

Buildings constitute 30% of global final energy consumption and contribute to 26% of global energy-related carbon dioxide emissions, with approximately half attributed to space and water heating<sup>1</sup>. Electrically-powered heat pumps (HPs), extracting heat energy from natural sources such as the ground, air, or water, offer a sustainable alternative to oil or gas-based heating, especially in regions with a high share of renewable electricity generation<sup>2</sup>. While already meeting 10% of global space heating needs in 2021, HPs have the potential to reduce global carbon dioxide emissions by at least 500 million tonnes by 2030, equivalent to the annual emissions from all cars currently in operation in Europe<sup>3</sup>. Yet, meeting the International Energy Agency's global non-binding target of 600 million HPs by 2030 necessitates an accelerated deployment of HPs, as current installation rates project a 58% shortfall<sup>4</sup>. The replacement of fossil fuel heating systems however

poses a significant financial challenge for homeowners due to the high upfront costs, even with available subsidies and the potential for better cost amortization through achieved savings<sup>5,6</sup>. Subsidies are prevalent in over 30 countries<sup>1</sup> but also represent a substantial financial burden for governments<sup>4,7,8</sup>. Moreover, regulations regarding heating systems in private households are sometimes entangled with emotional responses<sup>9</sup>. The German Building Energy Act serves as a notable example, reflecting public discontent arising from being compelled to invest in heating renovation<sup>5,10–12</sup>.

Additionally, HPs exhibit greater complexity compared to well-established gas and oil heating systems, and unlike these traditional systems, they have not undergone decades of optimization. The performance of HPs is influenced significantly by factors beyond design, such as occupant characteristics and HP system settings<sup>13–17</sup>, which is a

<sup>1</sup>Chair of Information Management, ETH Zurich, Weinbergstrasse 56/58, Zurich, Switzerland. <sup>2</sup>Institute of Technology Management, University of St. Gallen, Dufourstrasse 40a, St. Gallen, Switzerland. <sup>3</sup>Chair of Information Systems and Energy Efficient Systems, University of Bamberg, An der Weberei 5, Bamberg, Germany. ✉e-mail: [tbrudermuell@ethz.ch](mailto:tbrudermuell@ethz.ch); [felix.wortmann@unisg.ch](mailto:felix.wortmann@unisg.ch)

challenge for manufacturers, installers and owners. Consequently, the actual energy consumption of HPs in practice can deviate significantly from expectations, resulting in substantial additional operating costs. For example, Nolting et al.<sup>15</sup> report up to 24% lower performance than stated on the product certificate label. Additionally, in an analysis of 297 Swiss households with HPs, Weigert et al.<sup>18</sup> demonstrated that after on-site optimization by an energy consultant, half of them achieved average savings of 1805 kWh (15.2%) per year. As the operating costs of HPs are decisive in determining whether the technology is economically attractive compared to other heating solutions<sup>19–21</sup>, the discrepancy in performance can fuel dissatisfaction and, ultimately, pose a threat to the technology's acceptance<sup>22</sup>. The success of the heating transition is therefore closely tied to the performance of heat pumps in the field.

Maximizing the energy efficiency of HPs is also relevant for electricity grids, as these systems significantly increase both total and peak power demand. This heightened demand can necessitate the implementation of demand response programs and expensive upgrades to grid infrastructure<sup>23,24</sup>. For example, transitioning 10% of British households to HPs would increase peak demand by 2.5–3.75 GW (4.6–7.0%)<sup>25,26</sup>, and a transition of all British households would double it<sup>27</sup>. Similar studies are also available for other countries and scenarios<sup>28–32</sup>. In Switzerland, even with substantial already existing pump storage facilities that help control demand, a 34% increase in electricity storage capacity would be necessary if all fossil-based heating systems were replaced by HPs<sup>29</sup>.

Digitalization offers opportunities to tackle the current challenges of HP operations. As the majority of modern HP units are equipped with multiple sensors providing real-time data, it becomes possible to monitor their performance and control their operation effectively<sup>24,33,34</sup>. However, manufacturers are still in the early stages of offering services beyond displaying raw consumption data and activating alarms in case of heating failures<sup>35</sup>. While fault detection and diagnosis systems have been thoroughly explored in the literature<sup>36</sup>, there is a notable gap in research focusing on HPs operating without faults but potentially lacking optimization. Moreover, many existing studies lack the analysis of large sample size data sets of HPs in real-world settings, often relying on data from test houses, simulations, or laboratory experiments. For instance, Carroll et al.<sup>37</sup> conducted a structured literature review of field studies on air-source heat pumps and identified only 34 articles, with<sup>38</sup> being the largest field study with a sample size of 77. A more expansive and frequently cited field study was conducted by the UCL Energy Institute between 2013 and 2015, encompassing 292 air-source heat pumps and 92 ground-source heat pumps in the UK as part of the Renewable Heat Premium Payment (RHPP) program<sup>39</sup>. In addition to the limited number of field studies, there is a scarcity of studies developing methods to identify low-performing HPs in practical applications. Offering users individualized feedback about the energy efficiency of their HPs, however, has proven to significantly impact user satisfaction and can substantially enhance the acceptance of the technology<sup>22</sup>. Therefore, to improve the energy efficiency and reduce operational costs of HPs in real-world scenarios, gaining a profound understanding of their current performance and identifying systems with optimization potential, is imperative.

Despite this research gap, relevant studies in related domains can be classified into three categories: comparing real-world performance to product certificates, identifying underperforming states in individual systems, and developing optimal control strategies for individual systems. The first group of work compares in-situ performance of small sample sizes with corresponding benchmarks given in data sheets and finds significant differences between laboratory and on-site performance, e.g.,<sup>15,40,41</sup>. As explained by O'Hegarty et al.<sup>41</sup>, European regulations define fixed operational points for the evaluation and

reporting of performance in standardized product certificates. Therefore, the contribution of this type of related work lies mainly in the development of complex interpolation and extrapolation methods to compare real-world performance to product certificates. Although this type of study is able to identify HPs where performance largely deviates from specifications, it requires contextual information (e.g., the exact model of the HP) and highly detailed measurements, which makes it unsuitable for mass market applications. In contrast, the second category examines performance of individual systems without further reference, e.g.,<sup>42–45</sup>. This group of studies describes performance under perfect knowledge of the buildings and without comparison of HPs in large populations. If methods for detecting underperformance are proposed (e.g., in<sup>44</sup>), the focus is on detecting periods of inefficient operation or performance degradation of single systems. Lastly, the third group of work targets optimal control strategies for individual systems to improve their energy performance, e.g.,<sup>46–49</sup>.

The studies outlined above are not applicable to practical scenarios, as they fail to address variations in data availability, observation periods, and building characteristics. Moreover, these studies do not offer methods for identifying low-performing HPs within their specific installation environments. In practical settings, relevant contextual information such as building type, occupancy levels, and more, is typically unavailable. Additionally, HPs may not be controllable to operate precisely at the operational points defined in regulations, thereby hindering comparisons to known performance values derived from idealized laboratory conditions. This lack of control may be due to technical constraints or concerns about regular adjustments impacting occupant comfort. As a result, evaluating HP performance in real-world applications must rely solely on observations of actual operational conditions, without access to detailed contextual knowledge or the ability to directly interfere with its operation. Such assessments are critical in practice and are highly desired by HP users<sup>22</sup>, yet effective methods for conducting them have not been established previously. Additionally, clear benchmarks for good and poor performance in practical applications are lacking.

In this study, we address the existing gap by developing methods to evaluate HP performance in the field post-installation using real operational data. Additionally, we provide insights from a comprehensive performance analysis of 1023 HPs installed in residential buildings. Our findings reveal significant variability in performance among individual HPs, with a 2–3 fold difference between the lowest and highest efficiency systems. Moreover, 17% of air-source and 2% of ground-source HPs fall short of existing European efficiency standards. Approximately 10% of systems are oversized, while about 1% are undersized. These results highlight the critical need for standardized post-installation performance evaluation procedures and the development of digital tools to deliver actionable feedback for users and installers, ultimately improving operational efficiency and informing future installations.

## Results

### Real-world data set

The analyzed data encompasses a wide variety of HP models and configurations installed in residential buildings across 10 countries in Central Europe. Since the data originates from a single manufacturer, we acknowledge that our results should be further validated with data from HPs produced by other companies to ensure broader applicability and generalizability. Nonetheless, this study represents the largest field study conducted to date on the energy efficiency of HPs in residential buildings. The data set studied covers 1,023 HPs monitored between 2021-03-14 and 2023-04-30 (i.e., for up to 777 days), with 890 (87%) being air-to-water HPs and 133 (13%) brine-to-water HPs. There are no water-to-water HPs in the data set. While the descriptive

analyses encompass all systems, other analyses are limited to HPs with appropriate data to avoid distortions from poor model fits. Most analyses include around 600 to 700 systems, with the exact number of samples reported in each subsection. While other contextual information is unavailable, it is known that all HPs are installed in residential buildings in Germany: 434 (42.42%), the Netherlands: 211 (20.63%), Austria: 204 (19.94%), Czech Republic: 78 (7.62%), Sweden: 46 (4.50%), Denmark: 35 (3.42%), Poland: 3 (0.29%), Slovenia: 2 (0.20%), France: 1 (0.10%), Great-Britain: 1 (0.10%), and unknown: 8 (0.78%). In total, the data set contains 185,139 daily observations at outdoor temperatures of 15 °C or below, with each HP having an average of 182.22 days of data within this temperature range.

Each HP is connected to the internet and measures multiple parameters with a temporal resolution of a few seconds. To reduce complexity, in this study, we analyze daily aggregates of this data, using daily sums for electrical energy consumption and thermal energy production, and daily averages for all other parameters. Days with data gaps have been systematically excluded to ensure the highest attainable data quality and are not counted in the number of observations mentioned above. Days with more than three hours (12.5%) of missing measurements in total are removed. Further, any day lacking measurements of outdoor and supply temperatures, energy input, or energy output is neglected. Furthermore, measurements are categorized based on the operating modes for domestic hot water (DHW) production and space heating (SH). Since all HPs are used for SH but not all are used for DHW production, our analyses focus on SH to ensure comparability. The performance metrics reported in this study encompass final energy usage of the compressor, the fan or brine pump, and the electrical backup heater. This aligns with the European standard EN 14825<sup>50</sup> and adheres to the  $H_3$  system boundary taxonomy outlined in<sup>41,51</sup>. Further, it is worth noting that the energy values are not directly measured by energy sensors; rather, they are computed by the HPs themselves using operational sensors and the principles of physics. This computation relies on parameters such as pressure, volume flow, and power measurements and is common practice in most modern HPs<sup>15</sup>.

Due to potential errors in measurements, particularly during minor compressor modulations, inaccuracies in the recorded HP performance can occur. The specifics of the sensors used are unavailable to us, precluding detailed calculations of measurement uncertainties. However, a draft proposing updates to regulations by the European Union concerning HPs<sup>52</sup> suggests that the maximum permissible error for energy output should range from 7.5% to 15%, depending on the temperature difference. For energy input, a maximum permissible error of 5% is proposed. According to the HP manufacturer, the errors in the data under study already fall within these tolerances.

### Modeling and evaluating heat pump performance

This section provides essential fundamentals to support the descriptions of the results and outlines the methods developed in this study, which are designed for post-installation performance evaluation in practical applications.

**Explaining Carnot efficiency.** Our analyses rely on the coefficient of performance (COP) as a key metric for evaluating the efficiency of an HP, which is the ratio of thermal energy generated to electrical energy consumed in a fixed observation period. This metric is not constant and affected by operational conditions, as comprehensively reviewed in<sup>53</sup>. The maximum efficiency theoretically achievable by an HP is defined by the Carnot cycle and depends on the difference between the heat source temperature  $T_{\text{hsourse}}$  and the heat supply temperature  $T_{\text{hsupply}}$  in Kelvin. In practice, however, HPs typically operate at around half of their theoretical maximum efficiency or even lower, influenced by irreversible and non-ideal effects extensively studied in the

literature<sup>42,54,55</sup>. These effects can be represented by a correction factor  $\zeta$ , defining the COP as:

$$\text{COP} = \zeta \cdot \frac{T_{\text{hsupply}}}{T_{\text{hsupply}} - T_{\text{hsourse}}} \quad (1)$$

From Equation (1), it can be inferred that HPs are efficient when the temperature difference is small, achieved by using low flow temperatures for the water distributed by the HP to a space or system<sup>40</sup>. An in-depth analysis of the underlying reasons for the values of  $\zeta$  in practical applications is not the focus of this study. However, a comparison of observed COP values with Carnot efficiency and other models found in the literature is provided in Supplementary Note 2. Additionally, we note that several other factors are known to affect HP performance, such as the frequency of on-off transients<sup>14</sup>, the quantity of defrosting cycles<sup>53</sup>, the speed of the compressor<sup>56</sup>, and variations in temperature profiles<sup>44</sup> or part-load conditions<sup>57</sup>.

**Explaining part-load ratio and capacity ratio.** The performance of HPs, as reported in product certificates, assumes a fixed part-load ratio (PLR) at different operating points. The PLR is the ratio of the heating load at a specific temperature ( $T_j$ ) to the design heating load at the design temperature ( $T_{\text{design}}$ ), under the assumption of a linear relationship with outdoor temperature above the heating limit temperature ( $T_{\text{lim}}$ )<sup>58</sup>. According to EN 14825<sup>50</sup>, for an average climate,  $T_{\text{design}}$  is assumed to be -10 °C, for warmer climate it is 2 °C and for colder climate it is -22 °C. The heating limit temperature  $T_{\text{lim}}$  is assumed to be 16 °C. As formulated by Sieres et al.<sup>58</sup>, the PLR is given by:

$$\text{PLR}(T_j) = \begin{cases} (T_j - T_{\text{lim}})/(T_{\text{design}} - T_{\text{lim}}) & \text{if } T_j < T_{\text{lim}} \\ 0 & \text{if } T_j \geq T_{\text{lim}} \end{cases} \quad (2)$$

Another metric, closely related but not identical to the PLR, is the HP's capacity ratio (CR). While the PLR is independent of the HP capacity, the CR represents the HP's output capacity at  $T_j$  relative to its full load capacity. Consequently, depending on the design choice of the HP's full load capacity, the CR line may lie above or below the PLR line, but likely remains close to it.

**General approach for modeling HP performance.** Data availability can vary significantly among HPs in terms of observation periods and operation at different temperatures. Therefore, we assess and ensure comparability of HP performance by modeling each system's behavior and performance based on its in-situ measurements, facilitating simulation and evaluation. We accomplish this by fitting linear mixed-effects models, which include fixed effects for all HPs (including slope and intercept) and individual random effects (also including intercept and slope). These random effects capture the individual deviations of each HP from the mean of all systems. In the following sections, we denote a parameter associated with random effects using a superscript  $i$ , where  $i$  indexes a specific HP. As we proceed with the models defined below, several models based on existing literature were tested, with detailed results presented in the Methods section. Further note that the models are fitted and evaluated using observations from the SH mode exclusively, which is the primary application of focus (i.e., DHW is not included). The fitted model parameters for each individual HP are provided as supplementary material, enabling future studies to conduct simulations based on real-world data rather than product certificates. Finally, it is important to clarify that all subsequent models are fitted using the complete data available for each HP. However, to ensure robustness, additional tests were conducted by splitting the data for each system into training and test sets (see Supplementary Note 1). The model performance was then evaluated solely on the test data that was not seen during training. Since the differences in model fits however were minimal, we chose to use the models fitted with the

**Table 1 | Mean and standard deviation scores (in brackets) of the fits for individual models**

Model Type	MdAE	MAE	MSE	RMSE	MAPE	SMAPE	$R^2$
Heating Curve	1.02 (0.69)	1.20 (0.75)	2.99 (4.58)	1.48 (0.90)	3.32 (2.06)	3.30 (2.02)	0.43 (0.27)
COP	0.19 (0.11)	0.24 (0.13)	0.12 (0.14)	0.30 (0.18)	5.65 (2.48)	5.58 (2.38)	0.75 (0.17)
Heating Curve & COP	0.23 (0.13)	0.28 (0.15)	0.16 (0.18)	0.36 (0.19)	6.63 (2.79)	6.66 (2.83)	0.51 (0.31)
Utilization	4.66 (1.87)	5.59 (2.11)	57.48 (52.0)	7.11 (2.64)	21.18 (0.13)	17.49 (7.06)	0.65 (0.19)

These models are used in subsequent analyses and cover 708 heat pumps for the heating curve and coefficient of performance (COP) model, and 637 heat pumps for the utilization model.

entire data set to enhance the interpretability of the subsequent analysis.

**Modeling the heating curve.** The heating curve defines the supply temperature  $T_{\text{supp}}$  as a linear function of the outdoor temperature  $T_{\text{out}}$ , which most heating controllers allow to be set manually and is known to have a significant impact on performance<sup>17</sup>. Incorporating fixed and random slope and intercept terms, our heating curve model is expressed as:

$$T_{\text{supp}}^i(T_{\text{out}}) = (a_0^i - 0.270) \cdot T_{\text{out}} + (a_1^i + 38.244) \quad (3)$$

When comparing the fixed intercept and slope values to other heating curve models in the literature (e.g.,<sup>59,60</sup>), it becomes evident that these values can be interpreted in the context of a mixed distribution system involving radiators and floor heating. The individual models of each HP either exceed or fall below these baseline values adjusted by the random effects, as they are influenced by their respective distribution system and building insulation level, details of which are unknown in this study.

**Modeling the coefficient of performance.** Theoretically, a COP model could be designed to directly capture deviations from Carnot efficiency. However, this approach is impractical because when there are small differences between the heat source and heat sink temperatures, the denominator of Equation (1) becomes very small. This results in unrealistically high Carnot efficiency values that do not reflect real-world performance. Instead, several studies model the COP as a quadratic or linear function<sup>44,59,61,62</sup>. They either use the outdoor-to-supply temperature difference as a single independent variable or consider outdoor and supply temperatures separately as two independent variables. Note that we use outdoor temperature instead of brine temperature, even for ground-source heat pumps, as they also exhibit a dependence on outdoor temperature. We adopt this approach due to higher data availability, and to eliminate the potential influence of borehole depth on brine temperature measurements. While Fischer et al.<sup>61</sup> and Pospíšil et al.<sup>62</sup> utilize values from multiple HPs reported at operational points in product certificates, Sun et al.<sup>44</sup> employ real measurements but only from a single HP. However, no study has modeled COP using large sample size data sets from multiple HPs in the field. The COP model that performs best on our data set is a simple linear function, given by:

$$\text{COP}^i(T_{\text{out}}, T_{\text{supp}}) = (b_0^i + 0.098) \cdot T_{\text{out}} + (b_1^i - 0.104) \cdot T_{\text{supp}} + (b_2^i + 6.965) \quad (4)$$

**Modeling utilization as approximation for capacity ratio.** We evaluate the sizing of an HP based on its utilization. To this end, we use the compressor speed of an HP relative to its full-speed capability, expressed as a percentage, as an approximation of an HP's capacity ratio. Since the data set consists of daily aggregates, the average compressor speed for each day is calculated, encompassing total usage, i.e., it includes both space heating and domestic hot water modes. During periods of inactivity, the compressor speed is recorded

as 0%. Similar to the heating curve model, we fit a linear mixed-effects model to describe the utilization of each HP indexed  $i$  as a function of the outdoor temperature  $T_{\text{out}}$ , given by:

$$\text{Utilization}^i(T_{\text{out}}) = (c_0^i - 2.739) \cdot T_{\text{out}} + (c_1^i + 50.865) \quad (5)$$

**Evaluating model fits.** To ensure that the interpretation of results is not distorted by potentially poorly fitted models, we only evaluate HPs where the models provide an appropriate fit and where the mixed-effect slopes and intercepts accurately reflect physical properties. Hence, we consider only those HPs where the SMAPE score falls within the interquartile range and HPs with  $\text{SMAPE} \geq Q_3 + 1.5 \cdot (Q_3 - Q_1)$  are excluded from the analysis. A root-cause analysis of the reasons for poor model fits of individual HPs is beyond the scope of this paper but could be explored in future research. For example, as outdoor temperatures increase, the supply temperature and utilization of an HP must decrease, while the COP must increase. Due to this condition, 125 HPs (12.21%) are excluded from the heating curve and COP model analysis. Additionally, 190 HPs (18.57%) are excluded due to insufficient data, having fewer than 10 observations of supply temperatures and COP at outdoor temperatures below or equal to 15 °C. As a result, 708 HPs are used for energy efficiency evaluations and for calculating the effects of minor heating curve adjustments. Similarly, for analyses involving utilization models, 174 HPs (17.01%) lacked at least 10 measurements of average compressor speed at outdoor temperatures below or equal to 15 °C, and 212 HPs (20.72%) exhibited an insufficient model fit. Consequently, 637 HPs are included in the sizing evaluations. In contrast, the descriptive analyses in subsequent sections encompass all 1,023 HPs.

Table 1 shows the fits of the corresponding regression models. The values represent the mean and standard deviations of the individual scores of each HP included in the subsequent analyses. Note that, in addition to the individual models, we also provide the score for a combined model, where predictions from a heating curve model replace the original  $T_{\text{supp}}$  measurements as inputs for a COP model. This approach enables simulations that rely solely on outdoor temperature data. For completeness, the  $R^2$  value is also provided, indicating the variance in the data explained by the model. However, note that a small variance in the data can result in a low  $R^2$  value without necessarily indicating a poor fit.

**Calculating the seasonal coefficient of performance.** The European standard EN 14825<sup>50</sup> outlines a procedure for calculating the performance of an HP using a single metric known as the seasonal coefficient of performance (SCOP). This standard specifies a set of temperatures and corresponding weights to represent typical temperature conditions across three different climate zones: average, colder, and warmer. For the HPs in Sweden, the Czech Republic, Poland, and Slovenia, we use values corresponding to colder climate conditions. For the single HP in our data set located in France, we use values for warmer climate conditions, and for all other HPs, we assume average climate. The SCOP is determined by taking a weighted average of COP values at these predefined temperatures and is the metric reported on product labels. In addition to outdoor temperatures, the standard specifies

fixed supply temperatures and PLRs. These conditions are rarely met in practical applications without explicit intervention in HP operation, making it impractical to calculate SCOP as strictly defined by the standard<sup>41</sup>. Instead, to accurately assess the real-world performance of HPs, we calculate SCOP under real PLRs and using real supply temperatures obtained from in-situ measurements. We achieve this by sampling from each HP's heating curve and COP model (Equation (3) and Equation (4)) using the fixed outdoor temperatures  $T_{out}^j$  and corresponding weights  $w^j$  as defined in EN 14825<sup>50</sup> (see Supplementary Table 1). The real SCOP of an HP, indexed by  $i$ , is thus calculated as follows:

$$SCOP_{real}^i = \frac{\sum_j (w^j \cdot COP^i(T_{out}^j, T_{supp}^i(T_{out}^j)))}{\sum_j w^j} \quad (6)$$

According to O'Hegarty et al.<sup>41</sup>, the value calculated here is comparable to the  $SPF_{H3}$  reported in other studies, but it pertains only to space heating. For completeness, we also report results using the fixed supply temperatures defined in the standard. Although these may not accurately reflect real operating conditions, the calculated SCOP values are closer to those reported on product certificates. In this approach, no sampling from the heating curve is needed; instead, values can be directly sampled from the COP model using the fixed temperatures. Note that in this case, for ground-source heat pumps, the outdoor temperature is fixed at 0 °C, with only the supply temperatures varying, while we continue to use the actual part-load conditions. The complete definition of test points is provided in Supplementary Table 1.

**Simulating minor adjustments to the heating curve.** We simulate a reduction of the heating curve by simply subtracting 1 °C from the intercept. By combining the adjusted heating curve with the original COP model and applying Equation (6), we calculate a new SCOP value. This empowers users and installers to assess the impact on HP efficiency when maintaining the same heat output with lower supply temperatures, providing valuable guidance for optimizing settings. Moreover, this adjustment can be quantified in terms of energy consumption, enhancing its interpretability. Assuming the heat demand  $Q_{heat}$  is known, the difference in electricity consumption resulting from a change in the heating curve can be approximated by a function of the old and new SCOP, expressed as:

$$\Delta E = E_{new} - E_{old} = \frac{Q_{heat}}{SCOP_{new}} - \frac{Q_{heat}}{SCOP_{old}} \quad (7)$$

In practice, however,  $Q_{heat}$  may not be precisely known due to potential gaps in the measured data. Therefore, we calculate a percentage change relative to the old energy consumption, eliminating the dependency on the exact heat demand as follows:

$$\frac{\Delta E}{E_{old}} \cdot 100\% = \frac{SCOP_{old} - SCOP_{new}}{SCOP_{new}} \cdot 100\% \quad (8)$$

### Describing the observed performance of all heat pumps

Figure 1 illustrates the COP values and their temperature dependence across all HPs in our data set, showing only outdoor temperatures at or below 15 °C. This upper limit is consistent with other studies on HP performance, such as<sup>60</sup>, and aligns with the European standard EN 14825<sup>50</sup>. We distinguish between air-source heat pumps (ASHPs) and ground-source heat pumps (GSHPs), as well as observations related to the operating modes of SH and DHW production. In the graph, the quantity of observations among several HPs is presented at the top ( $N$ ), while the count of HPs is provided at the bottom ( $N_{HP}$ ). Due to different data availability between operating modes, these figures differ across subsets of samples. The graph reflects two insights that align with

common knowledge in HP literature. The first observation is that GSHPs are generally more efficient than ASHPs - in this case, by approximately 22%, with a mean COP of 4.90 compared to 4.03. This difference is statistically significant at a 99% confidence level, as indicated by the Welch t-test (statistic = -96.28, p-value = 0.0). The reason is that GSHPs do not need to perform defrosting cycles and benefit from stable and higher ground temperatures on cold days, which, although correlated with air temperatures, do not vary as significantly<sup>63,64</sup>. This is also evident in the contour plots shown in Fig. 1c) and d), where a linear interpolation on 500 levels of observed COP over outdoor and supply temperatures is depicted. The efficiency of GSHPs is influenced by outdoor temperature, though not to the extent seen in ASHPs. For GSHPs, the correlation between outdoor temperatures and COP in SH mode is 0.42, whereas for ASHPs, it is 0.49. The second observation is that the COP values are statistically significantly higher for SH compared to DHW, primarily because DHW requires higher flow temperatures<sup>13</sup> (Welch t-test: statistic = 340.81, p-value = 0.0). We address these differences by exclusively comparing HPs of the same type and modeling each HP individually. Furthermore, later assessments of energy efficiency specifically focus on applications related to SH.

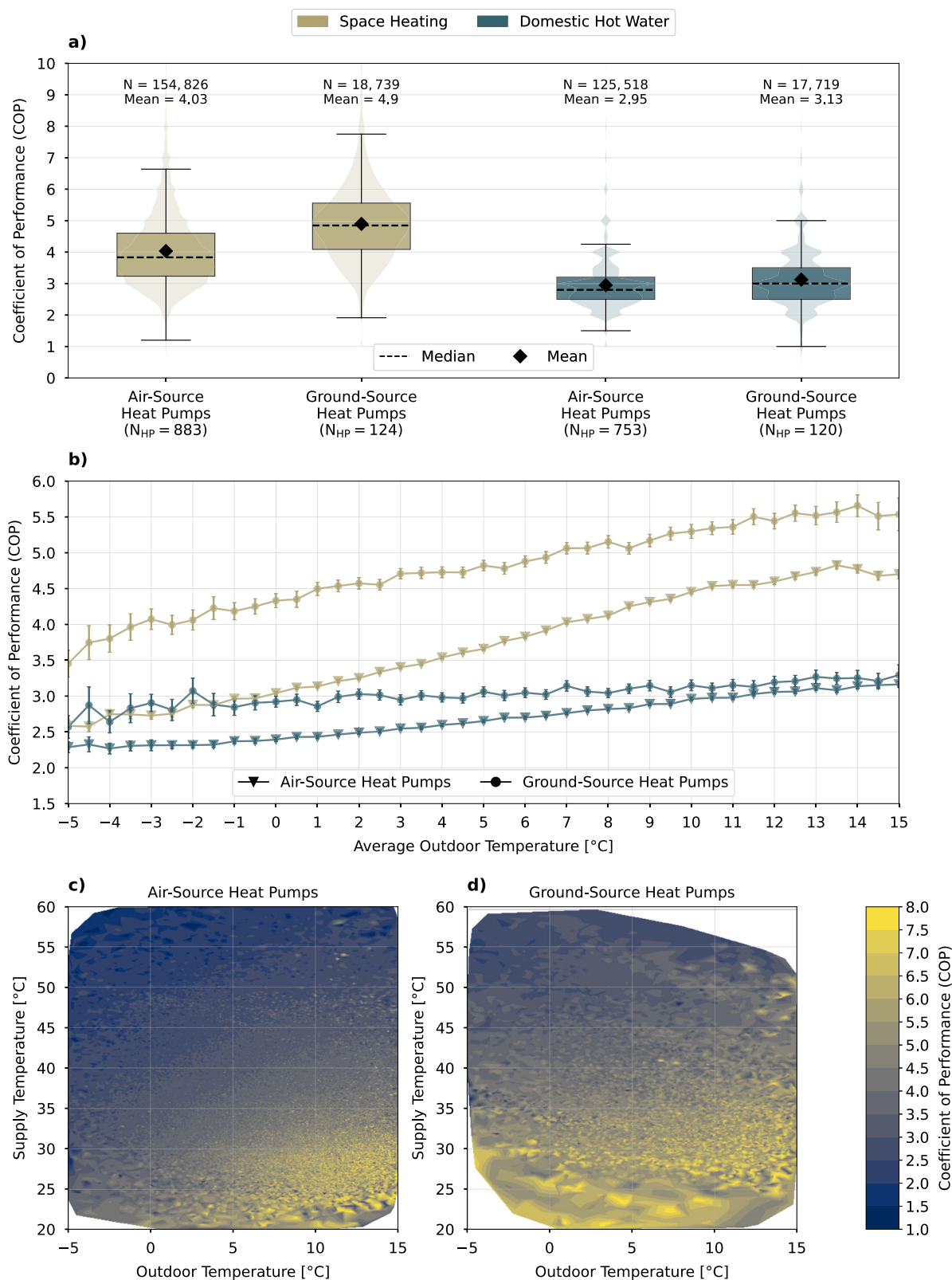
### Performance differences among individual heat pumps

Figure 2 visualizes the performance of HPs, considering the differences between individual systems (also see Supplementary Figs. 1 and 2 for additional graphs). For each specific HP, the median of COP values per operating mode was computed based solely on observations within a defined outdoor temperature range. The charts in Fig. 2 thus display histograms, where each vertical bar represents the distribution of individual HPs within a particular temperature range.

In this context, the  $N$ -values below the bars indicate the number of HPs used to calculate the proportions. This approach offers a comprehensive overview of the diverse performance and behavior of individual HPs in practice. For instance, while 18.3% of ASHPs still achieve a median COP of 3.0-3.5 at -6 to -3°C, 11.2% fall below 2.0 in this temperature range. Similarly, 11.5% of GSHPs reach a median COP above 5.5 in the temperature range of -3 to 0°C, while an equal percentage fall within the range of 3.0 to 3.5. Thus, HPs can exhibit significant variations in performance, sometimes differing by a factor of 2 to 3 even within the same temperature range, which underscores the importance of identifying low-performing systems.

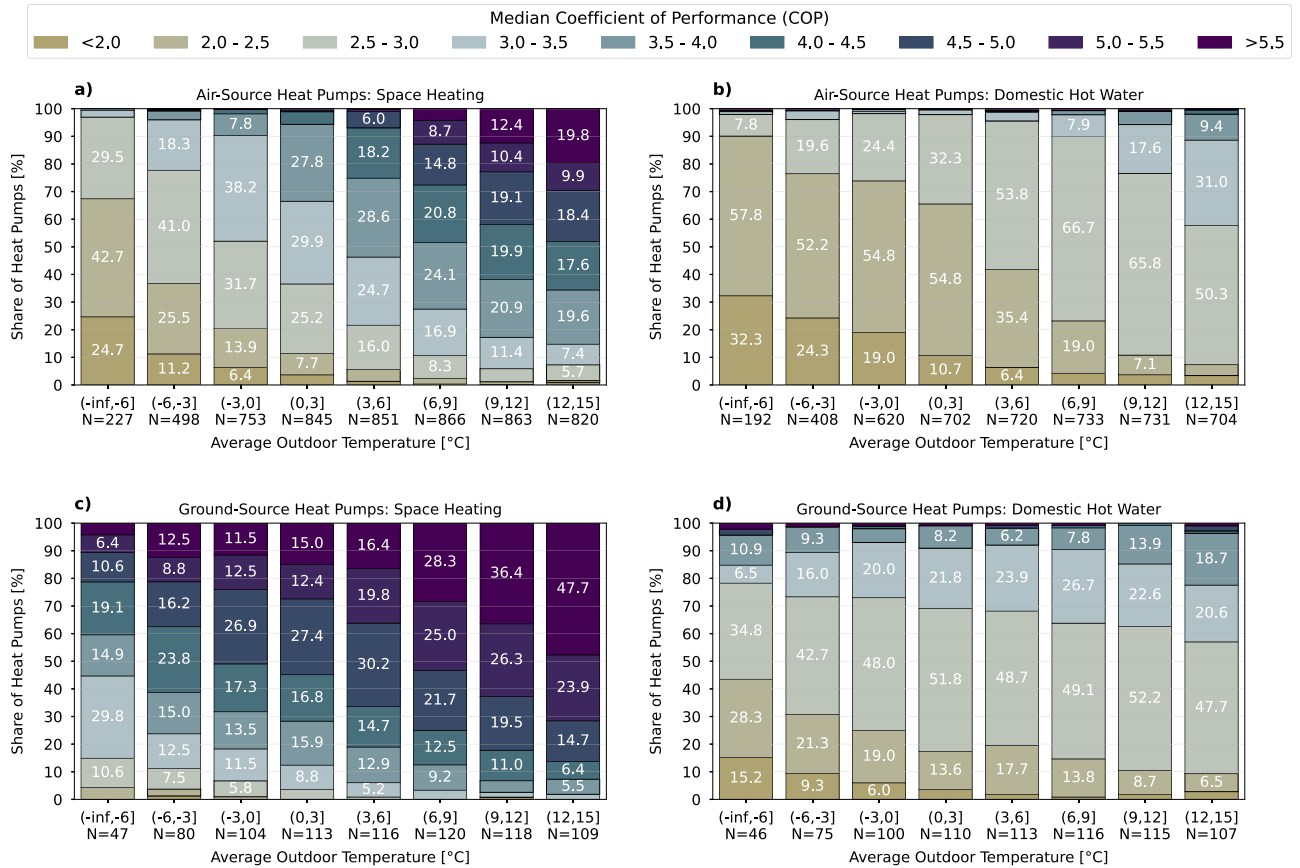
### Classifying heat pumps in terms of energy efficiency

Using  $SCOP_{real}^i$  (Equation (6)), HPs can be benchmarked against desired values or compared against each other by applying a distribution-based approach. This allows for the identification of low-performing systems. The Methods chapter provides a detailed explanation and derivation of the thresholds used to classify HP performance, with a brief summary of this procedure below. Existing regulations lack mandatory performance thresholds for HPs in real-world applications, as specified by official policies. However, the EN 14825 standard<sup>50</sup> defines minimum thresholds for the seasonal space heating energy efficiency that HPs should achieve under laboratory conditions during certification. These thresholds can be converted into SCOP limits specific to each HP type, indicating the minimum SCOP below which optimization is required. For HPs exceeding this limit, optimization remains optional but advisable. Additionally, Regulation 811/2013<sup>65</sup> offers a framework to classify HPs from A+++ to G, enhancing interpretability for HP owners to compare categorized performance labels. However, these classifications are again not mandatory for practical applications. The standard distinguishes between HPs designed for low (around 35 °C) and high (around 55 °C) temperature applications, which we average to categorize each HP because the intended application type is unknown in practice. Table 2 categorizes all HPs under assumptions of low, high, or average



**Fig. 1 | Performance and temperature dependence across heat pump types and operating modes.** **a** Distributions without accounting for temperatures. The whiskers follow Tukey's original boxplot definition, extending from  $Q_1 - 1.5 \cdot (Q_3 - Q_1)$  to  $Q_3 + 1.5 \cdot (Q_3 - Q_1)$ , where  $Q_1$  and  $Q_3$  are the first and third

quartiles. **b** Average performance considering outdoor temperature dependence. Error bars represent the 95% confidence interval. **c, d** Contour plots with linear interpolations of observed performance across outdoor and supply temperatures. Source data are provided as a Source Data file.



**Fig. 2 | Energy efficiency of individual heat pumps.** A single vertical bar represents a histogram of median coefficient of performance (COP) values calculated for individual heat pumps within specific temperature ranges and different operating modes. Distributions for all: (a) air-source heat pumps in space heating mode (b)

air-source heat pumps in domestic hot water mode (c) ground-source heat pumps in space heating mode (d) ground-source heat pumps in domestic hot water mode. Source data are provided as a Source Data file.

temperature applications, and we report evaluations with and without fixed supply temperatures. Figure 3 illustrates both the thresholds and results of the classification for average temperature applications, detailed below.

As described in the previous section, 708 systems with sufficient data and appropriate model fits are evaluated, including 612 ASHPs and 96 GSHPs. The average SCOP<sub>real</sub> of the ASHPs is 3.72, whereas for the GSHPs, it stands at 4.80. The maximum efficiency achieved by an ASHP is 5.55, whereas for GSHP, it is notably higher at 7.36. Among the ASHPs, 17.20% require optimization as they fall below a threshold of 3.01, whereas for GSHPs, only two systems (2.10%) fall below the corresponding threshold of 3.14. A significant proportion of HPs achieve high efficiency ratings, with 29.6% (8.3%) of ASHPs (GSHPs) reaching A+ level, 30.4% (17.7%) achieving A++ level, and 28.6% (72.9%) even reaching the highest A+++ level. These results underscore that HPs generally exhibit high energy efficiency even in real-world applications. However, the wide range between the lowest and highest performing systems (with a factor of two to three difference in SCOP<sub>real</sub>) highlights a significant performance gap. This underscores the importance of digital monitoring solutions, providing personalized feedback on HP efficiency, and identifying underperforming systems to optimize their operation. Furthermore, it is noteworthy that with 72.9% of GSHPs falling into the highest category A+++, there is a potential need for more refined definitions of classes within the top-performing segment, especially as devices on the market continue to achieve better performance. A further comparison of the observed

performance values with existing field studies utilizing small sample size data sets is provided in Supplementary Note 2.

### Evaluating the effects of adjustments to the heating curve

Most heating controllers allow the heating curve to be set manually, and as it is known to have a significant impact on performance<sup>17</sup>, reducing it is a simple measure to increase energy efficiency with little effort and at low cost (refer to Carnot efficiency in Equation (1)). For this reason, we investigate the effects of lowering the heating curve by shifting it parallel by 1 °C, achieved by a simple subtraction from the intercept (Equation (3)). In Fig. 4, we present the distribution of both the absolute change in SCOP and the relative change in energy consumption (Equation (8)) observed across the 708 HPs.

On average, the SCOP increases by 0.11, and the household energy consumption decreases by 2.61%. This result is consistent with a study with a smaller sample size by Lämmle et al.<sup>66</sup>, which analyzed data from 49 HPs and reported that each reduction of one Kelvin increases the seasonal performance factor by 0.10-0.13. With this improvement, 12 ASHPs (11.43% of this category) previously labeled as requiring optimization would now move to the category where optimization is optional, and the same applies to one GSHP (50% of this category). Similarly, 88 ASHPs (14.38% of all ASHPs) and 6 GSHPs (6.25% of all GSHPs) would achieve a better efficiency label. This highlights the substantial impact of HP settings on the energy efficiency achieved in practical applications.

**Table 2 | Statistics and categorization of the actual seasonal coefficient of performance (SCOP<sub>real</sub>) values**

		Without Fixed Supply Temperature			With Fixed Supply Temperatures		
		Average Temp. App.	Low Temp. App.	Medium Temp. App.	Average Temp. App.	Low Temp. App.	Medium Temp. App.
ASHPs (N = 612)	Mean	3.72	3.72	3.72	3.97	4.48	3.47
	Median	3.72	3.72	3.72	3.98	4.45	3.49
	Q <sub>1</sub>	3.17	3.17	3.17	3.75	4.05	3.31
	Q <sub>3</sub>	4.21	4.21	4.21	4.20	4.87	3.70
	Stdv	0.71	0.71	0.71	0.38	0.58	0.41
	Opt. Optional	507 (82.84%)	453 (74.02%)	543 (88.73%)	605 (98.86%)	606 (99.02%)	578 (94.44%)
	Opt. Required	105 (17.16%)	159 (25.98%)	69 (11.27%)	7 (1.14%)	6 (0.98%)	34 (5.56%)
	A+++	175 (28.59%)	106 (17.32%)	268 (43.79%)	188 (30.72%)	305 (49.84%)	69 (11.27%)
	A++	186 (30.39%)	162 (26.47%)	185 (30.23%)	366 (59.8%)	238 (38.89%)	446 (72.88%)
	A+	181 (29.58%)	197 (32.19%)	132 (21.57%)	54 (8.82%)	64 (10.46%)	80 (13.07%)
	A	30 (4.9%)	52 (8.5%)	18 (2.94%)	2 (0.33%)	2 (0.33%)	7 (1.14%)
	B	23 (3.76%)	46 (7.52%)	4 (0.65%)	1 (0.16%)	2 (0.33%)	5 (0.82%)
	C	9 (1.47%)	16 (2.61%)	4 (0.65%)	0 (0.0%)	1 (0.16%)	1 (0.16%)
D	8 (1.31%)	33 (5.39%)	1 (0.16%)	1 (0.16%)	0 (0.0%)	3 (0.49%)	
E	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
F	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
G	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.16%)	
GSHPs (N = 96)	Mean	4.80	4.80	4.80	4.80	5.52	4.29
	Median	4.84	4.84	4.84	4.92	5.44	4.38
	Q <sub>1</sub>	4.23	4.23	4.23	4.58	5.13	3.96
	Q <sub>3</sub>	5.30	5.30	5.30	5.15	5.84	4.65
	Stdv	0.87	0.87	0.87	0.57	0.74	0.54
	Opt. Optional	94 (97.92%)	93 (96.88%)	95 (98.96%)	96 (100.0%)	96 (100.0%)	96 (100.0%)
	Opt. Required	2 (2.08%)	3 (3.12%)	1 (1.04%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	A+++	70 (72.92%)	61 (63.54%)	77 (80.21%)	85 (88.54%)	87 (90.62%)	73 (76.04%)
	A++	17 (17.71%)	16 (16.67%)	17 (17.71%)	10 (10.42%)	8 (8.33%)	17 (17.71%)
	A+	8 (8.33%)	17 (17.71%)	2 (2.08%)	1 (1.04%)	1 (1.04%)	6 (6.25%)
	A	1 (1.04%)	1 (1.04%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	B	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	C	0 (0.0%)	1 (1.04%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
D	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
E	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
F	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
G	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	

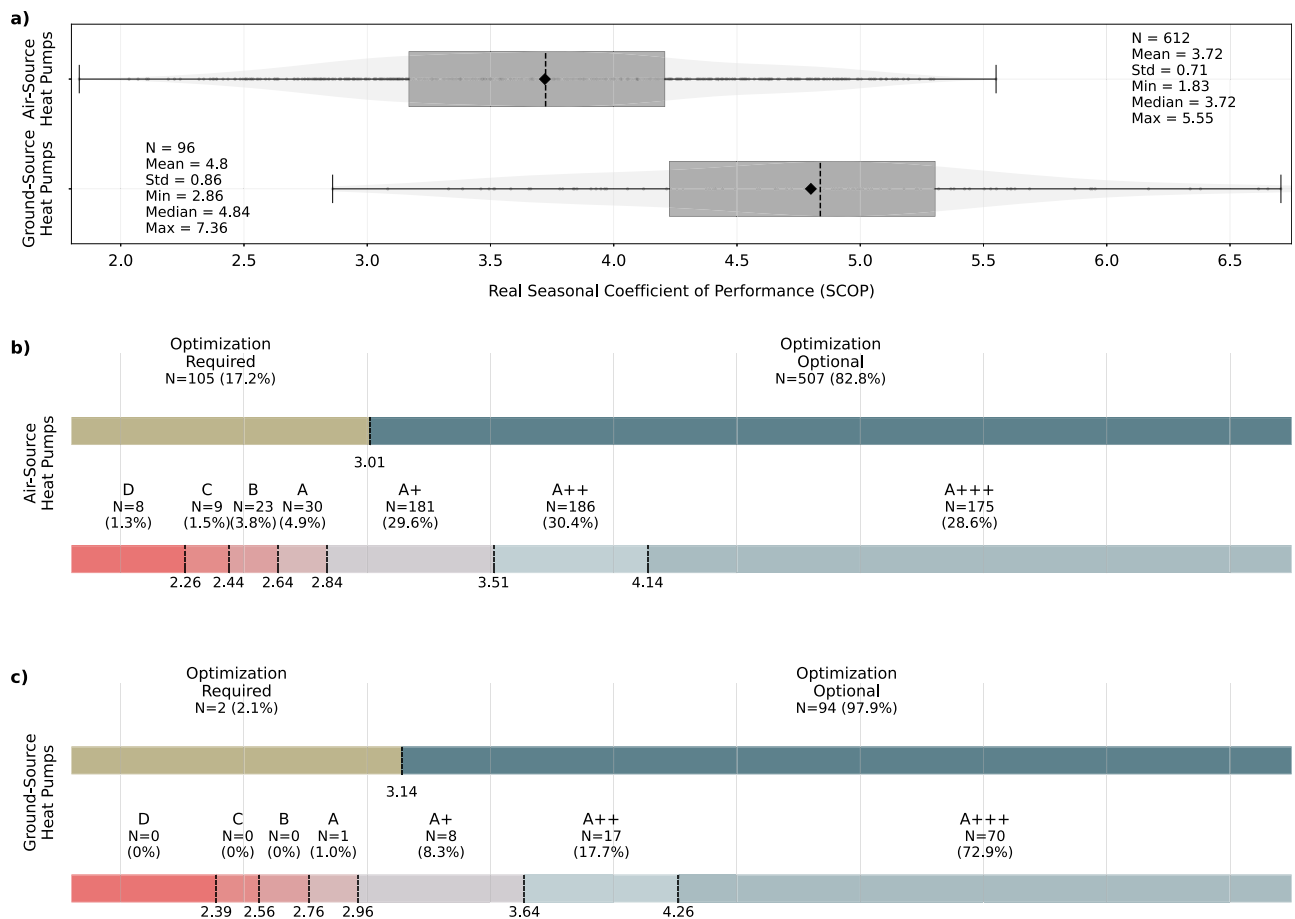
The classification evaluates efficiency classes and determines whether optimization (opt.) is required or optional. It is distinguished between temperature applications (temp. app.), and between evaluations using fixed supply temperatures defined in the standard and original supply temperatures sampled from heating curve models.

### Identifying inappropriately sized heat pumps

The size selection of an HP involves calculating the heating load required for the space it serves and matching it with a system of appropriate capacity<sup>67</sup>. Factors to consider are, for instance, building size, insulation levels, local climate conditions, chosen bivalent temperature, and manufacturer specifications. Properly sizing an HP is critical for maximizing performance throughout its operational life<sup>68</sup>. Despite its importance, installers often lack post-installation feedback on their choices, hindering opportunities for learning and improvement in future installations. Furthermore, detecting undersized systems is crucial to prevent damage. Undersized GSHPs, for example, can extract excessive energy from the ground, potentially leading to permafrost formation around the ground probe and causing it to break<sup>69,70</sup>. Early detection enables adjustments such as integrating additional heat sources to reduce the strain on the system. Accurately evaluating whether an HP is over- or undersized post-installation requires detailed contextual knowledge about building characteristics

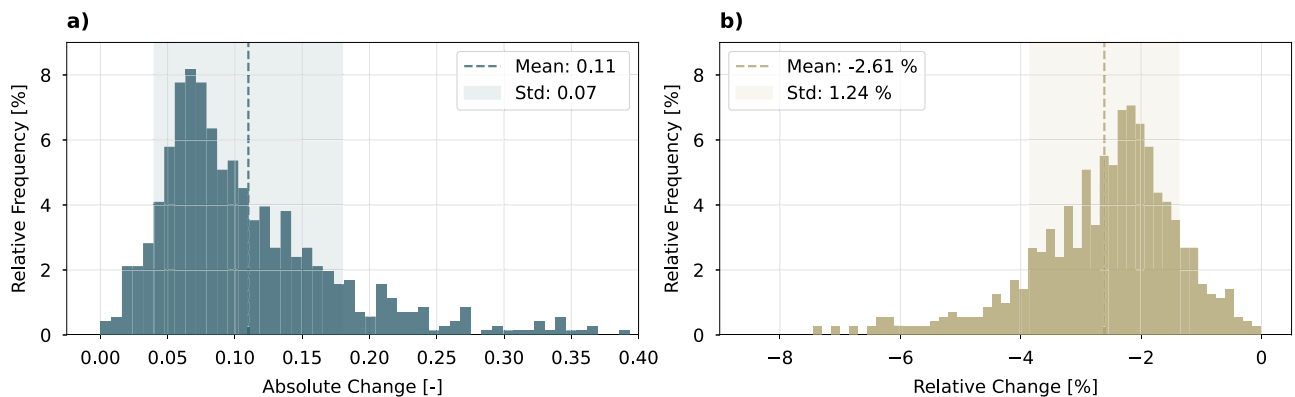
and design decisions, which is often unavailable in practice. Currently, there is no standardized method to assess inappropriate sizing using field data. However, utilization metrics are valuable indicators in this context, offering insights into HP performance under different conditions<sup>14,57</sup>. High utilization at moderate outdoor temperatures may indicate undersizing, while low utilization in cold conditions suggests potential oversizing. Utilization, expressed as a percentage, allows for standardized comparisons across HP sizes.

By sampling from the utilization model (Equation (5)), we assess the utilization of each HP at critical outdoor temperatures, such as -10 °C (the design temperature for average climate specified in EN 14825<sup>50</sup>) and 16 °C (the heating limit used in EN 14825<sup>50</sup>). These temperatures serve as conservative operational boundaries typically considered for HP performance. For instance, EN 14825<sup>50</sup> specifies -7 °C as the operational limit below which HP manufacturers do not need to guarantee their products' operation, and in well-insulated buildings, the heating limit is generally around 12 °C<sup>15</sup>. Figure 5a) presents a



**Fig. 3 | Classification of individual heat pumps by performance.** The heat pumps are categorized using seasonal coefficient of performance (SCOP) thresholds derived from the European standard EN 14825<sup>50</sup> and Regulation 811/2013<sup>65</sup>, incorporating a distinction between different heat pump types. **a** Distributions of real

SCOP values (see Equation (6)). The whiskers follow Tukey’s original boxplot definition, extending from  $Q_1 - 1.5*(Q_3 - Q_1)$  to  $Q_3 + 1.5*(Q_3 - Q_1)$ , where  $Q_1$  and  $Q_3$  are the first and third quartiles. **b** Evaluation of air-source heat pumps. **c** Evaluation of ground-source heat pumps. Source data are provided as a Source Data file.

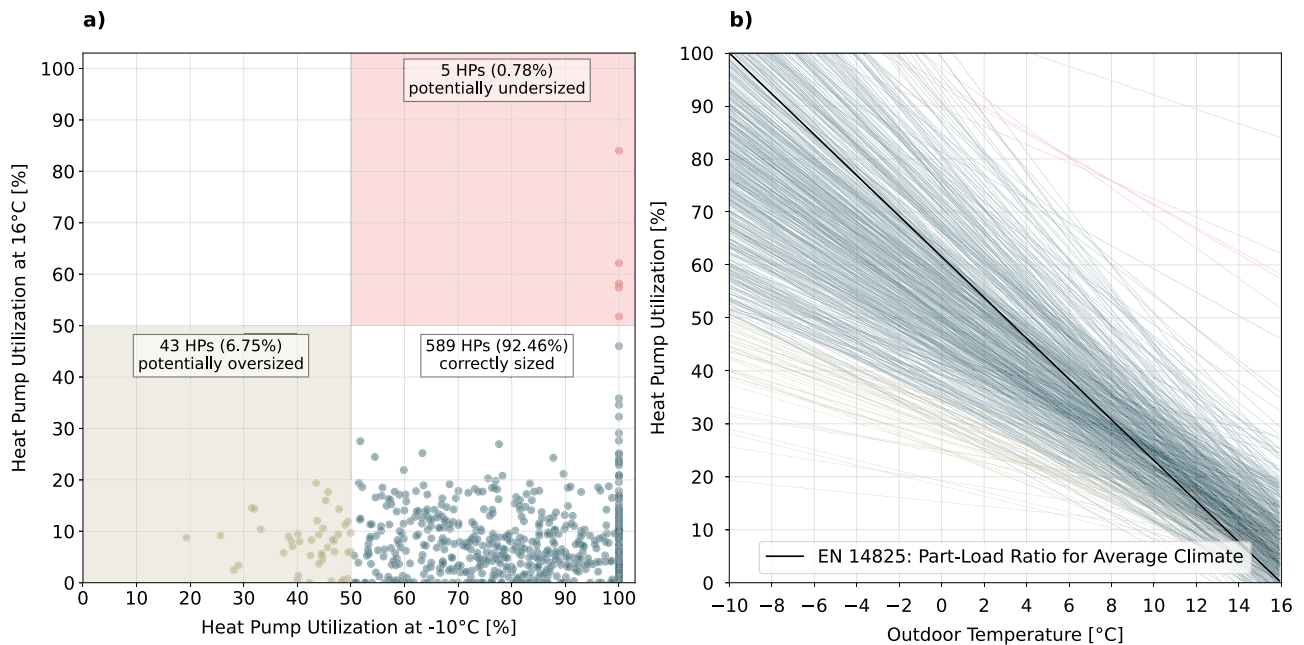


**Fig. 4 | Effect of reducing the heating curve by a 1 °C parallel shift on performance.** The graphs illustrate the distributions of the effects for each individual heat pump (N = 708). **a** Absolute change in seasonal coefficient of performance. **b** Relative change in energy consumption. Source data are provided as a Source Data file.

scatter plot illustrating the utilization of each HP sampled at -10 °C and 16 °C, while Fig. 5b) depicts the linear models of each HP alongside the PLR for average climate as defined in EN 14824<sup>50</sup>.

Rogea et al.<sup>57</sup> explored the effects of oversizing through simulation, dimensioning an HP to cover twice the original heating demand at the bivalent temperature. This suggests that an HP operating with 50% utilization at the bivalent temperature may indicate potential oversizing. Applying this criterion to the samples at -10 °C, we find that 43 HPs

(6.75%) show signs of potential oversizing. When assessed at -7 °C, this number increases to 71 HPs (11.15%). Conversely, we identify 5 HPs (0.78%) potentially undersized, as they would still operate with more than 50% utilization at 16 °C. At 12 °C, the assessment identifies 6 HPs (0.94%) potentially undersized. We summarize that inappropriate sizing of HPs may pose a more substantial issue in the field than previously reported in the literature. For example, a study by Weigert<sup>13</sup>, which analyzed 228 on-site inspection protocols, reported that only 5% of HPs



**Fig. 5 | Utilization of individual heat pumps as an indicator for appropriate sizing.** The utilization metric represents the daily average of the compressor speed relative to its full-speed capability, serving as an approximation for the heat pump's (HP) capacity ratio. **a** Evaluation of HP sizing based on utilization, sampled

at outdoor temperatures of -10 and 16 °C. **b** Fitted utilization models of each HP ( $N = 637$ ) alongside the part-load ratio for average climate defined in EN 14825 for comparison. Source data are provided as a Source Data file.

were either over- or undersized. In contrast, employing a conservative evaluation approach with our field data reveals approximately 7-11% of oversized HPs and around 1% that are undersized.

## Discussion

With an analysis of 1,023 HPs across 10 countries in Central Europe, this work represents the largest field study on HP performance. Our results and contributions can be summarized in four aspects: Firstly, we deepen the understanding of HP performance in real-world conditions through descriptive analyses and the development of models that enable future studies to simulate HPs based on actual operational data rather than relying solely on product certificates. Our findings reveal significant performance variability among individual HPs, with a 2-3 fold difference between low and high-efficiency systems. Secondly, we operationalize European regulatory thresholds into performance values that can be applied to HP field data, enabling the categorization of HPs into efficiency classes ranging from A+++ to G. Applying these thresholds to the 708 HPs examined, we identified 105 ASHPs (17.2%) and 2 GSHPs (2.1%) operating below the required energy efficiency specifications, underscoring the need for optimization. Further, with 72.9% of GSHPs falling under the highest category (A+++), our work also emphasizes the necessity for improved thresholds derived from real-world operation rather than laboratory conditions, integrated into standardized assessments defined by policymakers. Thirdly, to enhance current operations, we develop a method to evaluate the impact of reducing the heating curve by a 1 °C parallel shift, offering feedback on potential efficiency improvements. Our analysis shows an average improvement in SCOP by 0.11, corresponding to energy savings of 2.61%. With such a minor adjustment in the settings, 12 ASHPs (11.43%) and one GSHP (50%) requiring optimization could meet efficiency thresholds, while 88 ASHPs (14.38%) and 6 GSHP (6.25%) could qualify for improved efficiency labels. This emphasizes the substantial impact of configurations on HP efficiency and underscores the need for digital tools to provide feedback to users and installers. Fourthly, to guide installers in future installations, we propose a method that uses

operational utilization data to assess whether an installed HP is appropriately sized. Even with a conservative evaluation, we find that approximately 7-11% of systems may be oversized and around 1% may be undersized, indicating significant issues in planning and design practices. In the following sections, we discuss the implications of these results for various stakeholders, including policymakers, installers, users, manufacturers, and utilities.

The initial findings on variations in HP performance and the absence of suitable efficiency thresholds highlight the need for enhanced policies to accurately report HP performance, as energy efficiency labels and product certificates are key elements for user guidance<sup>71</sup>. Current certifications, derived from ideal laboratory conditions, often fail to reflect real-world HP operation<sup>15,41</sup>. This issue is similar to the inconsistencies in automotive fuel consumption labeling where lab tests do not capture real-world driving conditions<sup>72,73</sup>. Addressing the misalignment between observed and expected performance is crucial for building public confidence in HP technology and supporting the heating transition, as faster adoption requires positive word-of-mouth<sup>74,75</sup>. Comprehensive post-installation performance standards are urgently needed to bridge the gap in understanding real-world HP performance, especially in diverse building settings<sup>41,50</sup>. A proposed draft to update European regulations aims to make HP monitoring post-installation mandatory but lacks clear criteria for performance evaluation and responsibilities<sup>52</sup>. Closing these gaps is essential to ensure HPs remain economically viable, meet real-world performance expectations, and catalyze broader acceptance among stakeholders, thereby helping to achieve global installation targets<sup>4</sup>.

Additionally, the findings on the effects of reduced heating curve configurations are closely linked to the role end-users play in achieving HP efficiency in practice. To this end, more efforts are needed to improve user literacy concerning HP technology, as users with a deeper comprehension of their HPs achieve higher efficiencies<sup>22</sup>. An analysis of the experiences of 83 HP consumers showed that their level of satisfaction depends primarily on operating costs, including both

electricity consumption and maintenance costs<sup>22</sup>. When users were asked about possible improvements, 68% expressed a desire for a control system that provides feedback on cost savings and system efficiency, which underlines their expectation for guidance<sup>22</sup>. Further, troubleshooting through guided user support has been shown to lead to significantly lower maintenance costs compared to engaging energy consultants, technicians or hardware installers, who are also a limited resource<sup>13,76</sup>. This necessity for guidance is also supported by another study<sup>13</sup>, reporting that over 40% of users have limited knowledge of the heating control system and require training. The same study identifies that in 57% of the cases, the heating curve setting is set too high and could be reduced<sup>13</sup> and Narayanaswamy et al.<sup>77</sup> report that 40% of modern heating, ventilation and air conditioning systems are generally misconfigured. Addressing this significant prevalence of misconfigurations demands a fundamental shift in approach, necessitating users to possess basic knowledge of maintainability and a willingness to strike a balance between energy efficiency and heating comfort<sup>78</sup>. Instead of opting for excessively high settings to preempt heating comfort issues, installations should incorporate a testing phase. During this phase, settings should be gradually increased from the lowest point until comfort is achieved, balancing it with energy efficiency. Digital monitoring tools that offer feedback on configuration outcomes and demonstrate potential operating cost savings can greatly empower users through education<sup>78</sup>.

In addition to enhancing user literacy, it is also imperative to address HP installers, as our results show that many HPs in practice show signs of improper sizing. There is a critical need for enhanced guidance, vocational training, and feedback systems for installers and intermediaries<sup>67</sup>, as they constitute both significant drivers and barriers to the transition to energy-efficient and carbon-neutral housing<sup>79–81</sup>. Installers often serve as the primary points of contact for potential HP buyers and as advisors on HP operation. Their influence largely determines whether an HP is installed and, if so, whether it is designed appropriately and which settings are selected. However, installers' perspectives are not neutral, and they tend to opt for what is familiar to them to avoid situations where they lack the necessary skills for installation or advice-giving<sup>82</sup>. Further, related research has shown a poor correlation between installers' estimated and actual energy use of HPs<sup>83</sup>. This is largely due to the complex nature of heat demand calculations incorporating occupant preferences and other factors<sup>83</sup>. To avoid the risk of dissatisfying their customers, many installers tend to overestimate heat demand and choose oversized HPs, which can subsequently reduce operational performance<sup>68,84</sup>. Also, Decuyper et al.<sup>79</sup> report that many installers struggle to keep up with the rapid technological evolution and find it challenging and time-consuming to accurately assess energy efficiency. Digital guided support could offer installers feedback on system design and configurations, helping to optimize the operation of already installed equipment and improve learning for future installations.

To this end, HP manufacturers play a key role in offering such services that enable their appliances to be monitored and controlled<sup>33</sup>. These services should be cost-effective and privacy-preserving in order to achieve broad acceptance. Therefore, the HPs must be designed in such a way that they allow internet-based access to the sensor data. For older HPs where this option is not available in the field, data from the increasingly widespread smart electricity meters offers an alternative with great potential for standardized and manufacturer-independent performance monitoring, as addressed in<sup>14,18,56,85</sup>. This highlights the significant role utilities can play in monitoring HPs, particularly in conjunction with demand response programs and dynamic electricity tariffs. However, also the research community further needs to intensify its efforts on both sensor data and smart meter data to develop methods for HP performance evaluation and feedback. It is crucial that these methods are specifically designed to tackle practical challenges, including the absence of contextual information, handling

inaccurate measurements and data disruptions, and addressing privacy concerns.

### Limitations and future work

This study analyzes data from HPs installed in Central Europe and performance may differ in other geographic regions. Furthermore, we note that the installations are not evenly distributed across the countries included in this study (see details about the real-world data set). To ensure broader generalizability, future studies should validate our results using a data set that includes HPs from various countries and multiple manufacturers, as the current data is derived from a single manufacturer. In addition, the data comes from HPs with internet connectivity. This implies that our methods cannot be used where HPs lack sensors or do not transmit their measurements. In practice, some users may also withhold consent for data analysis due to privacy concerns, particularly regarding the HP's capability to provide real-time occupancy information. As our analyses are focused on SH, future research could extend this work to DHW and cooling applications. Another limitation of this study is that it does not analyze potential programs to exploit dynamic electricity tariffs, which, if prevalent, may influence HP operation. Considering time-of-use would enable field evaluations of the effects of HPs on electricity grids. However, this is beyond the scope of our current study and should be addressed by research focused on flexibility and demand response programs. Furthermore, additional investigations are needed to validate the quantification of inappropriately sized systems in the field, as our study is a starting point to report any figures on this issue. Similarly, we apply efficiency thresholds from European regulations to performance values observed in the field. However, these thresholds were originally intended for use under laboratory conditions. More community efforts are needed to refine these limits to better reflect real-world conditions. Future work could additionally consider integrating contextual details regarding buildings and heating systems, with an emphasis on exploring the utilization of open data for such purposes. This would allow for the use of more sophisticated models and would enable the comparison of HPs in clusters of similar buildings, while maintaining practical relevance. Additionally, real-world applications would benefit from methods for determining individual root causes of inefficient operation to increase user acceptance and provide guidelines for solving the underlying reasons of inefficiency in an automated manner. Nonetheless, this study marks a significant step toward leveraging the potential of digital monitoring solutions for improving energy efficiency of HPs in residential buildings in a scalable manner.

## Methods

### Modeling heat pump performance

For completeness, this section provides an overview of all models tested. Their definitions and parameters are detailed in Table 3. COP Model 6 is inspired by Pospíšil et al.<sup>62</sup> and Fischer et al.<sup>61</sup>, modeling COP as a quadratic function of the difference between supply and outdoor temperatures. Similarly, COP Model 1, used by Sun et al.<sup>44</sup>, employs a linear model with the same temperature difference. Heating Curve Model 1 follows the definition by Ruhnau et al.<sup>59</sup>. The models selected for further application in this study are Heating Curve Model 1 (see Equation (3)), COP Model 3 (see Equation (4)), and Utilization Model 1 (see Equation (5)). Note that in some models, the dummy variables  $d_{ASHP}^i$  and  $d_{GSHP}^i$  are used, where  $d_{ASHP}^i$  is 1 if the HP indexed  $i$  is an ASHP and 0 otherwise, and  $d_{GSHP}^i$  is 1 if it is an GSHP and 0 otherwise. These dummy variables allow for modeling even when the HP type is unknown in practical applications. All models incorporate random slopes and random intercepts for each HP, except for COP Model 5, which uses only a random intercept per HP. However, COP Model 5, along with COP Model 2 and COP Model 4, failed to converge, resulting in empty parameter estimates.

**Table 3 | List of all tested and evaluated models**

Variable	Coefficient	Std. Error	z	P >  z	[0.025, 0.975]
Heating Curve Model 1: $T_{\text{supp}}^i(T_{\text{out}}) = (a_0^i + a_0) \cdot T_{\text{out}} + (a_1^i + a_1)$					
$a_0$	-0.270	0.010	-26.239	0.000	[-0.290, -0.250]
$a_1$	38.244	0.241	158.915	0.000	[37.772, 38.715]
Heating Curve Model 2: $T_{\text{supp}}^i(T_{\text{out}}) = (a_0^i + a_0) \cdot T_{\text{out}} + (a_1^i + a_1) \cdot T_{\text{out}}^2 + (a_2^i + a_2)$					
$a_0$	-0.316	0.011	-28.040	0.000	[-0.338, -0.294]
$a_1$	0.003	0.001	5.262	0.000	[0.002, 0.005]
$a_2$	38.361	0.235	163.357	0.000	[37.901, 38.821]
COP Model 1: $\text{COP}^i(T_{\text{out}}, T_{\text{supp}}) = (b_0^i + b_0) \cdot (T_{\text{supp}} - T_{\text{out}}) + (b_1^i + b_1)$					
$b_0$	-0.099	0.002	-57.015	0.000	[-0.103, -0.096]
$b_1$	6.831	0.052	131.476	0.000	[6.729, 6.933]
COP Model 2: $\text{COP}^i(T_{\text{out}}, T_{\text{supp}}) = (b_0^i + b_0) \cdot (T_{\text{supp}} - T_{\text{out}}) + (b_1^i + b_1) + b_2 \cdot d_{\text{ASHP}}^i + b_3 \cdot d_{\text{GSHP}}^i$					
$b_0$	-0.099	0.002	-43.534	0.000	[-0.104, -0.095]
$b_1$	-19.097				
$b_2$	25.734				
$b_3$	27.056				
COP Model 3: $\text{COP}^i(T_{\text{out}}, T_{\text{supp}}) = (b_0^i + b_0) \cdot T_{\text{out}} + (b_1^i + b_1) \cdot T_{\text{supp}} + (b_2^i + b_2)$					
$b_0$	0.098	0.002	47.833	0.000	[0.094, 0.102]
$b_1$	-0.104	0.003	-37.432	0.000	[-0.109, -0.098]
$b_2$	6.965	0.099	70.354	0.000	[6.771, 7.159]
COP Model 4: $\text{COP}^i(T_{\text{out}}, T_{\text{supp}}) = (b_0^i + b_0) \cdot (T_{\text{out}}) + (b_1^i + b_1) \cdot T_{\text{supp}} + (b_2^i + b_2) + b_3 \cdot d_{\text{ASHP}}^i + b_4 \cdot d_{\text{GSHP}}^i$					
$b_0$	0.098	0.002	48.771	0.000	[0.094, 0.102]
$b_1$	-0.102	0.003	-40.543	0.000	[-0.107, -0.098]
$b_2$	-426.029				
$b_3$	432.762				
$b_4$	434.137				
COP Model 5: $\text{COP}^i(T_{\text{out}}, T_{\text{supp}}) = b_0 \cdot T_{\text{out}} + b_1 \cdot T_{\text{supp}} + (b_2^i + b_2) + b_3 \cdot d_{\text{ASHP}}^i + b_4 \cdot d_{\text{GSHP}}^i$					
$b_0$	0.099	0.000	266.505	0.000	
$b_1$	-0.088	0.001	-156.536	0.000	
$b_2$	-684.681				
$b_3$	691.141				
$b_4$	692.234				
COP Model 6: $\text{COP}^i(T_{\text{out}}, T_{\text{supp}}) = (b_0^i + b_0) \cdot (T_{\text{supp}} - T_{\text{out}}) + (b_1^i + b_1) \cdot (T_{\text{supp}} - T_{\text{out}})^2 + (b_2^i + b_2)$					
$b_0$	-0.123	0.016	-7.523	0.000	[-0.155, -0.091]
$b_1$	0.000	0.017	0.017	0.986	[-0.032, 0.033]
$b_2$	7.281	0.070	103.637	0.000	[7.143, 7.419]
Utilization Model 1: $\text{Utilization}^i(T_{\text{out}}) = (c_0^i + c_0) \cdot T_{\text{out}} + (c_1^i + c_1)$					
$c_0$	-2.739	0.041	-66.214	0.000	[-2.820, -2.658]
$c_1$	50.865	0.620	82.025	0.000	[49.650, 52.081]

The models selected for further use in this study are Heating Curve Model 1 (see Equation (3)), coefficient of performance (COP) Model 3 (see Equation (4)), and Utilization Model 1 (see Equation (5)).

### Deriving a classification scheme for heat pump energy efficiency

To evaluate HPs based on their energy efficiency, we calculate the  $\text{SCOP}_{\text{real}}^i$  (Equation (6)) of each HP according to the definition in the European standard EN 14825<sup>50</sup>, which describes performance as a single metric. Note that the definition of SCOP only considers SH, which means that this value is not calculated for DHW. Below, we provide a detailed explanation of how the categorization scheme is derived from European regulations. This scheme distinguishes between HPs where optimization is required or optional, and further categorizes them into distinct efficiency classes. However, it is important to clarify that these regulations form part of the certification and labeling process for HP products under laboratory conditions. Hence, they should not be interpreted as mandatory performance limits that HPs must achieve in practical usage, as such limits currently do not exist.

The standard EN 14825<sup>50</sup> does not specify direct thresholds for SCOP. However, it specifies minimum desired values for the seasonal space heating energy efficiency (SSHEE)  $\eta$ , expressed as a percentage. According to the definition, SSHEE can be calculated from SCOP as follows:

$$\eta = 0.4 \cdot \text{SCOP} \cdot 100\% - F(1) - F(2) \quad (9)$$

The value of 0.4 represents the average European grid power generation efficiency factor. Additionally,  $F(1) = 3\%$  serves as a correction factor accounting for contributions from temperature controls, while  $F(2) = 5\%$  acts as a correction factor specific to water-to-air or water-to-water systems<sup>41</sup>. As noted in<sup>60</sup>, for GSHPs, the combined correction factors  $F(1) + F(2) = 8\%$  apply, whereas for ASHPs, only  $F(1) = 3\%$  should be used. Consequently, the SCOP can be calculated

based on a given SSHEE  $\eta$  as follows:

$$\text{SCOP} = \begin{cases} \frac{\eta + 8\%}{0.4 \cdot 100\%} & \text{for GSHPs} \\ \frac{\eta + 3\%}{0.4 \cdot 100\%} & \text{for ASHPs} \end{cases} \quad (10)$$

According to EN 14825<sup>50</sup>,  $\eta$  shall not be lower than 110% for typical HP space heaters and HP combination heaters, unless they are low-temperature HPs, for which  $\eta$  should not be below 125%. A low-temperature HP is defined as “a heat pump space heater that is specifically designed for low-temperature application, and that cannot deliver heating water with an outlet temperature of 52 °C at an inlet dry (wet) bulb temperature of -7 °C (-8 °C) in the reference design conditions for average climate”<sup>50</sup>. Thus, the standard distinguishes between HPs designed for low-temperature applications (supply temperatures around 35°C) and medium-temperature applications (supply temperatures around 55°C)<sup>50</sup>. Using these thresholds as inputs into Equation (10), SCOP values can be derived, below which an HP requires optimization. As a result, GSHPs operating below an SCOP value of 2.95 (medium-temperature) or 3.325 (low-temperature) require optimization. Correspondingly, ASHPs should be optimized if their SCOP falls below 2.825 (medium-temperature) or 3.2 (low-temperature). In a real-world scenario, however, it is often unknown whether an HP was specifically designed for low-temperature applications. Therefore, the choice of the threshold also depends on how rigorous the benchmarking scheme should be. As a compromise, we utilize the average of these SCOP thresholds for each HP type to categorize whether optimization of HPs is necessary or optional. Thus, to evaluate whether an HP requires optimization, we apply a threshold of 3.14 for GSHPs and 3.01 for ASHPs.

Furthermore, the standard EN 14825<sup>50</sup> is complemented by Regulation 811/2013<sup>65</sup>, which establishes additional thresholds for SSHEE in the energy labeling of HP space heaters, categorized as A+++, A++, A+, and A to G. Following the same procedure as before, we calculate lower and upper boundaries for each category and HP type. For evaluation, we again utilize the corresponding averages of values from low and medium temperature applications. An overview of all thresholds is presented in Supplementary Table 2, where closed brackets denote that the value is included in the interval, while open brackets indicate that the value is excluded.

## Data availability

The raw data are protected and are not available due to data privacy laws. The processed data and the data generated in this study are provided in the Source Data file. Source data are provided with this paper.

## References

- International Energy Agency (IEA). Heat pumps. <https://www.iea.org/energy-system/buildings/heat-pumps> (International Energy Agency (IEA), Paris, 2023).
- Ruhnau, O., Hirth, L. & Praktiknjo, A. Heating with wind: economics of heat pumps and variable renewables. *Energy Econ.* **92**, 104967 (2020).
- International Energy Agency (IEA). Heating. <https://www.iea.org/energy-system/buildings/heating> (International Energy Agency (IEA), Paris, 2023).
- Rosenow, J., Gibb, D., Nowak, T. & Lowes, R. Heating up the global heat pump market. *Nat. Energy* **7**, 901–904 (2022).
- Bauermann, K. German Energiewende and the heating market—impact and limits of policy. *Energy Policy* **94**, 235–246 (2016).
- Pensini, A., Rasmussen, C. N. & Kempton, W. Economic analysis of using excess renewable electricity to displace heating fuels. *Appl. Energy* **131**, 530–543 (2014).
- Poblete-Cazenave, M. & Rao, N. D. Social and contextual determinants of heat pump adoption in the US: implications for subsidy policy design. *Energy Res. Soc. Sci.* **104**, 103255 (2023).
- Kokoni, S. & Leach, M. Policy mechanisms to support heat pump deployment: a UK case study based on techno-economic modelling. *Renew. Sustain. Energy Transit.* **1**, 100009 (2021).
- Tanil, G. et al. *Political and Public Perceptions*, 85–111 (Springer International Publishing, Cham, 2023).
- Engelen, K. C. Heat pump fiasco. *Int. Econ.* **37**, 10–13 (2023).
- Nast, M., Langniß, O. & Leprich, U. Instruments to promote renewable energy in the German heat market—renewable heat sources act. *Renew. Energy* **32**, 1127–1135 (2007).
- Decker, T. & Menrad, K. House owners’ perceptions and factors influencing their choice of specific heating systems in Germany. *Energy Policy* **85**, 150–161 (2015).
- Weigert, A. Identification and classification of heat pump problems in the field and their implication for a user-centric problem recognition. *Energy Inform.* **5**, 70 (2022).
- Brudermueller, T., Kreft, M., Fleisch, E. & Staake, T. Large-scale monitoring of residential heat pump cycling using smart meter data. *Appl. Energy* **350**, 121734 (2023).
- Nolting, L., Steiger, S. & Praktiknjo, A. Assessing the validity of European labels for energy efficiency of heat pumps. *J. Build. Eng.* **18**, 476–486 (2018).
- Faye Wade, M. S. & Hitchings, R. How installers select and explain domestic heating controls. *Build. Res. Inf.* **45**, 371–383 (2017).
- Tejeda, A., Milu, A., Riviere, P. & Marchio, D. Energy consequences of non-optimal heat pump parameterization. In *International High Performance Buildings Conference*. <https://docs.lib.purdue.edu/ihpbc/110/> (2014).
- Weigert, A., Hopf, K., Günther, S. A. & Staake, T. Heat pump inspections result in large energy savings when a pre-selection of households is performed: a promising use case of smart meter data. *Energy Policy* **169**, 113156 (2022).
- Ryland, M. & He, W. Heating economics evaluated against emissions: an analysis of low-carbon heating systems with spatio-temporal and dwelling variations. *Energy Build.* **277**, 112561 (2022).
- Shah, A., Krarti, M. & Huang, J. Energy performance evaluation of shallow ground source heat pumps for residential buildings. *Energies* **15**, 1025 (2022).
- Sadeghi, H., Ijaz, A. & Singh, R. M. Current status of heat pumps in Norway and analysis of their performance and payback time. *Sustain. Energy Technol. Assess.* **54**, 102829 (2022).
- Caird, S., Roy, R. & Potter, S. Domestic heat pumps in the UK: user behaviour, satisfaction and performance. *Energy Efficiency* **5**, 283–301 (2012).
- Müller, F. & Jansen, B. Large-scale demonstration of precise demand response provided by residential heat pumps. *Appl. Energy* **239**, 836–845 (2019).
- Lee, Z. E. et al. Providing grid services with heat pumps: a review. *ASME J. Eng. Sustain. Build. Cities* **1**, 011007 (2020).
- Lizana, J. et al. A national data-based energy modelling to identify optimal heat storage capacity to support heating electrification. *Energy* **262**, 125298 (2023).
- Love, J. et al. The addition of heat pump electricity load profiles to GB electricity demand: evidence from a heat pump field trial. *Appl. Energy* **204**, 332–342 (2017).
- Halloran, C., Lizana, J., Fele, F. & McCulloch, M. Data-based, high spatiotemporal resolution heat pump demand for power system planning. *Appl. Energy* **355**, 122331 (2024).
- Watson, S., Crawley, J., Lomas, K. & Buswell, R. Predicting future GB heat pump electricity demand. *Energy Build.* **286**, 112917 (2023).
- Rinaldi, A., Yilmaz, S., Patel, M. K. & Parra, D. What adds more flexibility? an energy system analysis of storage, demand-side

- response, heating electrification, and distribution reinforcement. *Renew. Sustain. Energy Rev.* **167**, 112696 (2022).
30. Protopapadaki, C. & Saelens, D. Heat pump and PV impact on residential low-voltage distribution grids as a function of building and district properties. *Appl. Energy* **192**, 268–281 (2017).
31. Terreros, O. et al. Electricity market options for heat pumps in rural district heating networks in Austria. *Energy* **196**, 116875 (2020).
32. Bernath, C., Deac, G. & Sensfuß, F. Influence of heat pumps on renewable electricity integration: Germany in a European context. *Energy Strategy Rev.* **26**, 100389 (2019).
33. International Energy Agency - Technology Collaboration Programme on Heat Pumping Technology (IEA - HPT TCP). Annex 56 on digitalization and IoT for heat pumps. <https://heatpumpingtechnologies.org/annex56/> (2023).
34. Larsen, S. P. A. K. & Gram-Hanssen, K. When space heating becomes digitalized: investigating competencies for controlling smart home technology in the energy-efficient home. *Sustainability* **12**, 6031 (2020).
35. Song, Y., Rolando, D., Marchante Avellaneda, J., Zucker, G. & Madani, H. Data-driven soft sensors targeting heat pump systems. *Energy Convers. Manag.* **279**, 116769 (2023).
36. Bellanco, I., Fuentes, E., Vallès, M. & Salom, J. A review of the fault behavior of heat pumps and measurements, detection and diagnosis methods including virtual sensors. *J. Build. Eng.* **39**, 102254 (2021).
37. Carroll, P., Chesser, M. & Lyons, P. Air source heat pumps field studies: a systematic literature review. *Renew. Sustain. Energy Rev.* **134**, 110275 (2020).
38. Huchtemann, K. & Müller, D. Evaluation of a field test with retrofit heat pumps. *Build. Environ.* **53**, 100–106 (2012).
39. Lowe, R. et al. Final report on analysis of heat pump data from the Renewable Heat Premium Payment (RHPP) scheme. [https://assets.publishing.service.gov.uk/media/5a82b8faed915d74e62374d8/DECC\\_RHPP\\_161214\\_Final\\_Report\\_v1-13.pdf](https://assets.publishing.service.gov.uk/media/5a82b8faed915d74e62374d8/DECC_RHPP_161214_Final_Report_v1-13.pdf) (2017).
40. Miara, M., Günther, D., Langner, R., Helmling, S. & Wapler, J. 10 years of heat pumps monitoring in Germany. outcomes of several monitoring campaigns. from low-energy houses to un-retrofitted single-family dwellings. In *12th IEA Heat Pump Conference*, 11 (2017).
41. O'Hegarty, R., Kinnane, O., Lennon, D. & Colclough, S. Air-to-water heat pumps: review and analysis of the performance gap between in-use and product rated performance. *Renew. Sustain. Energy Rev.* **155**, 111887 (2022).
42. Lämmle, M. et al. Heat pump systems in existing multifamily buildings: a meta-analysis of field measurement data focusing on the relationship of temperature and performance of heat pump systems. *Energy Technol.* **11**, 2300379 (2023).
43. Deng, J., Wei, Q., Liang, M., He, S. & Zhang, H. Does heat pumps perform energy efficiently as we expected: field tests and evaluations on various kinds of heat pump systems for space heating. *Energy Build.* **182**, 172–186 (2019).
44. Sun, X. et al. Seasonal heating performance prediction of air-to-water heat pumps based on short-term dynamic monitoring. *Renew. Energy* **180**, 829–837 (2021).
45. Gao, B. et al. Operation performance test and energy efficiency analysis of ground-source heat pump systems. *J. Build. Eng.* **41**, 102446 (2021).
46. Deng, J. et al. How to improve the energy performance of mid-deep geothermal heat pump systems: optimization of heat pump, system configuration and control strategy. *Energy* **285**, 129537 (2023).
47. Zhou, X., Lin, W., Cui, P., Ma, Z. & Huang, T. An unsupervised data mining strategy for performance evaluation of ground source heat pump systems. *Sustain. Energy Technol. Assess.* **46**, 101255 (2021).
48. Noye, S., Mulero Martinez, R., Carnieletto, L., De Carli, M. & Castelruiz Aguirre, A. A review of advanced ground source heat pump control: Artificial intelligence for autonomous and adaptive control. *Renew. Sustain. Energy Rev.* **153**, 111685 (2022).
49. Nolting, L. & Praktijnjo, A. Techno-economic analysis of flexible heat pump controls. *Appl. Energy* **238**, 1417–1433 (2019).
50. European Standard EN 14825. 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; German version EN 14825:2022 (2022).
51. Nordman, R. et al. SEasonal PErformance factor and MOonitoring for heat pump systems in the building sector SEPOMO-Build: Final report. (Intelligent Energy Europe 2012).
52. Norwegian Ministry of Energy. The Norwegian Water Resources and Energy Directorate (NVE). [XXX/XXXX] Ecodesign regulation space / combination heaters COMMISSION REGULATION (EU) No [XXX/XXXX] of [XX/XX/XXXX] implementing Directive 2009/125/EC of the European Parliament and of the Council with regard to ecodesign requirements for space heaters and combination heaters, repealing Commission Regulation (EU) No 813/2013 and Council Directive 92/42/EEC. [https://www.nve.no/media/15330/space-heaters\\_ed\\_27032023.pdf](https://www.nve.no/media/15330/space-heaters_ed_27032023.pdf) (2023).
53. Staffell, I., Brett, D., Brandon, N. & Hawkes, A. A review of domestic heat pumps. *Energy Environ. Sci.* **5**, 9291–9306 (2012).
54. Verley, G., Esposito, M., Willaert, T. & Van den Broeck, C. The unlikely carnot efficiency. *Nat. Commun.* **5**, 4721 (2014).
55. O'Donovan, A. & O'Sullivan, P. In-use performance of air-to-water heat pumps: are the standards robust? In *E3S Web of Conferences*, **246**, 06002 (EDP Sciences, 2021).
56. Brudermueller, T., Wirth, F., Weigert, A. & Staake, T. Automatic differentiation of variable and fixed speed heat pumps with smart meter data. In *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 412–418 (2022).
57. Rogeau, A., Vieubled, J., Ruche, L. & Girard, R. A generic methodology for mapping the performance of various heat pumps configurations considering part-load behavior. *Energy Build.* **318**, 114490 (2024).
58. Sieres, J., Ortega, I., Cerdeira, F., Álvarez, E. & Santos, J. M. Seasonal efficiency of a brine-to-water heat pump with different control options according to ecodesign standards. *Clean. Technol.* **4**, 542–554 (2022).
59. Ruhnau, O., Hirth, L. & Praktijnjo, A. Time series of heat demand and heat pump efficiency for energy system modeling. *Sci. Data* **6**, 1–10 (2019).
60. Palkowski, C., Zottl, A., Malenkovic, I. & Simo, A. Fixing efficiency values by unfixing compressor speed: dynamic test method for heat pumps. *Energies* **12**, 1045 (2019).
61. Fischer, D., Wolf, T., Wapler, J., Hollinger, R. & Madani, H. Model-based flexibility assessment of a residential heat pump pool. *Energy* **118**, 853–864 (2017).
62. Pospíšil, J., Špiláček, M. & Kudela, L. Potential of predictive control for improvement of seasonal coefficient of performance of air source heat pump in Central European climate zone. *Energy* **154**, 415–423 (2018).
63. Mouzeviris, G. A. & Papakostas, K. T. Comparative analysis of air-to-water and ground source heat pumps performances. *Int. J. Sustain. Energy* **40**, 69–84 (2021).
64. Violante, A. C., Donato, F., Guidi, G. & Proposito, M. Comparative life cycle assessment of the ground source heat pump vs air source heat pump. *Renew. Energy* **188**, 1029–1037 (2022).
65. The European Commission. *Commission Delegated Regulation (Eu) No 811/2013 Of 18 February 2013 Supplementing Directive 2010/30/ eu Of The European Parliament And Of The Council With Regard To The Energy Labelling Of Space Heaters, Combination Heaters, Packages Of Space Heater, Temperature Control And Solar Device And Packages Of Combination Heater, Temperature Control And Solar Device.* <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013R0811> (2013).

66. Lämmle, M. et al. Performance of air and ground source heat pumps retrofitted to radiator heating systems and measures to reduce space heating temperatures in existing buildings. *Energy* **242**, 122952 (2022).
67. Gleeson, C. P. Residential heat pump installations: the role of vocational education and training. *Build. Res. Inf.* **44**, 394–406 (2016).
68. Bagarella, G., Lazzarin, R. & Noro, M. Sizing strategy of on-off and modulating heat pump systems based on annual energy analysis. *Int. J. Refrig.* **65**, 183–193 (2016).
69. Hein, P., Kolditz, O., Görke, U.-J., Bucher, A. & Shao, H. A numerical study on the sustainability and efficiency of borehole heat exchanger coupled ground source heat pump systems. *Appl. Therm. Eng.* **100**, 421–433 (2016).
70. Klingebiel, J., Hassan, M., Venzik, V., Vering, C. & Müller, D. Efficiency comparison between defrosting methods: a laboratory study on reverse-cycle defrosting, electric heating defrosting, and warm brine defrosting. *Appl. Therm. Eng.* **233**, 121072 (2023).
71. Stadelmann, M. & Schubert, R. How do different designs of energy labels influence purchases of household appliances? A field study in Switzerland. *Ecol. Econ.* **144**, 112–123 (2018).
72. Ntziachristos, L. et al. In-use vs. type-approval fuel consumption of current passenger cars in Europe. *Energy Policy* **67**, 403–411 (2014).
73. Yu, R., Ren, H., Liu, Y. & Yu, B. Gap between on-road and official fuel efficiency of passenger vehicles in China. *Energy Policy* **152**, 112236 (2021).
74. Michelsen, C. C. & Madlener, R. Homeowner satisfaction with low-carbon heating technologies. *J. Clean. Prod.* **141**, 1286–1292 (2017).
75. Flower, J., Hawker, G. & Bell, K. Heterogeneity of UK residential heat demand and its impact on the value case for heat pumps. *Energy Policy* **144**, 111593 (2020).
76. Oikonomou, E., Zimmermann, N., Davies, M. & Oreszczyn, T. Behavioural change as a domestic heat pump performance driver: Insights on the influence of feedback systems from multiple case studies in the UK. *Sustainability* **14**, 16799 (2022).
77. Narayanaswamy, B., Balaji, B., Gupta, R. & Agarwal, Y. Data driven investigation of faults in HVAC systems with model, cluster and compare (MCC). In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, BuildSys '14, 50–59 (Association for Computing Machinery, New York, NY, USA, 2014).
78. de Wilde, M. A heat pump needs a bit of care: on maintainability and repairing gender-technology relations. *Sci., Technol., Hum. Values* **46**, 1261–1285 (2021).
79. Decuyper, R., Robaey, B., Hudders, L., Baccarne, B. & Van de Sompel, D. Transitioning to energy efficient housing: drivers and barriers of intermediaries in heat pump technology. *Energy Policy* **161**, 112709 (2022).
80. Bergman, N. Why is renewable heat in the UK underperforming? a socio-technical perspective. *Proc. Inst. Mech. Eng., Part A: J. Power Energy* **227**, 124–131 (2013).
81. European Heat Pump Association (EHPA). European heat pump market and statistics report 2021. <https://www.ehpa.org/market-data/market-report-2021/> (2023).
82. Scheuer, C. W. et al. *Adoption of Residential Green Building Practices: Understanding the Role of Familiarity*. Ph.D. thesis, University of Michigan (2007).
83. Gleeson, C. et al. *Analysis Of Heat Pump Data From The Renewable Heat Premium Payment Scheme (Rhpp) To The Department Of Business, Energy And Industrial Strategy: Compliance With Mcs Installation Standards*. (2017).
84. Dongellini, M., Naldi, C. & Morini, G. L. Sizing effects on the energy performance of reversible air-source heat pumps for office buildings. *Appl. Therm. Eng.* **114**, 1073–1081 (2017).
85. Brudermueller, T., Breer, F. & Staake, T. Disaggregation of heat pump load profiles from low-resolution smart meter data. In *Proceedings of*

*the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '23, 228–231 (Association for Computing Machinery, New York, NY, USA, 2023).

## Acknowledgements

This research was funded and supported by the Swiss Federal Office of Energy under the grant number SI/502257 (T.B., T.S., E.F.) and the Bosch IoT Lab at the University of St. Gallen and ETH Zurich (U.P., F.W.).

## Author contributions

T.B.: conceptualization; data curation; formal analysis; investigation; methodology; software; visualization; writing - original draft; writing - review & editing. U.P.: investigation; methodology; software; writing - review & editing. E.F., F.W., & T.S.: writing - review & editing; supervision; project administration.

## Funding

Open access funding provided by Swiss Federal Institute of Technology Zurich.

## Competing interests

T.S. & E.F. declare that they are supervisory board members of companies (Hoval and Bosch, respectively) that, among other products and services, manufacture and sell heating systems. T.B., U.P., & F.W. declare no competing interests. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policy of the sponsors or partners, either expressed or implied. The funding agencies and partners had no control over the design, conduct, data, analysis, review, reporting, or interpretation of the research conducted.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58014-y>.

**Correspondence** and requests for materials should be addressed to Tobias Brudermueller or Felix Wortmann.

**Peer review information** *Nature Communications* thanks Paula Carroll, Lars Nolting and Yifeng Hu for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025