Secondary Publication

Brudermüller, Tobias; Kreft, Markus; Fleisch, Elgar; Staake, Thorsten

Large-scale monitoring of residential heat pump cycling using smart meter data

Date of secondary publication: 19.09.2023

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-909184

Primary publication

Brudermüller, Tobias; Kreft, Markus; Fleisch, Elgar; Staake, Thorsten (2023): „Large-scale monitoring of residential heat pump cycling using smart meter data“. In: Applied energy, Jg. 350, 15 November 2023, Amsterdam [u.a.]: Elsevier, 121734, doi: 10.1016/j.apenergy.2023.121734.

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-holder(s).

This document is made available under a Creative Commons license.

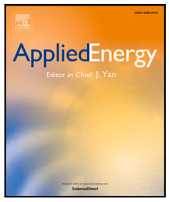
The license information is available online:
https://creativecommons.org/licenses/by/4.0/legalcode

Applied Energy 350 (2023) 121734

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



Large-scale monitoring of residential heat pump cycling using smart meter data

Tobias Brudermueller [a,*], Markus Kreft [a], Elgar Fleisch [a,b], Thorsten Staake [a,c]

[a] ETH Zurich, Chair of Information Management, Weinbergstrasse 56/58, Zurich, 8092, Switzerland
[b] University of St. Gallen, Institute of Technology Management, Dufourstrasse 40a, St. Gallen, 9000, Switzerland
[c] University of Bamberg, Chair of Information Systems and Energy Efficient Systems, An der Weberei 5, Bamberg, 96047, Germany

ARTICLE INFO

Keywords:
Smart meter data
Heat pump
Cycling behavior
outlier detection
Residential buildings
Energy efficiency
Appliance monitoring
Real-world operation
Machine learning

ABSTRACT

Heat pumps play an essential role in decarbonizing the building sector, but their electricity consumption can vary significantly across buildings. This variability is closely related to their cycling behavior (i.e., the frequency of on–off transitions), which is also an indicator for improper sizing and non-optimal settings and can affect a heat pump's lifetime. Up to now it has been unclear which cycling behaviors are typical and atypical for heat pump operation in the field and importantly, there is a lack of methods to identify heat pumps that cycle atypically. Therefore, in this study we develop a method to monitor heat pumps with energy measurements delivered by common smart electricity meters, which also cover heat pumps without network connectivity. We show how smart meter data with 15-minute resolution can be used to extract key indicators about heat pump cycling and outline how atypical behavior can be detected after controlling for outdoor temperature. Our method is robust across different building characteristics and varying times of observation, does not require contextual information, and can be implemented with existing smart meter data, making it suitable for real-world applications. Analyzing 503 heat pumps in Swiss households over a period of 21 months, we further describe behavioral differences with respect to building and heat pump characteristics and study the relationship between heat pumps' cycling behavior, energy efficiency, and appropriate sizing. Our results show that outliers in cycling behavior are more than twice as common for air-source heat pumps than for ground-source heat pumps.

1. Introduction

Alongside increasing energy prices and ambitious targets to reduce greenhouse gas emissions, considerable attention is being paid to the energy needs of the building sector. The International Energy Agency (IEA) states that almost half of buildings' energy demand is used for space and water heating. The associated CO_2 emissions reached an all-time high of 2.5 Gt in 2021 [1]. Consequently, a large-scale deployment of clean and efficient heating technologies is needed to reach emission targets. Electric HPs are a central pillar in this context because they can reduce greenhouse gas emissions when replacing natural gas or oil furnaces [2,3]. The amount of savings depends on the primary energy for the electricity that is consumed by the HP [4]. Accordingly, this favors countries with a high share of electricity generation from renewables, nuclear power and/or hydropower. The authors of [5] show that per TJ of heating energy, HPs in Belgium could save up to 47 t of CO_2, whereas in Switzerland it could even be up to 61 t. Consequently, HPs have become a backbone of many policymakers' carbon mitigation plans. As a result, the global stock of HPs has achieved an average annual growth rate of 10% over the past five years, sometimes with the help of subsidies. More than 190 million HPs are in operation worldwide [1], and the IEA estimates that 600 million HP installations are required by 2030 to cover 20% of buildings' heating needs [6].

In theory, HPs are highly efficient devices. In practice however, many HPs have a significantly higher electricity consumption and lower efficiency than stated by their manufacturer [3,7–10]. Hence, there is a large potential to optimize HPs in the field. The authors of [10] showed that half of the 297 Swiss households studied achieved average savings of 1,805 kWh (15.2%) per year after their HPs were optimized by an energy consultant. The large gap in HP performance has multiple reasons. First, the seasonal performance depends on weather and climate conditions [11–13]. Second, non-optimal planning in the form of over- or undersized HPs leads to inefficient operation [12,14,15]. Third, heat

* Corresponding author.
E-mail addresses: tbrudermuell@ethz.ch (T. Brudermueller), mkreft@ethz.ch (M. Kreft), efleisch@ethz.ch (E. Fleisch), thorsten.staake@uni-bamberg.de (T. Staake).

https://doi.org/10.1016/j.apenergy.2023.121734
Received 28 April 2023; Received in revised form 16 June 2023; Accepted 6 August 2023
Available online 18 August 2023
0306-2619/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

pumps can show faulty or unwanted behavior [16,17]. Lastly and most commonly, HPs can operate without faults but suffer from misconfigurations. A study estimates that 40% of modern heating, ventilation and air conditioning (HVAC) systems are misconfigured [18]. Such misconfigurations include unsuitable heating curve settings, unfavorable cut-off temperatures and wrong bivalent temperatures, which are all closely linked to HP performance [19,20].

The consequences of inefficient HP operation are severe. As the energy consumption of an HP accounts for 90% of its carbon dioxide equivalent emissions [21], the emissions savings associated with HPs highly depend on their electricity consumption. In this context, a study [21] finds a 4% decrease of life cycle climate performance when the coefficient of performance of an HP is improved by 5%. Moreover, HPs are typically associated with high upfront costs, so operational costs play a critical role in making HPs financially attractive compared to fossil-based heating solutions such as gas boilers [22]. High operational costs can create mistrust in the technology and can slow down the endeavors to decarbonize the heating sector. Lastly, the inefficiencies have implications for the grid. Already now, utilities need solutions to respond to the increased electricity demand from HPs [23,24]. A study which simulates an HP penetration of 100% in Great Britain [25] found that this scenario could cause the national annual electricity demand to rise to 189 TWh, which corresponds to an increase of around 60%. Another UK-based study found that the peak grid demand increases by 14% if 20% of households use HPs [26] - a phenomenon that is amplified when many HPs operate outside of their optimal working conditions.

As a result, many manufacturers develop digital solutions for remotely monitoring the systems of their own customer base. However, these solutions still face several problems. First, existing products focus more on breakdown prevention and less on efficiency improvements [27]. Second, there are no standards for cross-manufacturer data exchange [28], nor is there a unified service, limiting the ability of owners to monitor their HPs if their manufacturer does not offer a service. Moreover, most of the currently installed HPs do not have a network connection [29]. Consequently, these HPs are not covered by digital service offerings even though they will remain in operation for years or decades.

However, a parallel development in recent years can serve as an important transitional technology. The deployment of smart electricity meters (SMs) could form the basis for more widespread efficiency services. A study reports that in 2019, 94.8 million SMs were installed in the United States, and in 2016, China already counted more than 350 million devices [30]. Another recent report from 2022 also shows significant progress in smart meter deployment in Europe [31]. Most EU countries are expected to achieve a deployment rate of at least 80% by 2025. Ten countries have already achieved this and are in the process of adopting second-generation devices. As a result, smart meter data (SMD) is becoming increasingly available at the household level, offering new opportunities for monitoring high-load devices in the grid, such as HPs [27].

In this context, our work is a first step towards a unified service for remotely monitoring HPs in residential buildings using SMD. We focus on evaluating HP cycling behavior (i.e., the frequency of switch-on and switch-off operations) as it affects performance and mean time to failure, and is an indicator of improper sizing and non-optimal settings [32–35]. Therefore, we analyze and describe the cycling of 503 HPs in the field in Switzerland over a 21-month period. To the best of our knowledge, this is the first study to monitor HP operation on such a large scale with SMD. To this end, our work particularly makes the following contributions:

- We show that cycling behavior can be derived from common smart meter data. We present an algorithm that extracts key indicators from 15-minute resolution data available for the vast majority of smart electricity meter installations.
- We show that operational characteristics related to cycling allow for the detection of outliers with respect to energy consumption and (in)appropriate sizing of HPs. This is advantageous because it does not require contextual information.
- We identify thresholds for atypical cycling behavior of HPs and describe behavioral differences with respect to building and HP characteristics.

The results are valuable to HP manufacturers and energy consultants, who can benefit from decision boundaries that allow them to identify malfunctioning equipment. The work is also relevant for companies developing remote services. The algorithms we propose can serve as a basis for monitoring HPs, even if they are connected not through the internet but through traditional SMs. This makes it possible to include most HPs in modern service offerings, since a large proportion of households is equipped with a SM. Finally, our work presents a promising use-case for utilities, which can use SMD to monitor HP activity and leverage their flexibility for demand response programs or targeted energy efficiency campaigns. A more detailed description of the benefits of our approach in terms of applicability to real-world use-cases is given in Section 5.3.

The remainder of this paper is structured as follows: First, we provide an overview of related work (Section 2). Afterwards, we introduce our data set (Section 3.1), cluster households by their building and HP characteristics (Section 3.2), and evaluate them in terms of energy efficiency and appropriate sizing (Section 3.4). Next, we derive information about each HP's individual cycling behavior (Sections 3.6 and 3.7) before including the effect of outdoor temperature (Section 3.8). This information is then used to detect HPs that cycle atypically (Section 3.9). In Sections 4 and 5, we report and discuss corresponding results considering different building and HP characteristics. Lastly, Section 6 provides a summary and conclusion.

2. Related work

In the following, an overview of existing work on HP cycling and HPs in the context of SMD is provided.

2.1. Heat pump cycling

The heat load of a building increases with decreasing outdoor temperatures. In order to ensure heating comfort, the *heating curve* defines increasing set-point temperatures as outdoor temperatures decrease [36]. Usually, the heating control monitors the return water temperature to adjust it to the set-point temperature [33]. Therefore, HPs show heating activity while the return temperature is lower than the set-point temperature. This leads to regular switch-on and switch-off events and thus, a cycling behavior. In the following, we refer to the time between consecutive switch-on and switch-off events as a cycle.

Power consumption during standby and off phases is negligible [37]. Instead, a significant amount of energy is consumed during heating cycles. Therefore, a few studies have investigated the relationship between heating cycles and energy consumption and found a strong correlation between short cycles and energy losses [32–34]. The study [32] puts these losses at 5%–30%, while [14] reports 12%. The reason for the cycling losses is that during the start-up phase of HPs, the delivered heating power is reduced until a steady state is reached [33]. This time interval depends on the characteristics of an HP. For example, in the study of [32], the HP reaches 90% of the steady-state value after 3 min, which the authors consider relatively short. The same study [32] concludes that a minimum time of 15 min between two consecutive cycles should be ensured. The work of [38] comes to a similar conclusion with a minimum run time of 20 min per cycle. Since a recurring cause of short cycles is inefficient defrosting schedules [35], short cycles occur more frequently in air source heat pumps (ASHPs) than in ground source heat pumps (GSHPs) [34]. Moreover, frequent

cycles reduce the lifetime of HPs [39]. However, it is not only start-up losses that reduce performance; long run times [32], which often relate to weather, do as well. As outdoor temperatures decrease, the thermal output of an HP system also decreases, and consequently the run time of the HP increases to compensate for this [34]. It can be concluded that neither short nor long cycles are desirable and that run time monitoring and optimization can improve the overall system performance [40,41]. Further, an evaluation of cycling behavior must take outdoor temperatures into account.

In addition to outdoor temperatures, other factors influence the cycle length and number of cycles; for example, HP sizing, as described in [42,43]. On the one hand, an undersized HP may run almost continuously, sometimes even alongside conventional water heaters, likely offsetting or even reversing the energy savings associated with the smaller HP size. On the other hand, oversized HPs tend to cycle on and off frequently, resulting in shorter cycles and negatively affecting performance. Another factor of influence is the type of HP installed in terms of its modulation capacity [44]. In the work of [33], a 12% drop in performance is observed when a fixed speed HP is used compared to a variable speed HP. A fixed speed HP can only rotate the compressor at a single speed and tends to have more and shorter cycles [29]. In contrast, a variable speed HP can modulate compressor speed to regulate for heat demand, which typically results in higher total run times but fewer on–off transients [29]. Finally, the installation of a buffer tank can also increase the length of a single cycle and decrease the number of cycles [34].

The studies presented give a good indication of the fundamental importance of HP cycling for energy efficiency. However, they all have in common that they analyze only a few individual HPs under laboratory conditions or in test houses and up to now it has been unclear how HPs operate in different homes. This is a fundamental difference to our study because we analyze 503 HPs in the field to investigate commonalities and differences in terms of cycling behavior. Additionally, none of the existing studies explains which cycling behaviors are typical or atypical for HPs, whereas we identify HPs that cycle atypically. Moreover, our work differs from previous studies in terms of the type of data used. Existing work uses HP sensor data and known contextual information because both are available for test houses or in laboratory environments. However, for real-world applications, methods are needed that can deal with missing contextual information or use publicly available data sources. Therefore, we derive context using a national building register database and a geographic information system. While this information is useful for evaluating energy efficiency and appropriate sizing of an HP, our method to evaluate the HP cycling behavior does not require this contextual information. Further, instead of sensor data, we use smart meter data because it is commonly available, can cover HPs that are not connected to the internet, and is independent of the HP manufacturer, making it suitable for wide-spread energy efficiency services. In this context, below, we briefly review existing related work at the intersection of SMD and HPs.

2.2. Heat pumps and smart meter data

The presence of HP systems in residential buildings can be predicted by combining SMD with weather data [45,46]. While [45] uses data at 15-minute resolution and [46] at daily resolution, both have in common that they extract features from time series to use as input for classification algorithms. A similar approach is used in [29], where variable and fixed speed heat pumps are distinguished using SMD. An extension of HP detection with SMD is non-intrusive load monitoring, also known as load disaggregation. In the context of heating, the goal is to isolate heating-related patterns from the SMD when the heating system is measured along with other appliances. Since this area is reviewed in [47–49] in general and in [50] with a particular focus on HVAC systems, we do not discuss it in detail here. However, we note that only a few works disaggregate HP patterns from low resolution energy data (e.g., 15 min as in our case). Most studies either use data with higher frequencies and dimensions [51–55], or they focus on other devices with rather constant power consumption such as electric resistance heaters [56–58] or electric water heaters [59–61]. However, two studies extract HP patterns from low-resolution SMD using clustering algorithms [62,63]. The first work [62] uses SMD with a resolution of 5 minutes and clusters dominant on and off activities found by a peak detection algorithm. This also means that the proposed solution cannot cope with variable speed HPs, where power consumption modulates in small steps proportional to compressor speed [29], and the study lacks an evaluation using SMD with a more common resolution of 15 min. The second work [63] uses hourly data, but does not aim to isolate and obtain a complete HP pattern. Instead, an HP pattern is automatically divided into flexible and non-flexible loads. What both studies have in common is that they are ultimately used to estimate flexibility as part of demand response programs. Other studies also look at HP scheduling in this context [64–68]. For example, a study from the Netherlands investigates the flexibility of HP schedules to reduce operating costs by using dynamic tariffs [69] and another work uses SMD and Wi-Fi data to detect occupancy and propose HVAC schedules in commercial buildings [70].

In summary, previous work on HPs and SMD has focused on detecting HP installations, isolating their patterns from comparably high frequency measurements, or estimating flexibility for demand response programs. There is a lack of work monitoring active HPs in operation with SMD. However, more importantly there is no study to identify HPs which operate energy inefficiently, are inappropriately sized, or cycle atypically using SMD, which is addressed in this study.

3. Methods

In this section, the data set is described, the methods for evaluating the HP cycling behavior are explained, and the baseline models serving as ground truth with respect to HP energy efficiency and appropriate sizing are derived.

3.1. The data set

The data set includes 503 single-family houses in Switzerland with an observation period from January 2021 to September 2022 (21 months). Each household has an HP for heating purposes, but no photovoltaic system.

Smart Meter Data: For each household, we use the measurements of HP electricity consumption in kWh in 15-minute resolution. The HP is measured separately from other appliances. Each SMD time series differs in terms of start and end dates and missing data. The average data availability per household over the entire period is 61%. Each HP has an average of 395 days of data without outages.

Temperature Data: To enrich the SMD with data on average outdoor temperature in daily resolution, we use the address of each household and query a paid weather service. It individually finds the nearest weather station and returns the temperature measurements (in °C) as integers without decimal places, e.g. $t_{avg} = 5°C$.

Meta Data: For 342 households (i.e., 68%), the installed electric power of the HPs in kW is known through the utility company. Additionally, we use the address of each household to look up the building year and heated floor area from the Swiss building register [71]. A study on behalf of the Swiss Federal Office of Energy [72] suggests applying a correction factor of 1.15 to the heated floor area and treating floor areas outside the range of 70–400m^2 as outliers. We follow this advice and treat the outlier households as if the information was not available. Accordingly, we do the same in terms of the building year with buildings that were built before 1700. Additionally, for some households, no entry is found in the building register, or the corresponding fields are not filled. In total, this leads to the heated floor area being available for 393 households (i.e., 78%) with a median of

Table 1
Characteristics of the clusters of households, when applying K-Means clustering to each of the meta data variables individually.

Category	Cluster	Count	Minimum	Mean	Median	Maximum
Building Size (Floor Area in m^2)	small	157	74.75	151.00	151.80	188.60
	medium	167	189.75	226.79	224.25	267.95
	big	69	269.10	309.35	305.90	391.00
	unknown	110	–	–	–	–
Building Age (Building Year)	old	21	1732	1813	1820	1876
	medium	60	1888	1947	1954	1974
	new	308	1976	2002	2004	2016
	unknown	114	–	–	–	–
HP Size (Electric Power in kW)	small	229	1.00	2.89	3.00	4.34
	medium	95	4.40	5.83	5.50	9.00
	big	18	9.70	12.67	12.00	18.00
	unknown	161	–	–	–	–
HP Type	air source	188	–	–	–	–
	ground source	125	–	–	–	–
	unknown	190	–	–	–	–

207 m^2. The building year is available for 389 households (i.e., 77%) with a median of 2002. Lastly, we again use each household's address to extract the existence of a drilling profile through a publicly available geographic information system [73]. If an entry is found, we define a HP with a drilling profile as ground source and without as air source heat pump. For 190 households (i.e., 38%), no entry is found. We mark the corresponding HPs to be of unknown type. The other 313 (i.e., 62%) are composed of 125 ground source (i.e., 40%), and 188 air source heat pumps (i.e., 60%).

3.2. Categorizing households by characteristics

Next, we form groups of similar households to describe and compare the behavior of their HPs in a later step. For this purpose, we use the contextual information of each household as derived earlier: the building year, the heated floor area, and the installed electric HP power (see the previous section on the meta data of our data set). If this information is unknown for a household, we immediately assign the household to a cluster named *unknown*. For all other households, we apply K-Means clustering in a one-dimensional manner to each variable individually and set the cluster size to 3 (initialization of K-parameter). The reasons are that the clusters become easily interpretable by their cluster centers and that they can be combined in a flexible way. Hence, a household is clustered multiple times but only once for each variable. The clustered heated floor area refers to the size of the house (small, medium, big, unknown). Similarly, the installed electric HP power serves as metric for the size of HP installation (small, medium, big, unknown), and the building year as metric for the building age (old, medium, new, unknown). Additionally, since the HP type is categorical, we interpret it as already being clustered (air source, ground source, unknown). Table 1 shows the results of this process and provides the final characteristics of each cluster in terms of cluster size, and minimum, mean, median, and maximum of the observations within a cluster.

3.3. Daily observations and heating applications

In what follows, the behavior of HPs is analyzed in daily observations for several reasons: First, by choosing a daily resolution for the extraction of cycling metrics, they can be matched to the weather variables that are available in daily resolution. Second, daily behavior is easier to interpret; for example, behavior on a cold winter day can be compared to behavior on a spring day with mild temperatures. Third, smart meters in our setting report the measurements of an entire day all at once. Hence, if data is missing, it is missing for an entire day and there are no gaps in SMD within a day. Fourth, heating systems have some inertia and do not respond immediately to changes in outdoor temperature because insulation keeps heat in the building. Therefore, interpreting the cycling response with a resolution of a few minutes or hours is not useful without detailed information about the underlying exergy models of individual buildings.

In addition, this work focuses on HPs in heating mode, but not on cooling or hybrid applications. We limit our descriptions and analyses to identify conspicuous HPs on days with average outdoor temperatures of 0–12 °C. We choose the upper limit of 12 °C because according to the Swiss standard, days with an average outdoor temperature below this value are heating days [74]. Accordingly, we choose the limit of 0 °C because below this value, an HP may not operate in monovalent mode but may be supported by an electric auxiliary heater [19]. In addition, days with average temperatures below 0 °C are rare in central Europe, so incorrect conclusions may be drawn from distributions with few observations at cold temperatures.

3.4. Evaluating energy efficiency and appropriate sizing of heat pumps

To evaluate the energy efficiency and the appropriate sizing of each HP, we use the corresponding distribution of daily energy consumption and contextual information on heated floor area and electrical power.

3.4.1. Normalizing daily energy by degree day:

Typically, an HP's energy consumption increases with decreasing outdoor temperatures. This means that the distribution of daily energy depends on temperature. To make the HPs comparable without weather effects, we use a common approach that normalizes the daily energy sums by a simple division by degree day, as done for example in [75]. We use a base temperature t_{base} of 20°C, which is the Swiss norm value of an artificial indoor temperature setting [74]. Then we compute the degree day (in °C) from the average outdoor temperature t_{avg} as follows:

$$\text{degree day} = |t_{avg} - t_{base}| \text{ with } t_{base} = 20°\text{C} \quad (1)$$

3.4.2. Calculating energy intensity as a measure of energy efficiency

A common approach to assess the energy efficiency of residential buildings is to calculate energy intensity [75]. This approach assumes that a building's energy consumption is proportional to its floor size. Therefore, for each HP, we divide the daily energy values, already normalized by degree day (Eq. (1)), by the heated floor area. The median is then calculated from the resulting distribution. For better readability, we refer to this single value in the following only as *energy intensity* (in $kWh/m^2/°C$). It can be viewed as a measure of energy efficiency that makes the HPs comparable in energy consumption regardless of building size and temperature. An HP with high energy intensity can be considered less efficient than an HP with low energy intensity.

3.4.3. Calculating utilization as a measure of appropriate sizing

To assess whether an HP is appropriately sized, we calculate its daily utilization u_{day} (in %) using the daily energy E_{day} (in kWh) and the installed electrical heat pump power P_{HP} (in kW) as follows:

$$u_{day} = \frac{E_{day}}{P_{HP} * 24h} * 100\% \quad (2)$$

Again, we divide the daily utilization values by degree day values (Eq. (1)) to eliminate the effects of outdoor temperature and then calculate the median daily utilization of each HP from the corresponding distribution. In what follows, we refer to this value only as *utilization* (in %/°C) and consider it a measure of appropriate sizing that is independent of weather and HP size. An HP with very low utilization can be considered rather oversized, while an HP with very high utilization can be considered rather undersized.

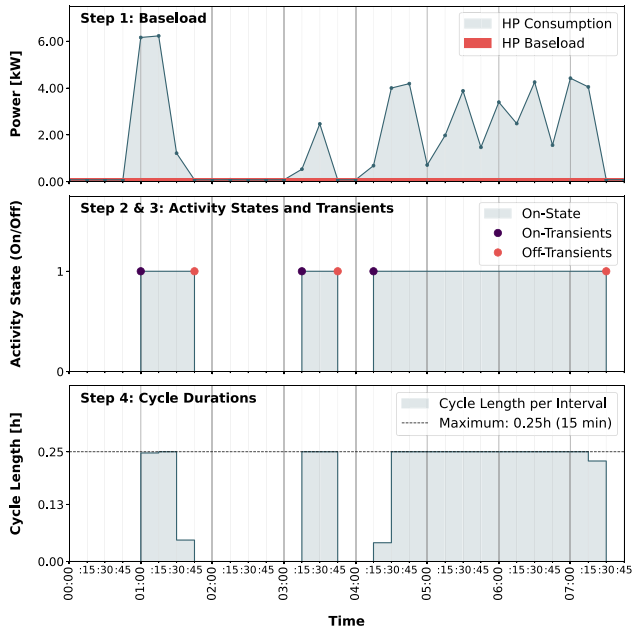


Fig. 1. Example of the different steps of extracting cycles from smart meter data. For better understanding, in the first graph the energy measurements (in kWh) are converted into power (in kW) by multiplying them by four.

3.5. The role of contextual information

The evaluation of energy efficiency and appropriate sizing of HPs is possible with SMD and contextual information. However, usually such required information (e.g., about the heated floor area and electrical power of each HP) is not known or difficult to obtain [76]. In this study, the address of each household is used to extract these parameters from the Swiss building register. Nevertheless, even in this advantageous scenario, the heated floor area is unknown for about 22% of households, and the electrical power for about 32% of HPs. An even more common scenario is that due to privacy constraints, the address is either unknown or cannot be used, making the use of national building register databases impossible [77]. Also, most countries do not offer a well-maintained building database like Switzerland. Therefore, we develop a solution to assess whether an HP requires special monitoring or exhibits conspicuous behavior by relying solely on SMD at common 15-minute resolution and temperature data at daily resolution measured by a local weather station nearby. Not only the SMD but also the weather data is usually available to utilities because the city or region from which the SMD originates is known. The cycling behavior of an HP is a meaningful indicator of whether it is energy efficient and appropriately sized (see Section 2.1). Therefore, in the following, we develop a novel method to extract HP cycles from SMD, derive key indicators to describe the cycling behavior, and consequentially identify HPs that cycle atypically. In this way, our method is independent of additional contextual information and therefore suitable for large-scale real-world applications.

3.6. Extracting heat pump cycles

In the following, we explain how we extract heating cycles of the individual HPs from SMD. For better readability, we only briefly describe each step in this section and provide detailed descriptions of the algorithms in the appendix (Appendix A.1). Fig. 1 serves as supporting graphic, which provides an example of the different steps.

Step 1 – Estimating Baseload: Even in standby mode, an HP has an energy consumption that is slightly above zero [29]. The control unit causes a baseload, which we estimate individually for each HP from the distribution of SMD measurements. *(Algorithm in Appendix A.1.1)*.

Step 2 – Deriving Activity States: We assume that any energy consumption above the estimated baseload is due to the HP performing a cycle. Therefore, we compare the energy consumption at each timestamp to this threshold to derive binary activity states (on-state: HP performs a cycle; off-state: HP is in standby or off). *(Algorithm in Appendix A.1.2)*.

Step 3 – Determining On- and Off-Transients: Next, we determine the changes in activity states to find the time points associated with an on-transient and an off-transient of each cycle. It is possible that at certain times the energy consumption exceeds the threshold, but the previous and subsequent measurements do not. In this case, we mark the corresponding timestamp as both an on-transient and an off-transient and assume that it is caused by a cycle shorter than the 15-minute measurement interval. *(Algorithm in Appendix A.1.3)*.

Step 4 – Calculating Durations of Individual Cycles: From the activity states and transients, we calculate the duration of each cycle in hours. An HP can turn on or off at any time within the 15-minute measurement interval. Therefore, the measured energy of the switch-on and switch-off operations is usually lower than during the cycles [29]. We assume a constant electricity uptake between an on-transient and the subsequent measurement, or an off-transient and the previous measurement. Therefore, by accounting for the difference between the observed energy measurements, we can calculate the exact times within a 15-measurement interval that the HP turns on or off. While this assumption holds for fixed-speed HPs, it may be incorrect for variable-speed HPs because they can modulate compressor speed from one measurement to the next. However, we argue that the estimates using our method better reflect the actual cycle durations, since otherwise each on- and off-transient would be counted as a full 15 min, resulting in large overestimates. For better readability, we report all cycle durations in hours rounded to two decimals. *(Algorithm in Appendix A.1.4)*.

Table 2
Description of the metrics that serve as key indicators for daily cycling behavior.

Metric	Description	Unit
Operating Hours	Sum of cycle durations. Describes total hours of cycling activity per day.	h
Cycles	Sum of on-transients. Describes total number of cycles per day.	#
Cycles/Operating Hours	Sum of on-transients divided by sum of cycle durations. Describes number of switch-on operations per hour of cycling.	#/h
Average Cycle Length	Average of cycle durations. Describes average length of a single cycle. Equal to the inverse of cycles/operating hours.	h

3.7. Calculating daily cycling metrics

From the procedure described in the previous section, for each HP, a list is obtained, indicating the start and end time and duration of each cycle. Now, we convert this information into aggregated metrics that describe individual days and serve as key indicators for cycling behavior (recall that in Section 3.3, it is explained why a daily resolution is used). The metrics calculated per day are operating hours, number of cycles, ratio of number of cycles to operating hours, and average cycle length. Table 2 provides a description of these metrics, and the algorithm in Appendix A.2 of the appendix formalizes the calculation.

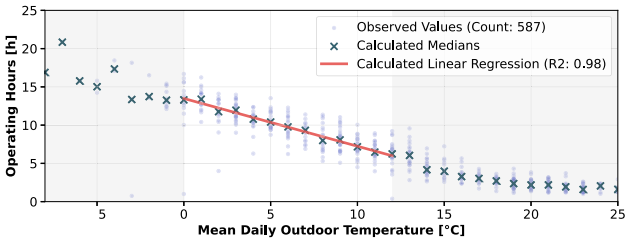


Fig. 2. Example of daily operating hours observed for a single HP. The graph shows the observations grouped by the average daily outdoor temperature to obtain distributions, their medians, and a linear regression on the medians. The latter is limited to the temperature range of 0–12 °C (see explanations in Section 3.3).

3.8. Deriving temperature curves

For days with the same average outdoor temperatures, a single HP shows similar but slightly varying cycling behavior. Therefore, the daily cycling metrics can be interpreted as distributions that depend on the average outdoor temperature. Below it is explained how to handle this, while Fig. 2 serves as a supporting graph.

First, for each HP and metric, we group the daily observations by mean daily outdoor temperature to derive the distributions. We then reduce the dimensionality and the effects of outliers by calculating the median of each distribution. In this way, one value per temperature and per metric is obtained. When grouping the medians by each metric and then sorting them in descending order by outdoor temperature, we obtain what we refer to as a *temperature curve*. Note that the analyses of this paper focus on 0–12 °C temperature range (as previously explained in Section 3.3), where an almost linear behavior can be observed (see descriptions in Section 5.2). Therefore, for each metric, we compute a linear regression on the medians between 0–12 °C. In this way, the cycling behavior of HPs becomes comparable through the slopes and intercepts of the linear approximations and independent of the observation period.

3.9. Detecting conspicuous heat pumps

We hypothesize that atypical cycling is a good indicator of whether an HP requires special monitoring or system optimization. Therefore, we determine which systems cycle typically and which cycle atypically by examining each metric individually and interpreting the slopes and intercepts of the linear regressions as bivariate distributions which span two-dimensional feature spaces. This allows us to apply an outlier detection method and interpret a system within the normal range as typical and an outlier system as atypical. We determine outliers in an unsupervised manner using the local outlier factor (LOF) [78]. LOF considers the local density of each sample and compares it to the local density of its nearest neighbors. Then, it computes a score for how isolated the object is in the local neighborhood and identifies samples with much lower density as outliers. We use the implementation in [79], which relies on the k-nearest neighbor algorithm and the Minkowski distance to calculate the local density. We also use the default setting of [79] that considers 20 neighboring data points and algorithmically determines the number of outliers. However, note that the parameters can be set to classify a predefined percentage of data points as outliers, making the approach suitable for household selection in campaigns. The fact that an outlier score is returned for each data point also makes it possible to sort the HPs according to how typical or atypical they behave with respect to a single metric.

We apply LOF individually to each cycling behavior metric and classify an HP as conspicuous if it is determined to be an outlier by any of the cycling metrics. This is a rather sensitive approach that does not allow HPs to deviate in any parameter. To be less sensitive, a minimum number of parameters could be introduced that an HP must violate to be considered atypical. In any case, for each HP it can be listed which metrics cause it to be considered conspicuous. In addition, for each metric, we calculate the means and standard deviations of the slopes and intercepts of the identified inlier points. Based on the position of an outlier point relative to the ranges of mean ± standard deviation, additional clues can be provided as to what makes the system an outlier. Hence, it can be inferred whether the slope, the intercept, or both are atypical (too high or too low, or rather okay).

Fig. 3 shows an example of the method when applied to the ratio of cycles to operating hours. Each point in the figure refers to a single HP. In the first plot, the radius of a circle represents the outlier value, i.e., the larger the circle, the more isolated the point. The color encodes whether an HP has been determined to be an outlier and thus conspicuous. The second graph shows only the identified outliers and indicates whether the slope or intercept is too high compared to the inlier points (indicated by the corresponding range of mean ± standard deviation). The third graph shows a two-dimensional kernel density estimate of outlier and inlier points for further insight.

Table 3
Statistics describing the distributions of R^2 values of the linear regressions on the temperature curves as derived in Section 3.8 and used in Section 3.9. The statistics are reported separately for each cycling metric across the distributions of the 243 HPs under evaluation. They therefore indicate how well a metric can be linearly approximated in the 0–12 °C temperature range.

Metric	Median R^2	Mean R^2	Standard deviation of R^2
Operating Hours	0.96	0.88	0.19
Cycles	0.73	0.62	0.31
Cycles/Operating Hours	0.82	0.71	0.28
Average Cycle Length	0.80	0.70	0.28

4. Results

In the following, the methods proposed in the previous section are applied to the data set presented in Section 3.1. We compare the results of our outlier detection based on cycling behavior (Section 3.9) with the ground truth obtained from the baseline models (Section 3.4) that evaluate energy efficiency and appropriate sizing of HPs. Because the baseline models require contextual information on heated floor area and electrical power, the assessments in this chapter are based on a subset of 243 HPs for which both parameters are available. For all of these HPs, observations are available for at least 6 different average daily outdoor temperatures of the 13 possible values in the 0–12 °C temperature range.

4.1. Evaluating linear fit of temperature curves

To detect HPs that cycle atypically, we use the slopes and intercepts of the linear regressions on temperature curves (see Section 3.8). Therefore, we first assess the fit of these regressions by examining their *residuals-versus-fitted plots* and calculating the R^2 value of each regression. The descriptive statistics of the resulting R^2 distributions for each metric are reported in Table 3. All medians of the R^2 distributions range from 0.73 to 0.96, which we interpret as an indication that all cycling metrics can be well approximated linearly in the 0–12 °C temperature range. The best approximation is obtained for the hours of operation and the worst for the number of cycles.

4.2. Identified conspicuous heat pumps

The algorithm automatically finds a total of 41 conspicuous HPs, which corresponds to 16.9% of the 243 HPs evaluated. Note that the algorithm allows a system to be classified as an outlier with respect to several key indicators simultaneously. In 12 of the 41 cases (i.e., 29.3%), an outlier system is classified as such by more than

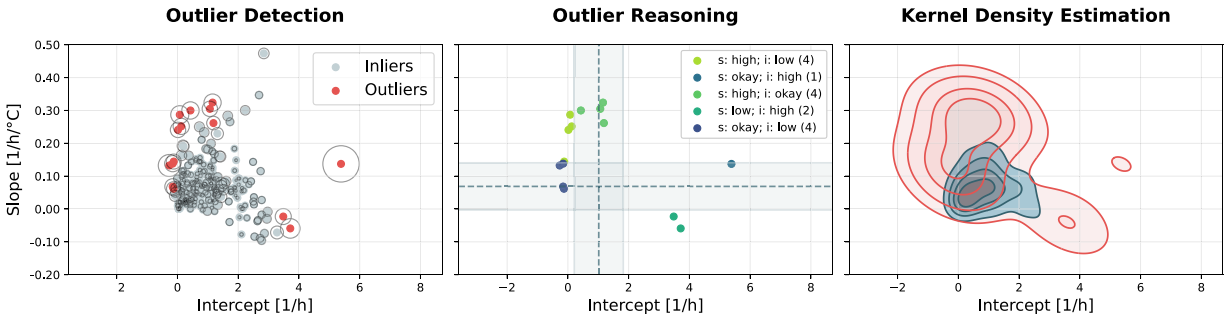


Fig. 3. Example of determining HPs that cycle atypically by applying outlier detection to the ratio of cycles to hours of operation. We apply this procedure to each cycling metric individually and classify an HP as conspicuous if it is an outlier by any metric.

Table 4
The 41 conspicuous HPs by clusters. The percentages represent how many systems of each population have been identified to cycle atypically.

Category	Cluster	Population size	Outliers count	Outliers percentage (%)
Building Size	small	104	19	18.3
	medium	101	17	16.8
	big	38	5	13.2
Building Age	old	14	2	14.3
	medium	37	10	27.0
	new	163	24	14.7
	unknown	29	5	17.2
HP Size	small	165	25	15.2
	medium	67	14	20.1
	big	11	2	18.2
HP Type	air source	98	23	24.0
	ground source	65	6	9.2
	unknown	80	12	15.0

one key indicator. For average cycle length, 26 outliers are identified, followed by the ratio of number of cycles to operating hours (17 outliers), hours of operation (9 outliers), and number of cycles (5 outliers). Table 4 lists the clusters to which the identified systems belong. To account for the different cluster sizes, not only the absolute numbers are reported but also the percentage of conspicuous HPs in each population. It can be noticed that the proportion of identified outliers that are ASHPs is more than twice the proportion of GSHPs (24.0% vs. 9.2%). In addition, the proportion of identified systems installed in middle-aged buildings is significantly higher than in buildings of other ages (medium: 27.0%; new: 14.7%; old: 14.3%).

In Fig. 4, examples of the original SMD of noticeable outlier and inlier HPs by operating hours and number of cycles are shown. The figure shows heat maps, with rows representing time of day and columns representing dates. This improves the ability to observe cycling patterns over time. The visualizations show that our algorithms are able to distinguish HPs with short cycles and very frequent on–offs from HPs with longer and healthier cycles.

4.3. Evaluating performance

Our method identifies atypical HPs by analyzing cycling behavior without requiring contextual information. To evaluate performance, we compare our results to baseline models that assess energy efficiency and appropriate sizing of HPs but require contextual information (Section 3.4). To this end, we treat the problem as a binary classification with binary labels indicating whether an HP is salient (label: 1; minority class) or not (label: 0; majority class). Consequently, we label the 41 HPs that are conspicuous for cycling behavior as salient and the other 202 HPs as not salient. At this point, note that the LOF-based outlier detection (as introduced in Section 3.9) can automatically select HPs with high or low activity as long as they behave atypically. Therefore, for ground truth, we treat the best and worst 10% of HPs in terms of utilization and energy intensity as atypical to ensure a fair comparison with the baseline models. Fig. 5 shows the resulting confusion matrices

Table 5
Performance scores of outlier detection by cycling behavior when compared to baseline models evaluating energy efficiency and appropriate sizing but requiring contextual information.

Baseline model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Cohen's Kappa
Energy Intensity	0.80	0.51	0.42	0.46	0.66	0.40
Utilization	0.81	0.54	0.44	0.48	0.67	0.37

with absolute and relative scores. Table 5 shows the corresponding performance scores typically used for classification tasks. In addition, the table includes the Cohen's Kappa statistic, which is a measure of agreement between two annotators, rather than comparing a classifier's predictions to the ground truth values.

According to [80], the Cohen's Kappas of 0.40 for energy intensity and cycling behavior and 0.37 for utilization and cycling behavior indicate a fair agreement in both cases. In a machine learning setting, the high accuracy values of the model (0.80 and 0.81) combined with the rather low F1-scores (0.46 and 0.48) could be interpreted as an insufficient fit to an unbalanced data set. However, in this case, a more complex investigation and interpretation is required. The model does not necessarily fail to find conspicuous HPs that require special monitoring. Instead, it focuses on a different aspect of optimization. By identifying HPs that are conspicuous in cycling, it finds a proportion of HPs that are atypical in energy intensity or utilization, but some HPs can also cycle atypically without being inefficient in terms of energy use or inappropriately sized. In this case, cycling behavior could still have an impact on the lifetime of the entire device or its components, and improvements in operation could still be possible. Therefore, we consider special monitoring of HPs that are conspicuous by their cycling behavior to be appropriate in any case.

5. Discussion

In the following sections, additional descriptive analyses are provided as well as insights into factors that influence cycling. Further, limitations and future work are listed. Unlike the evaluation in the previous chapter, in this chapter all analyses are conducted for all 503 HPs rather than for just a subset of HPs.

5.1. Analyzing correlations

To investigate how the cycling metrics relate to each other, to temperature, to utilization, and to energy intensity, we calculate the correlations of the daily measures across all households. Fig. 6 shows the resulting correlation matrix.

We find that all measures of cycling are correlated with outdoor temperature, most strongly with hours of operation with a negative correlation of −0.60, followed by number of cycles (−0.43). In contrast, utilization and energy intensity show little correlation with temperature. This confirms that the normalization by degree day introduced

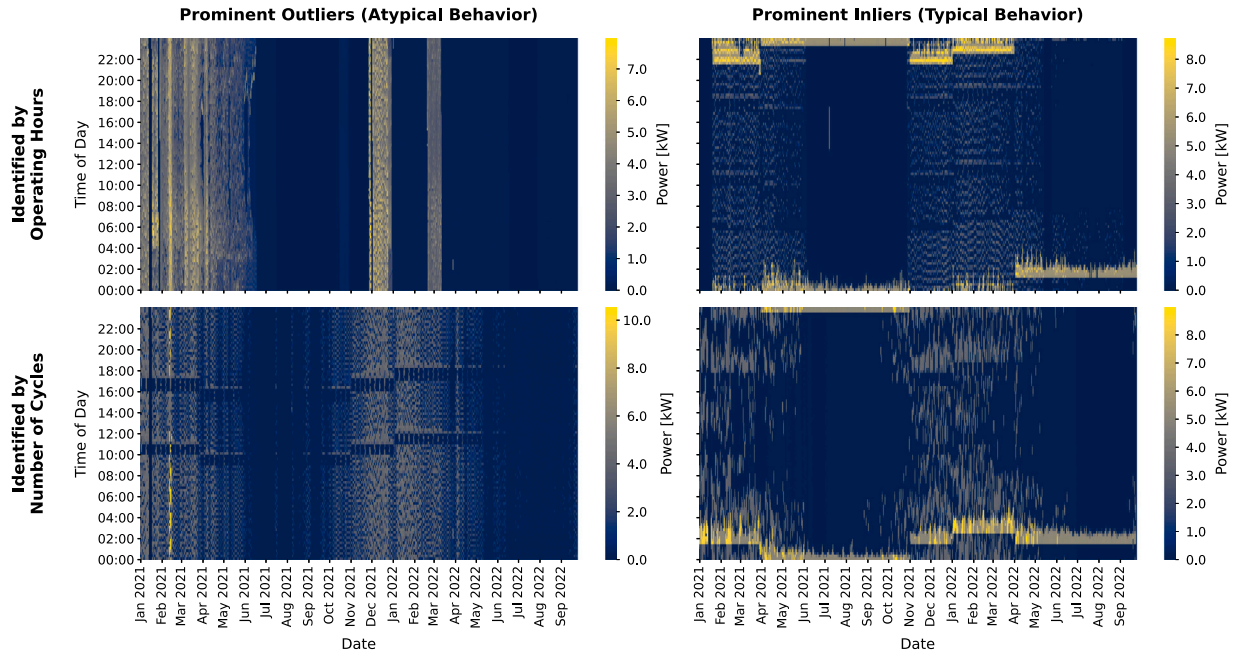


Fig. 4. Examples of HPs that the algorithm determines as typical or atypical with respect to operating hours and cycles. The visualizations show original smart meter data in 15-min resolution but transformed to power in kW (multiplication by four) and heat maps are used to display the cycling patterns over time.

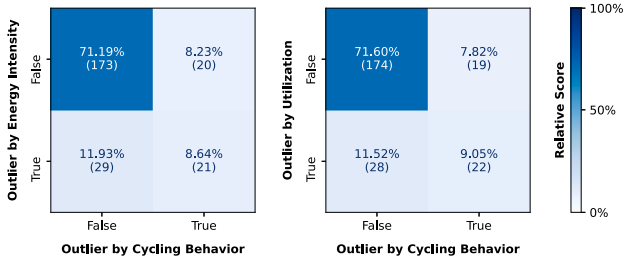


Fig. 5. Confusion matrices with absolute and relative scores. The matrices compare the conspicuous HPs identified by our method with the best and worst 10% of HPs in terms of utilization and energy intensity as given by the baseline models.

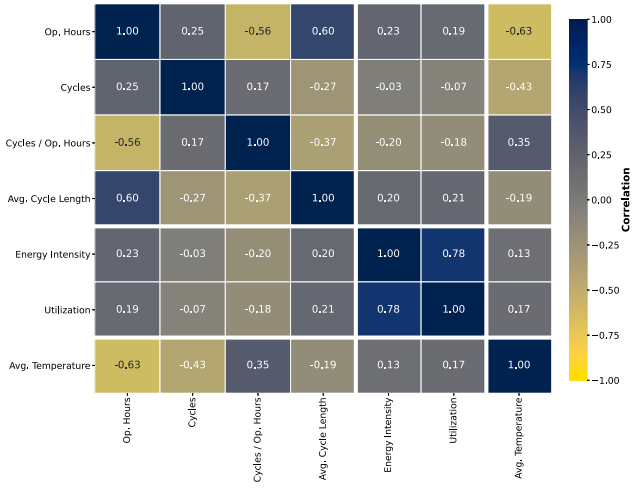


Fig. 6. Correlations of cycling metrics, energy intensity, utilization, and outdoor temperature at daily resolution for all HPs.

in Section 3.4 has the desired effect of removing weather dependency. However, utilization and energy intensity are strongly correlated, with a factor of 0.78. In contrast, the metrics for cycling behavior are relatively independent of utilization and energy intensity, as reflected in the low correlation factors. It could be that a combination of these metrics (as used in the outlier detection of our approach) has a stronger relationship with the energy intensity and utilization. However, this may not be well reflected here but supports the interpretation of the performance values reported in Section 4.3, where it has been concluded that HPs can also cycle atypically without a direct link to energy efficiency and sizing. Finally, operating hours and number of cycles are positively correlated (0.25) and so are average cycle length and operating hours (0.60).

5.2. Analyzing temperature curves

In Section 3.2, clusters of HPs are formed. Here, we use these to describe cycling behavior that is typical for each cluster. Since the features are static and categorical, correlations with daily cycling metrics cannot be computed as done previously. Therefore, we derive the temperature curves (see Section 3.8), but with different observations. Instead of calculating the medians for each HP separately, we calculate them again using all observations of the HPs that are part of the same cluster. This way, the plots in Fig. 7 are obtained, where each point reflects a median value and where error bars are reported at the 95% interval. In what follows, we explain our observations of typical cycling in all populations using this visualization. We restrict the descriptions to the temperature range of 0–12 °C (see Section 3.3). Note that also daily energy intensity and utilization from the baseline models (Section 3.4) are included, but without normalization by degree day, as this would eliminate the temperature effect, which is of interest at this point.

5.2.1. Typical heat pump cycling

From Fig. 7, it can be seen that most of the measured variables behave almost linearly in the 0–12 °C temperature range. Only the number of cycles shows saturation effects, as the behavior is linear between 5–12 °C and almost constant between 0–5 °C. This also explains why in Table 3, which evaluates the fit of the linear regressions, the number of cycles has the smallest R^2 values. The hours of operation increase with decreasing outdoor temperatures (about 4–7 h at 12 °C to

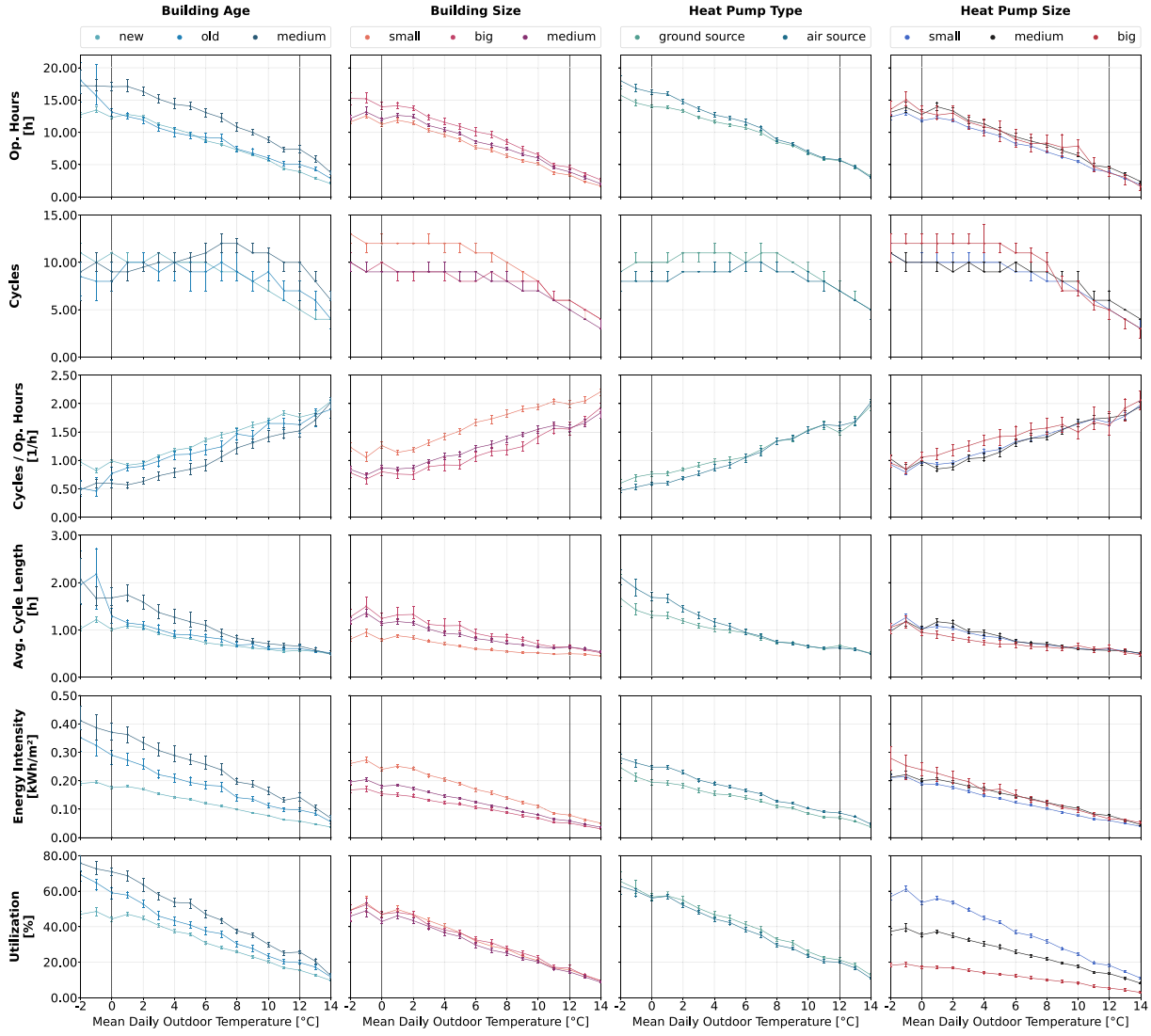


Fig. 7. Visualization of HP cycling behavior, energy intensity and utilization of all 503 HPs over mean daily outdoor temperatures by different building and HP characteristics. The points represent the medians of all observations across each population, while the error bars are at 95% intervals. To not remove the temperature effect in the graph, the daily utilization and energy intensity values are not normalized by degree day as otherwise done in Section 3.4.

13–17 h at 0 °C). The number of cycles also increases, but to a lesser extent (from about 5 cycles at 12 °C to about 12 cycles at 0 °C). Also, the average cycle length increases with decreasing outdoor temperatures, while the ratio of cycles to operating hours behaves antiproportionally and decreases. As a rule of thumb, it can be deduced that for days with an average outdoor temperature of 12 °C, the typical ratio of cycles to operating hours of an HP is about 1.7 to 2.0 and decreases linearly until it is 0.8 to 1.2 at an average temperature of 0 °C. Finally, the utilization also increases linearly with decreasing outdoor temperatures, starting at about 20% at 12 °C. A similar behavior can be observed for energy intensity.

5.2.2. *Influence of characteristics on cycling*

Although in Fig. 7 the HPs behave very similarly across all populations, we describe below some notable differences we find between clusters.

Building Age: For almost all metrics, HPs behave similarly in buildings of different ages. However, since HPs are a technology mainly suited for well-insulated buildings, the clusters are quite unevenly populated with 309 new and 21 old buildings (see Table 1). We observe that energy intensity (i.e., total daily energy per heated floor area) behaves as we would expect it. In general, newer buildings are better insulated, resulting in lower energy consumption. In this context, newer buildings have up to half the energy intensity than older buildings. The fact that this factor is not more extreme is most likely due to the fact that a building must have a certain level of insulation for an HP to even be considered as a heating technology. We therefore assume that a comparison of the cycling behavior of different HPs is possible quite independently of the building age.

Building Size: In terms of building size, the diagram shows that HPs in smaller buildings tend to have more cycles with fewer hours of operation at the expense of higher energy intensity. One possible reason for this behavior could be that smaller buildings tend to have oversized HPs, which typically have more but shorter cycles, as described in Section 2.1. This hypothesis cannot be proven by looking at the utilization, which shows almost no difference between buildings of different sizes. Another reason could be that larger buildings may use other additional energy sources for heating or domestic hot water production. However, also this hypothesis cannot be proven with our data set.

Heat Pump Type: Fig. 7 shows that ASHPs have more operating hours, fewer cycles, and a slightly longer average cycle length than GSHPs. It is known that ASHPs are less energy efficient than GSHPs [43]. Therefore, we conclude that ASHPs must operate more to compensate for the lower heat dissipation. However, we might have

expected more cycles, since defrosting is required. Instead, the number of cycles of GSHPs below 5 °C even exceeds that observed for ASHPs. Nevertheless, the known overall energy efficiency is reflected in the energy intensity curves. The curve of GSHPs is below the curve of ASHPs.

Heat Pump Size: Larger HPs have more cycles with a shorter average cycle length compared to smaller HPs. As explained earlier in the context of building size, larger HPs may be slightly oversized, resulting in more and shorter cycles. Therefore, the striking difference in utilization is as expected. Small heat pumps are up to four times more utilized than large ones, especially at cold temperatures.

5.3. Contributions and benefits of our approach

The methods presented here are focused on practical applications to monitor real-world HPs in operation. We have demonstrated how common smart meter data in 15-min resolution can be used to derive key indicators about HP cycling and how HPs that cycle atypically can be identified. Additionally, we have shown that HPs which are conspicuous in terms of cycling behavior are often also outliers with respect to energy efficiency and appropriate sizing. However, while evaluating a heat pump's energy efficiency and sizing requires contextual information, its cycling behavior can serve as indicator when the context is unknown. This makes our approach suitable for many applications in practice. Further, we designed our methods in such a way that they are robust to common difficulties in practice as listed below.

1. **Data Availability:** The approach takes advantage of the large amounts of smart meter data that become available through advanced metering infrastructure for residential buildings of all types. In particular, it also covers HPs without network connectivity. Our method is designed to be able to deal with missing data and different observation periods of different households. It only requires measurements of a few cold days.
2. **Comprehensive View:** By identifying heat pumps that are outliers in terms of energy efficiency, appropriate sizing, and cycling behavior, and comparing them against each other in these categories, the approach incorporates a comprehensive perspective on HP optimization.
3. **Contextual Information:** To properly evaluate performance and sizing of HPs, contextual information on the heated floor area and electrical power is required. Usually, this relevant additional information is unavailable. In this case, we show that the cycling behavior of an HP alone can serve as a stand-alone indicator to identify and prioritize critical systems. Therefore, it is also suitable for situations where the context is not known, e.g., due to privacy constraints.
4. **Interpretability:** The information derived from SMD (e.g., cycles per day) is easy to understand and can provide accurate feedback on HP behavior.
5. **Benchmarking:** Our method makes it easy to sort, compare and benchmark HPs by the metrics of interest. This makes it a versatile approach suitable for any type of campaign.
6. **Expandability:** The algorithm can be easily adapted by introducing additional key indicators of interest and treating them in the same way as the metrics already introduced. Furthermore, it can be easily adapted to handle other types of data, e.g., sensor data or different resolutions.

5.4. Limitations

This study is a first step towards investigating real heat pump cycling on a large scale. However, below we list some limitations. First, the analyses only cover single-family homes in Switzerland that are not equipped with a photovoltaic system, and the results could be different in other geographic regions or for other types of buildings. Second, this study focuses on HPs used for heating and does not further examine whether the HPs under investigation are used in hybrid applications, i.e., also for cooling. Third, it is not distinguished whether or not an HP is responsible for the production of domestic hot water, since this information is unavailable. Finally, the 15-minute resolution limits the scope of what can be observed. For example, two cycles that are shorter than 15 min but run consecutively would be observed as a single cycle. However, given these limitations, this study shows that SMD is a powerful source of data for monitoring HPs, especially given its already high and increasing availability. It may be impossible to obtain detailed information on cycling, and there may be errors in individual cycles, but since these problems are shared by the entire population, the overall trends of individual households can still be observed. These can provide valuable feedback to HP owners, utilities, and other stakeholders.

5.5. Future work

The results of this work are based on SMD, which measures HPs separately from other devices. Therefore, the methods presented can be extended to aggregate measurements, which requires an additional step of disaggregating the HP load from the total load. On the other hand, future work on non-intrusive load monitoring focusing on HP applications may also incorporate the typical HP behavior that has been identified in this study as prior information. Additionally, it remains an open task to investigate the impact of daily temperature range on HP cycling. We use only the average outdoor temperature of a day, but do not consider the difference between the maximum and minimum temperature of a day. However, the most important remaining part of the work to be addressed by the research community is to investigate the reasons for differences in cycling behavior. There is a lack of work explaining under which conditions short and long cycling occurs and how it can be optimized. Here, we see a limitation in the use of SMD, but believe that sensor data with higher resolution and more parameters can help. Additionally, more contextual information about the underlying buildings and system components is needed.

6. Conclusion

In this study, we aim to contribute to a longer lifetime and higher energy efficiency of heat pumps (HPs) in residential buildings. To this end, we use smart meter data (SMD) with a resolution of 15 min to monitor 503 HPs installed in Swiss single-family homes over a period of 21 months. We show how heating cycles can be extracted from the corresponding time series of energy measurements. Then, we use the identified cycles to calculate the following daily key indicators for each HP: operating hours, number of cycles, ratio of cycles to operating hours, and average cycle length. When grouping the daily indicators of each HP by average daily outdoor temperatures and calculating the median for each temperature, it can be found that the behavior is nearly linear in the temperature range of 0–12 °C. Therefore, we calculate a linear regression for each HP and indicator. Using the derived slopes and intercepts as inputs to an outlier detection algorithm, the algorithm identifies conspicuous HPs that behave atypically and differently than the overall population. In addition to this approach, which does not require contextual information, we also evaluate the energy efficiency and appropriate size of each HP. However, this requires additional information on heated floor area and electrical power, which is typically unavailable and here, is only available for 243 of the 503 HPs (i.e., 48%). To evaluate the approach, we apply it to this subset and compare the HPs that cycle atypically to the best and worst 10% of the HPs in terms of energy intensity and utilization. The resulting Cohen's Kappas of 0.40 and 0.37, a common measure to assess inter-rater reliability, indicate a fair agreement between cycling, energy intensity, and utilization [80]. Therefore, the approach to assess cycling

is also suitable for identifying conspicuous HPs in situations where the context is unknown, e.g., due to privacy constraints or missing data. In addition to identifying conspicuous HPs, extensive descriptive analyses of typical cycling across all 503 HPs are provided considering building age, building size, HP type, and HP size. While this work is the first to monitor real-world HP cycling with SMD on a large scale, our analyses are limited to the geographic conditions and heating scenarios of Switzerland. Future work could focus on using HP sensor data with higher resolution and with more parameters to verify the results and investigate the reasons for differences in cycling behavior.

CRediT authorship contribution statement

Tobias Brudermueller: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Markus Kreft:** Writing – review & editing, Visualization. **Elgar Fleisch:** Writing – review & editing, Supervision, Project administration. **Thorsten Staake:** Writing – review & editing, Supervision, Project administration, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tobias Brudermueller reports financial support was provided by Swiss Federal Office of Energy.

Data availability

Data will be made available on request.

Acknowledgments

We want to thank Marina González Vayá and Angelos Selviaridis for their help on the data collection and curation, and the valuable feedback and input to improve this work. Further, Tobias Graml and Thomas Losinger have provided us with new ideas during fruitful discussions. The research project is funded by the Swiss Federal Office of Energy under the grant number SI/502257. Therefore, we also want to acknowledge the strong support of other people involved in the research project: Jan Marckhoff, Dominik Bilgeri, Lorenz Deppeler, Hardy Schröder, Michael Peter, Andreas Weigert, Stephan Renz, Carina Alles, Rita Kobler, and Roland Brüniger.

Appendix A

A.1. Algorithms for extracting heat pump cycles

In this section, detailed descriptions of the algorithms to derive the HP cycles from SMD measurements are provided. We start with a formal definition of SMD and afterwards explain each step in a separate subsection. The algorithms are applied independently to all households in the data set and the HP of each household is measured separately from all other appliances. All formulas in the following are from the authors of this paper.

The SMD is represented by a time series E that is a sequence of non-negative energy consumption values e_t in kWh. The sequence is ordered by the index t, which represents the timestep with a step size of 0.25 h (i.e., 15 min). Hence, with T observations, we can write:

$$E = \{e_t \in \mathbb{R}^{\geq 0} \mid 0 \leq t \leq T-1\} \tag{3}$$

A.1.1. Step 1: Estimating baseload

Even during periods of non-heating, HPs typically show slightly non-zero energy consumption in the SMD. The reason is that the control unit remains active and thus, creates a small baseload [29]. We interpret this baseload as an energy threshold beyond which the HP is considered to perform a cycle. As the HPs differ among households, we derive the baseload b for each household individually. We use a distribution-based approach that was found empirically:

1. Round all values e_t to two decimals.
2. Calculate the maximum $e_{\max}$ from the set of values $\{e_t\}$ that occur more than once.
3. Calculate the threshold as follows:

$$b = \begin{cases} 0.10 & \text{if } e_{\max} > 1.0 \\ 0.06 & \text{if } e_{\max} < 0.6 \\ 0.10 \times e_{\max} & \text{otherwise} \end{cases} \tag{4}$$

A.1.2. Step 2: Deriving activity states

From the original time series E, we create a binarized version A. While the timesteps t remain identical, the values are now binary to indicate activity states. Hence, it can be written as:

$$A = \{a_t \in \{0,1\} \mid 0 \leq t \leq T-1\} \tag{5}$$

To derive a_t, we compare the SMD values e_t against the baseload estimation b:

$$a_t = \begin{cases} 1 & \text{if } e_t > b \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

If a_t is equal to zero, the HP is off or in standby. If a_t is equal to one, it indicates that the HP in an on-state and performs a cycle.

A.1.3. Step 3: Determining on- and off-transients

Next, we use A to derive the timesteps at which the HP switches on and off (i.e., changes of state). We refer to these as on- and off-transients of each cycle and proceed as follows:

1. For each a_t that does not represent the first and last value of the time series ($1 \leq t \leq T-2$), we calculate the difference to the previous binary state:

$$\Delta a_t = a_t - a_{t-1} \tag{7}$$

2. The set of indices referring to on-transients is then defined as follows:

$$S_{\text{on}} = \{0\} \cup \{t \mid \Delta a_t = 1\} \tag{8}$$

   Note that this way, we always assume the start of the time series to represent an on-transient.
3. Correspondingly, we assume that the end of the time series is a switch-off activity and calculate the off-transients as follows:

$$S_{\text{off}} = \{t \mid \Delta a_{t+1} = -1\} \cup \{T-1\} \tag{9}$$

At this point, it is important to note that by using Δa_t for the on-transients and Δa_{t+1} for the off-transients, a timestep t can occur in both sets S_{on} and S_{off}. Hence, t can represent a switch-on and switch-off activity at the same time, which occurs when a heating cycle is shorter than the measurement interval of 15 min. Further, treating $t = 0$ as an on-transient and $t = T-1$ as an off-transient ensures that an on-transient is always followed by an off-transient. Hence, both sets S_{on} and S_{off} contain the same number of elements.

A.1.4. Step 4: Calculating durations of individual cycles

In the following, we treat S_{on} and S_{off} as ordered sets with ascending values, where each element receives an index i according to its position. Hence, with $I = |S_{\text{on}}| = |S_{\text{off}}|$ elements in each set, we can rewrite:

$$S_{\text{on}} = \{t_i^{\text{on}} \mid 0 \leq i < I\} \tag{10}$$

$$S_{\text{off}} = \{t_i^{\text{off}} \mid 0 \leq i < I\} \tag{11}$$

Now, we can interpret i as index referring to a single cycle and I as the total number of observed cycles. We define a single cycle to have a switch-on timestep t_i^{on}, switch-off timestep t_i^{off} and a cycle duration in hours d_i. Assuming that a cycle covers all t within the interval $[t_i^{\text{on}}, t_i^{\text{off}}]$, we can write the set of cycles S_{cycle} as follows:

$$S_{\text{cycle}} = \{\langle t_i^{\text{on}}, t_i^{\text{off}}, d_i\rangle \mid 0 \leq i < I \ \wedge \ t_i^{\text{on}} \in S_{\text{on}} \ \wedge \ t_i^{\text{off}} \in S_{\text{off}}\} \tag{12}$$

The easiest way to define d_i would be to count each t within the interval $[t_i^{\text{on}}, t_i^{\text{off}}]$ as full 15 min, formalized as:

$$d_i = (t_i^{\text{off}} - t_i^{\text{on}} + 1) \times 0.25 \tag{13}$$

However, as explained in Section 3.6, this would lead to overestimates because an HP can switch on or off at any time within a 15-minute interval. Therefore, we assume a constant electricity uptake between an on-transient and the subsequent measurement, or an off-transient and the previous measurement to calculate the exact durations at t_i^{on} and t_i^{off}. We count all other measurements of a cycle $t_i^{\text{on}} < t < t_i^{\text{off}}$ as full 15 min. Hence, instead of using Eq. (13), we define d_i to be *"more exact"* as:

$$d_i = \begin{cases} 0.125 & \text{if } t_i^{\text{on}} == t_i^{\text{off}} \\ d_i^{\text{on}} + d_i^{\text{off}} & \text{if } t_i^{\text{off}} == t_i^{\text{on}} + 1 \\ d_i^{\text{on}} + d_i^{\text{off}} + \sum_{t_i^{\text{on}}+1}^{t_i^{\text{off}}-1} 0.25 & \text{otherwise} \end{cases} \tag{14}$$

In this context, we define d_i^{on} and d_i^{off} as follows:

$$d_i^{on} = \min\left(\frac{e_{t_i^{\text{on}}}}{e_{t_i^{\text{on}}+1}} \times 0.25, 0.25\right) \tag{15}$$

$$d_i^{off} = \min\left(\frac{e_{t_i^{\text{off}}}}{e_{t_i^{\text{off}}-1}} \times 0.25, 0.25\right) \tag{16}$$

Note that the definitions above handle some special cases where calculating exact durations is difficult:

- If t_i^{on} equals t_i^{off}, a cycle is shorter than 15 min and the exact duration cannot be determined. Then, we count the cycle duration as half a 15 min interval, i.e. 0.125 h.
- If an on-transient is immediately followed by an off-transient, the cycle duration is between 15 and 30 min and cannot be well estimated. However, we treat this case as any other cycle with longer durations by considering the proportions of the two consecutive energy measurements.
- If the energy of the on-transient is higher than the energy of the consecutive measurement, or the energy of the off-transient is higher and the energy of the previous measurement, we cannot determine the exact duration of a switch-on or switch-off activity. In this case, we count it as full 15 min.

A.2. Calculating daily metrics as key indicators

In Section 3.7, we explain that we calculate daily metrics to match them with average outdoor temperatures. To this end, we apply step 1–4 of the previous procedure to each day of SMD individually, such that we gain a set of cycles S_{cycle} per day. Then, we can calculate the daily metrics as follows:

$$\begin{aligned} \text{Operating Hours} &= \sum d_i \\ \text{Cycles} &= |S_{\text{cycle}}| \\ \text{Cycles/Operating Hours} &= \frac{|S_{\text{cycle}}|}{\sum d_i} \\ \text{Average Cycle Length} &= \frac{\sum d_i}{|S_{\text{cycle}}|} \end{aligned} \tag{17}$$

References

[1] IEA. Heating. 2022, IEA, URL https://www.iea.org/reports/heating. [Accessed: 15 December 2022].

[2] Walker IS, Less BD, Casquero-Modrego N. Carbon and energy cost impacts of electrification of space heating with heat pumps in the US. Energy Build 2022;259:111910. http://dx.doi.org/10.1016/j.enbuild.2022.111910.

[3] Yin P, Pate M, Battaglia F. In-field performance evaluation and economic analysis of residential ground source heat pumps in heating operation. J Build Eng 2019;26:100932. http://dx.doi.org/10.1016/j.jobe.2019.100932.

[4] Jenkins D, Tucker R, Rawlings R. Modelling the carbon-saving performance of domestic ground-source heat pumps. Energy Build 2009;41(6):587–95. http://dx.doi.org/10.1016/j.enbuild.2008.12.002.

[5] Bayer P, Saner D, Bolay S, Rybach L, Blum P. Greenhouse gas emission savings of ground source heat pump systems in Europe: A review. Renew Sustain Energy Rev 2012;16(2):1256–67. http://dx.doi.org/10.1016/j.rser.2011.09.027.

[6] IEA. Installation of about 600 million heat pumps covering 20% of buildings heating needs required by 2030. 2022, IEA, URL https://www.iea.org/reports/installation-of-about-600-million-heat-pumps-covering-20-of-buildings-heating-needs-required-by-2030. [Accessed: 15 December 2022].

[7] Caird S, Roy R, Potter S. Domestic heat pumps in the UK: user behaviour, satisfaction and performance. Energy Efficiency 2012;5(3):283–301. http://dx.doi.org/10.1007/s12053-012-9146-x.

[8] Qiao Z, Long T, Li W, Zeng L, Li Y, Lu J, et al. Performance assessment of ground-source heat pumps (GSHPs) in the southwestern and northwestern China: In situ measurement. Renew Energy 2020;153:214–27. http://dx.doi.org/10.1016/j.renene.2020.02.024.

[9] Willem H, Lin Y, Lekov A. Review of energy efficiency and system performance of residential heat pump water heaters. Energy Build 2017;143:191–201. http://dx.doi.org/10.1016/j.enbuild.2017.02.023.

[10] Weigert A, Hopf K, Günther SA, 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 2022;169:113156. http://dx.doi.org/10.1016/j.enpol.2022.113156.

[11] Vocale P, Morini GL, Spiga M. Influence of outdoor air conditions on the air source heat pumps performance. Energy Procedia 2014;45:653–62. http://dx.doi.org/10.1016/j.egypro.2014.01.070.

[12] Dongellini M, Naldi C, Morini GL. Seasonal performance evaluation of electric air-to-water heat pump systems. Appl Therm Eng 2015;90:1072–81. http://dx.doi.org/10.1016/j.applthermaleng.2015.03.026.

[13] Morrison G, Anderson T, Behnia M. Seasonal performance rating of heat pump water heaters. Sol Energy 2004;76(1–3):147–52. http://dx.doi.org/10.1016/j.solener.2003.08.007.

[14] Bagarella G, Lazzarin R, Noro M. Sizing strategy of on–off and modulating heat pump systems based on annual energy analysis. Int J Refrig 2016;65:183–93. http://dx.doi.org/10.1016/j.ijrefrig.2016.02.015.

[15] Dongellini M, Naldi C, Morini GL. Influence of sizing strategy and control rules on the energy saving potential of heat pump hybrid systems in a residential building. Energy Convers Manage 2021;235:114022. http://dx.doi.org/10.1016/j.enconman.2021.114022.

[16] 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 2021;39:102254. http://dx.doi.org/10.1016/j.jobe.2021.102254.

[17] Madani H, Roccatello E. A comprehensive study on the important faults in heat pump system during the warranty period. Int J Refrig 2014;48:19–25. http://dx.doi.org/10.1016/j.ijrefrig.2014.08.007.

[18] 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. 2014, p. 50–9. http://dx.doi.org/10.1145/2674061.2674067.

[19] Tejeda A, Milu A, Riviere P, Marchio D. Energy consequences of non-optimal heat pump parameterization. 2014.

[20] Doseva N, Chakyrova D. Life cycle cost analysis of different residential heat pump systems. In: E3S web of conferences, vol. 207. EDP Sciences; 2020, p. 01014. http://dx.doi.org/10.1051/e3sconf/202020701014.

[21] Li G. Comprehensive investigations of life cycle climate performance of packaged air source heat pumps for residential application. Renew Sustain Energy Rev 2015;43:702–10. http://dx.doi.org/10.1016/j.rser.2014.11.078.

[22] Barnes J, Bhagavathy SM. The economics of heat pumps and the (un) intended consequences of government policy. Energy Policy 2020;138:111198. http://dx.doi.org/10.1016/j.enpol.2019.111198.

[23] Baeten B, Rogiers F, Helsen L. Reduction of heat pump induced peak electricity use and required generation capacity through thermal energy storage and demand response. Appl Energy 2017;195:184–95. http://dx.doi.org/10.1016/j.apenergy.2017.03.055.

[24] Nykamp S, Molderink A, Bakker V, Toersche HA, Hurink JL, Smit GJ. Integration of heat pumps in distribution grids: Economic motivation for grid control. In: 2012 3rd IEEE PES innovative smart grid technologies Europe. IEEE; 2012, p. 1–8. http://dx.doi.org/10.1109/ISGTEurope.2012.6465605.

[25] Watson S, Crawley J, Lomas K, Buswell R. Predicting future GB heat pump electricity demand. Energy Build 2023;286:112917. http://dx.doi.org/10.1016/j.enbuild.2023.112917.

[26] Love J, Smith AZ, Watson S, Oikonomou E, Summerfield A, Gleeson C, et al. The addition of heat pump electricity load profiles to GB electricity demand: Evidence from a heat pump field trial. Appl Energy 2017;204:332–42. http://dx.doi.org/10.1016/j.apenergy.2017.07.026.

[27] Weigert A. Identification and classification of heat pump problems in the field and their implication for a user-centric problem recognition. Energy Inform 2022;5(1):70. http://dx.doi.org/10.1186/s42162-022-00250-3.

[28] Scotton S, Nowak T. A digital revolution for the built environment? Flexibility, comfort, and new business models for heat pumps. European Heat Pump Association (EHPA); 2021, URL https://www.ehpa.org/fileadmin/user_upload/EHPA-Digitalisation_White_Paper-06-compressed.pdf. [Accessed: 04 May 2022].

[29] 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. 2022, p. 412–8. http://dx.doi.org/10.1109/SmartGridComm52983.2022.9961055.

[30] Adams JN, Bélafi ZD, Horváth M, Kocsis JB, Csoknyai T. How smart meter data analysis can support understanding the impact of occupant behavior on building energy performance: A comprehensive review. Energies 2021;14(9):2502. http://dx.doi.org/10.3390/en14092502.

[31] Efkarpidis N, Geidl M, Wache H, Peter M, Adam M. Smart metering applications: main concepts and business models, vol. 88. Springer Nature; 2022, http://dx.doi.org/10.1007/978-3-031-05737-3.

[32] Uhlmann M, Bertsch S. Measurement and simulation of startup and shut down of heat pumps. 2010.

[33] Dongellini M, Morini GL. On-off cycling losses of reversible air-to-water heat pump systems as a function of the unit power modulation capacity. Energy Convers Manage 2019;196:966–78. http://dx.doi.org/10.1016/j.enconman.2019.06.022.

[34] Green R. The effects of cycling on heat pump performance. Confidential Report, EA Technology, EA Technology, Chester, UK, Project 46640, 2012, URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/65695/7389-effects-cycling-heat-pump-performance.pdf. [Accessed: 25 January 2023].

[35] Song M, Mao N, Xu Y, Deng S. Challenges in, and the development of, building energy saving techniques, illustrated with the example of an air source heat pump. Therm Sci Eng Prog 2019;10:337–56. http://dx.doi.org/10.1016/j.tsep.2019.03.002.

[36] Huchtemann K, Müller D. Simulation study on supply temperature optimization in domestic heat pump systems. Build Environ 2013;59:327–35. http://dx.doi.org/10.1016/j.buildenv.2012.08.030.

[37] Corberán JM, Donadello D, Martínez-Galván I, Montagud C. Partialization losses of ON/OFF operation of water-to-water refrigeration/heat-pump units. Int J Refrig 2013;36(8):2251–61. http://dx.doi.org/10.1016/j.ijrefrig.2013.07.002.

[38] Waddicor DA, Fuentes E, Azar M, Salom J. Partial load efficiency degradation of a water-to-water heat pump under fixed set-point control. Appl Therm Eng 2016;106:275–85. http://dx.doi.org/10.1016/j.applthermaleng.2016.05.193.

[39] Chen X-D, Li L, Tseng M-L, Tan K, Ali MH. Improving power quality efficient in demand response: Aggregated heating, ventilation and air-conditioning systems. J Clean Prod 2020;267:122178. http://dx.doi.org/10.1016/j.jclepro.2020.122178.

[40] Xu Z, Sun X, Li X, Wang Z, Xu W, Shao S, et al. On-off cycling model featured with pattern recognition of air-to-water heat pumps. Appl Therm Eng 2021;196:117317. http://dx.doi.org/10.1016/j.applthermaleng.2021.117317.

[41] Xu Z, Li H, Shao S, Xu W, Wang Z, Wang Y, et al. A semi-theoretical model for energy efficiency assessment of air source heat pump systems. Energy Convers Manage 2021;228:113667.

[42] 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 2022;14(24):16799. http://dx.doi.org/10.3390/su142416799.

[43] Staffell I, Brett D, Brandon N, Hawkes A. A review of domestic heat pumps. Energy Environ Sci 2012;5(11):9291–306. http://dx.doi.org/10.1039/C2EE22653G.

[44] Dongellini M, Abbenante M, Morini GL. A strategy for the optimal control logic of heat pump systems: impact on the energy consumptions of a residential building. In: Proceedings of the 12th IEA heat pump conference, vol. 2017. 2017.

[45] Weigert A, Hopf K, Weinig N, Staake T. Detection of heat pumps from smart meter and open data. Energy Inform 2020;3(1):1–14. http://dx.doi.org/10.1186/s42162-020-00124-6.

[46] Fei H, Kim Y, Sahu S, Naphade M, Mamidipalli SK, Hutchinson J. Heat pump detection from coarse grained smart meter data with positive and unlabeled learning. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. 2013, p. 1330–8. http://dx.doi.org/10.1145/2487575.2488203.

[47] Faustine A, Mvungi NH, Kaijage S, Michael K. A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem. 2017, arXiv preprint arXiv:1703.00785.

[48] Zhuang M, Shahidehpour M, Li Z. An overview of non-intrusive load monitoring: Approaches, business applications, and challenges. In: 2018 international conference on power system technology. IEEE; 2018, p. 4291–9. http://dx.doi.org/10.1109/POWERCON.2018.8601534.

[49] Klemenjak C, Goldsborough P. Non-intrusive load monitoring: A review and outlook. 2016, arXiv preprint arXiv:1610.01191.

[50] Rafati A, Shaker HR, Ghahghahzadeh S. Fault detection and efficiency assessment for HVAC systems using non-intrusive load monitoring: A review. Energies 2022;15(1):341. http://dx.doi.org/10.3390/en15010341.

[51] Harell A, Makonin S, Bajić IV. Wavenilm: A causal neural network for power disaggregation from the complex power signal. In: ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing. IEEE; 2019, p. 8335–9. http://dx.doi.org/10.1109/ICASSP.2019.8682543.

[52] Lee ZE, Zhang KM. Scalable identification and control of residential heat pumps: A minimal hardware approach. Appl Energy 2021;286:116544. http://dx.doi.org/10.1016/j.apenergy.2021.116544.

[53] Wang H, Yang W. An iterative load disaggregation approach based on appliance consumption pattern. Appl Sci 2018;8(4):542. http://dx.doi.org/10.3390/app8040542.

[54] Yuan Q, Wang H, Wu B, Song Y, Wang H. A fusion load disaggregation method based on clustering algorithm and support vector regression optimization for low sampling data. Future Internet 2019;11(2):51. http://dx.doi.org/10.3390/fi11020051.

[55] Zeinal-Kheiri S, Shotorbani AM, Mohammadi-Ivatloo B. Residential load disaggregation considering state transitions. IEEE Trans Ind Inf 2019;16(2):743–53. http://dx.doi.org/10.1109/TII.2019.2925323.

[56] Deb C, Frei M, Hofer J, Schlueter A. Automated load disaggregation for residences with electrical resistance heating. Energy Build 2019;182:61–74. http://dx.doi.org/10.1016/j.enbuild.2018.10.011.

[57] Hedegaard RE, Kristensen MH, Petersen S. Experimental validation of a model-based method for separating the space heating and domestic hot water components from smart-meter consumption data. In: E3S web of conferences, vol. 172. EDP Sciences; 2020, p. 12001. http://dx.doi.org/10.1051/e3sconf/202017212001.

[58] Liang H, Ma J. Separation of residential space cooling usage from smart meter data. IEEE Trans Smart Grid 2020;11(4):3107–18. http://dx.doi.org/10.1109/TSG.2020.2965958.

[59] Green C, Garimella S. Non-intrusive load monitoring of water heaters using low-resolution data. In: Proceedings of the 5th international workshop on non-intrusive load monitoring. 2020, p. 54–8. http://dx.doi.org/10.1145/3427771.3427853.

[60] Basu K, Debusschere V, Bacha S, Maulik U, Bondyopadhyay S. Nonintrusive load monitoring: A temporal multilabel classification approach. IEEE Trans Ind Inf 2015;11(1):262–70. http://dx.doi.org/10.1109/TII.2014.2361288.

[61] Zufferey T, Valverde G, Hug G. Unsupervised disaggregation of water heater load from smart meter data processing. In: The 12th mediterranean conference on power generation, transmission, distribution and energy conversion, vol. 2020. IET; 2020, p. 283–8. http://dx.doi.org/10.1049/icp.2021.1269.

[62] Müller FL, Jansen B. Large-scale demonstration of precise demand response provided by residential heat pumps. Appl Energy 2019;239:836–45. http://dx.doi.org/10.1016/j.apenergy.2019.01.202.

[63] Kouzelis K, Tan ZH, Bak-Jensen B, Pillai JR, Ritchie E. Estimation of residential heat pump consumption for flexibility market applications. IEEE Trans Smart Grid 2015;6(4):1852–64. http://dx.doi.org/10.1109/TSG.2015.2414490.

[64] Péan TQ, Salom J, Costa-Castelló R. Review of control strategies for improving the energy flexibility provided by heat pump systems in buildings. J Process Control 2019;74:35–49. http://dx.doi.org/10.1016/j.jprocont.2018.03.006.

[65] Schibuola L, Scarpa M, Tambani C. Demand response management by means of heat pumps controlled via real time pricing. Energy Build 2015;90:15–28. http://dx.doi.org/10.1016/j.enbuild.2014.12.047.

[66] Fischer D, Wolf T, Wapler J, Hollinger R, Madani H. Model-based flexibility assessment of a residential heat pump pool. Energy 2017;118:853–64. http://dx.doi.org/10.1016/j.energy.2016.10.111.

[67] Patteeuw D, Henze GP, Helsen L. Comparison of load shifting incentives for low-energy buildings with heat pumps to attain grid flexibility benefits. Appl Energy 2016;167:80–92. http://dx.doi.org/10.1016/j.apenergy.2016.01.036.

[68] Efkarpidis NA, Vomva SA, Christoforidis GC, Papagiannis GK. Optimal day-to-day scheduling of multiple energy assets in residential buildings equipped with variable-speed heat pumps. Appl Energy 2022;312:118702. http://dx.doi.org/10.1016/j.apenergy.2022.118702.

[69] Klaassen E, Asare-Bediako B, De Koning C, Frunt J, Slootweg J. Assessment of an algorithm to utilize heat pump flexibility-theory and practice. In: 2015 IEEE Eindhoven powertech. IEEE; 2015, p. 1–6. http://dx.doi.org/10.1109/PTC.2015.7232783.

[70] Nweye K, Nagy Z. MARTINI: Smart meter driven estimation of HVAC schedules and energy savings based on Wi-Fi sensing and clustering. Appl Energy 2022;316:118980. http://dx.doi.org/10.1016/j.apenergy.2022.118980.

[71] Merkmalskatalog, no. 22905270. Bundesamt für Statistik (BFS): Bundesamt für Statistik (BFS); 2022, URL https://dam-api.bfs.admin.ch/hub/api/dam/assets/22905270/master, Backup Publisher.

[72] Kulawik M, Bucher P. Gebäude-heizenergiebedarf: Methodik zur schätzung des heizenergiebedarfs der wohngebäude mittels kantonalem gebäude- und wohnungsregister. 2013, Bau-, Umwelt- und Wirtschaftsdepartement Umwelt und Energie (uwe), Kanton Luzern, URL https://uwe.lu.ch/-/media/UWE/Dokumente/publikationen/Publikationen_02_G_bis_L/Gebaeude_Heizenergiebedarf_GWR.pdf.

[73] Amt für Raumentwicklung, Abteilung Geoinformation, Fachstelle GIS, GIS-Browser. Geografisches informationssystem des kantons Zürich (GIS-ZH). 2022, https://www.im.ethz.ch and https://www.bitstoenergy.com. [Accessed: 06 January 2022].

[74] SIA Standard. 381/3: Heating degree-days in Switzerland. 1982, Swiss Association of Engineers and Architects (in German), Zurich, Switzerland.

[75] Wang E, Shen Z, Alp N, Barry N. Benchmarking energy performance of residential buildings using two-stage multifactor data envelopment analysis with degree-day based simple-normalization approach. Energy Convers Manage 2015;106:530–42. http://dx.doi.org/10.1016/j.enconman.2015.09.072.

[76] Hopf K, Weigert A, Staake T. Value creation from analytics with limited data: A case study on the retailing of durable consumer goods. J Decis Syst 2022;1–37. http://dx.doi.org/10.1080/12460125.2022.2059172.

[77] Asghar MR, Dán G, Miorandi D, Chlamtac I. Smart meter data privacy: A survey. IEEE Commun Surv Tutor 2017;19(4):2820–35. http://dx.doi.org/10.1109/COMST.2017.2720195.

[78] Breunig MM, Kriegel H-P, Ng RT, Sander J. LOF: Identifying density-based local outliers. SIGMOD Rec 2000;29(2):93–104. http://dx.doi.org/10.1145/335191.335388.

[79] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[80] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74. http://dx.doi.org/10.2307/2529310, URL http://www.jstor.org/stable/2529310.