Brudermueller, Tobias; Fleisch, Elgar; González Vayá, Marina; u. a.

HEAPO : An Open Dataset for Heat Pump Optimization with Smart Electricity Meter Data and On-Site Inspection Protocols

Date of secondary publication: 19.08.2026

Version of Record (Published Version), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-116869x

Primary publication

Brudermueller, Tobias; Fleisch, Elgar; González Vayá, Marina; u. a. (2025): HEAPO: An Open Dataset for Heat Pump Optimization with Smart Electricity Meter Data and On-Site Inspection Protocols, in: Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems, New York: Association for Computing Machinery, pp. 699–711, doi: 10.1145/3679240.3734637

HEAPO – An Open Dataset for Heat Pump Optimization with Smart Electricity Meter Data and On-Site Inspection Protocols

Tobias Brudermueller
ETH Zurich
Chair of Information Management
Zurich, Switzerland
tbrudermuell@ethz.ch

Elgar Fleisch*
ETH Zurich
Chair of Information Management
Zurich, Switzerland
University of St. Gallen
Institute of Technology Management
St. Gallen, Switzerland
efleisch@ethz.ch

Marina González Vayá
Elektrizitätswerke des Kantons Zürich
New Technology Department
Zurich, Switzerland
marina.gonzalezvaya@ekz.ch

Thorsten Staake†
University of Bamberg
Chair of Information Systems and Energy Efficient Systems
Bamberg, Germany
ETH Zurich
Chair of Information Management
Zurich, Switzerland
thorsten.staake@uni-bamberg.de

Abstract

Heat pumps are essential for decarbonizing residential heating but consume substantial electrical energy, impacting operational costs and grid demand. Many systems run inefficiently due to planning flaws, operational faults, or misconfigurations. While optimizing performance requires skilled professionals, labor shortages hinder large-scale interventions. However, digital tools and improved data availability create new service opportunities for energy efficiency, predictive maintenance, and demand-side management. To support research and practical solutions, we present an open-source dataset of electricity consumption from 1,408 households with heat pumps and smart electricity meters in the canton of Zurich, Switzerland, recorded at 15-minute and daily resolutions between 2018-11-03 and 2024-03-21. The dataset includes household metadata, weather data from 8 stations, and ground truth data from 410 field visit protocols collected by energy consultants during system optimizations. Additionally, the dataset includes a Python-based data loader to facilitate seamless data processing and exploration.

CCS Concepts

• **Hardware → Smart grid**; • **Software and its engineering →** *Software libraries and repositories.*

*Also with Institute of Technology Management, University of St. Gallen.
†Also with Chair of Information Management, ETH Zurich.

E-ENERGY '25, Rotterdam, Netherlands

ACM ISBN 979-8-4007-1125-1/25/06
https://doi.org/10.1145/3679240.3734637

Keywords

heat pump, optimization, smart electricity meter, open energy data

ACM Reference Format:
Tobias Brudermueller, Elgar Fleisch, Marina González Vayá, and Thorsten Staake. 2025. HEAPO – An Open Dataset for Heat Pump Optimization with Smart Electricity Meter Data and On-Site Inspection Protocols. In *The 16th ACM International Conference on Future and Sustainable Energy Systems (E-ENERGY '25), June 17–20, 2025, Rotterdam, Netherlands.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3679240.3734637

1 Introduction

Electric heat pumps (HPs) are a key solution for decarbonizing residential space and water heating. However, their complexity compared to traditional gas or oil-based systems makes real-world efficiency heavily reliant on proper planning [16, 44], optimal settings [50], and fault-free operation [24, 34], which many users and installers find difficult to manage [12, 15]. Consequently, real-world performance varies widely, and many HPs have substantial room for optimization [8]. For example, Nolting et al. [41] found real-world efficiency to be up to 24% lower than rated values. Likewise, a study of 297 Swiss households showed that after optimization, half reduced annual electricity use by 1,805 kWh (15.2%) on average [52]. Unexploited optimization potential from HPs can strain electricity grids [32, 33, 45, 48] and drive up homeowners' costs, undermining economic viability and hindering the transition to sustainable heating [2, 14, 17, 40]. In this context, digital tools and data-driven solutions provide valuable opportunities for predictive maintenance, enhanced energy efficiency, and demand-side flexibility to address these challenges. Studies show that HPs can even enhance grid stability, with operational flexibility supporting frequency regulation and reserve capacity in demand response programs [3, 30, 31]. For example, throttling 300+ HPs reduced load by 40-65% [38], and flexible HPs in Germany lowered homeowners' electricity costs by 4-9% in demand response scenarios [21].

Digital solutions for HP optimization are even more important in the light of a shortage of skilled workers. The European Union estimates a need for 750,000 additional installers, with at least 50% of the current workforce requiring reskilling [18]. In the UK, where HPs can only be installed by certified professionals, only 1,800 installers were certified in 2020, far below the 69,500 needed by 2035 [5]. This labor shortage affects both the deployment of new HPs and the optimization of existing systems. To address this, digital tools are essential for assessing system performance, identifying households with optimization potential, and prioritizing them for on-site interventions by professionals. While many modern HPs already transmit high-quality sensor data, including pressure, volume flow, and energy input/output, a significant number of existing units lack internet connectivity and advanced measurement capabilities [19, 25]. Yet, they will remain in operation for decades. In such cases, advanced metering infrastructure (AMI) with smart electricity meters (SMs) provides a scalable alternative for remote HP monitoring [6, 9, 38]. For example, a study using real-world smart meter data (SMD) from 503 Swiss households has already demonstrated that inefficient HPs can be identified by analyzing their on-off cycling behavior [7].

To drive research advancements and scale real-world services, which still remain rarely deployed, high-quality data must be freely available to capture the variability of real-world system performance [22, 23, 36]. Understanding the factors that influence HP behavior and performance, as well as training algorithms to capture diverse usage patterns, requires a dataset that combines operational data with contextual information about the building and heating system, along with reliable ground-truth data on HP settings and optimization potential. There are several open-source datasets related to residential electricity consumption and HP usage, but none provide all the functionalities required for developing algorithms tailored to the aforementioned purposes [1, 28, 42, 43, 46, 47]. For instance, the dataset by Schlemminger et al. [47] offers residential electricity and HP load profiles from single-family homes in Northern Germany, recorded at resolutions ranging from 10 seconds to 60 minutes between 2018 and 2020. However, with data from only 38 households, the sample size is small, and the dataset includes very limited metadata on household characteristics, with no details on HP settings. Another well-known dataset in the HP domain is *When2Heat* by Ruhnau et al. [46], which provides synthetic national time series for heat demand and HP coefficient of performance across 16 European countries rather than real-world measurements. Other commonly used SMD datasets, often applied in non-intrusive load monitoring research, include *REDD* [29], *UK-DALE* [27], *GREEND* [37], and *AMPds* [35]. Among these, only *AMPds* contains HP measurements, providing one-minute power readings for a single house in Canada over two years. In summary, there is currently no publicly and freely available dataset with a large sample size that integrates SMD, metadata, and ground-truth data from residential buildings using HPs for heating or cooling.

To address these challenges, this work presents *HEAPO (Heat Pump Optimization)* – a real-world, open-source dataset designed to support researchers, practitioners, utilities, and other stakeholders in advancing both research and practical solutions for optimizing HP operations. For instance, *HEAPO* can be utilized to develop advanced algorithms, applications, and services for improving energy efficiency, disaggregating loads, leveraging flexibility in demand response programs, and estimating residential photovoltaic (PV) feed-in to the grid. The dataset provides cleaned and ready-to-use energy data at both 15-minute and daily resolutions, recorded by standard smart electricity meters in 1,408 real households that use HPs for space and water heating. These households are located in the canton of Zurich, Switzerland. The field data was collected between 2018-11-03 and 2024-03-21, with each household having an average of 721.13 days of available data in daily resolution and 638.75 days in 15-minute resolution. Additionally, the dataset includes temperature data from the 8 nearest weather stations matched to household locations, metadata provided by households through an online tool from their utility company, and 410 field visit protocols recorded by professional energy consultants. These consultants, with over a decade of experience conducting energy audits, optimized the HPs during their on-site visits. For 214 of the households with available SMD, a corresponding audit protocol is available. Therefore, the protocols provide valuable insights into common issues observed in real-world installations and enable the analysis of changes in electricity consumption following the optimization, offering a unique opportunity to evaluate the impact of targeted interventions. To facilitate seamless use of the dataset, a Python-based dataloader including visualization functionalities is provided, enabling efficient data processing and software development without delay. Both the dataset and the accompanying code are publicly available on GitHub: https://github.com/tbrumue/heapo

2 Dataset Description

This section outlines data collection and pre-processing, dataset structure, dataloader functionality, and technical validation.

2.1 Data Collection and Pre-Processing

The dataset integrates information from multiple sources into a user-friendly structure, assigning each household a unique identifier (*Household_ID*) for seamless cross-referencing across all data instances. To ensure compliance with data privacy regulations, identifiers are randomly assigned integer values, and any information that could reveal a household's exact location or private details has been removed. Details on data collection and pre-processing steps for each source are provided in the following subsections, with an overview illustrated in Figure 1.

2.1.1 Protocol Data. The dataset includes 410 reports from on-site energy audits conducted between 2015 and 2024 by experienced energy consultants specializing in HP optimization. These audits are part of a paid service provided by a local energy consulting company within the utility and can take place on any day of the year. During visits, consultants assess the HP and heat distribution system, addressing minor issues that do not require an installer. If basic settings such as the heating curve, time programs, or domestic hot water temperature are found to be suboptimal, they may adjust them with customer consent. However, consultants do not perform retrofits, though they may recommend them. Since customers can adjust settings before or after the visit, the configuration details reflect the system's status at the time of inspection, while metadata about the building typically remains stable. We emphasize that

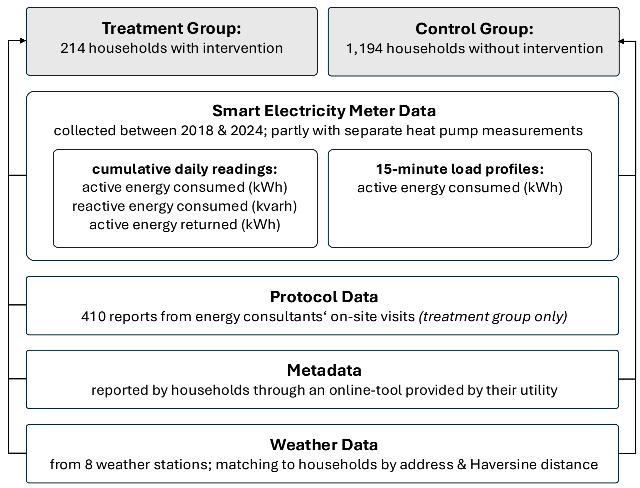

Figure 1: Overview of available data sources.

the provided reports make this dataset unique, offering valuable insights into common issues observed in field-installed HPs and providing high-quality ground-truth data to develop algorithms that can anticipate these issues and assess HP performance in operation. Each protocol is assigned a unique identifier, referred to as *Report_ID*. For 214 households, both protocol data and SMD are available, forming the treatment group, while the remaining 1,194 households with only SMD serve as the control group. Note that the protocol data also includes 196 reports from households without associated SMD, which are not part of either the treatment or control group. For each protocol associated with a treatment group household, a mapping between the report identifier and the household identifier is provided. Within the treatment group, 3 households booked the HP service multiple times, resulting in reports from different periods. Additionally, 10 households have SMD only before the visit, 53 only after, and 151 for both before and after the visit. In particular, the latter enable a comparison of consumption patterns before and after intervention, allowing for the analysis of associated treatment effects. A detailed list of variables included in the protocols can be found in Table 5 in the Appendix. The reports are provided in a machine-readable format, using categorical, binary, or numerical values. To preserve household privacy, any commentary data in German has been excluded.

2.1.2 Smart Meter Data. Each SM in the study's households transmits recorded measurements to the utility company on a daily basis. Consequently, missing or corrupted data is limited to full daily records and may appear as continuous periods of zero values or negative readings for grid consumption. Such erroneous measurements have been excluded to ensure a cleaned dataset. All timestamps in the dataset files are reported in coordinated universal time (UTC). To facilitate data filtering, each SMD file includes a *Group* column, indicating whether a household belongs to the treatment or control group, and an *AffectsTimePoint* column, specifying whether each timestamp falls before, during, or after an energy consultant's visit for treatment group households, or is marked as unknown for control group households. The SMs provide cumulative daily meter readings since installation, ensuring high data quality for billing by reconciling any missing interval data in subsequent readings. These cumulative daily readings capture a household's interaction with the power grid, including both energy drawn from the grid and self-generated energy fed back into it in homes with a PV system. However, they do not capture self-consumed PV energy that does not reach the grid. For the cumulative daily values of received energy, the meters record both active consumption (in kWh) and reactive consumption (in kvarh), distinguishing between inductive (first quadrant) and capacitive (fourth quadrant) components. The dataset includes the original cumulative readings along with pre-processed files for daily and monthly consumption. Daily consumption is calculated by subtracting the previous day's counter value, while monthly consumption is derived from the difference between the last recorded values of each month, where available. In addition to the daily cumulative readings for billing purposes, the SMs also provide 15-minute load profiles of active energy consumption (in kWh), which are calculated by multiplying the average active power (in kW) over each 15-minute interval by the time duration. This calculation is performed directly on the device by each SM before transmission. While SMs are technically capable of providing higher-resolution measurements, practical constraints related to data transmission, storage, and processing typically limit the available data in real-world applications to a resolution between 15 minutes and one hour [54]. Note that energy returned to the grid and reactive energy received from the grid are only available in the daily cumulative readings, not in the 15-minute load profiles. Additionally, some households have two SMs, one exclusively measuring HP electricity consumption and the other covering all other appliances. In such cases, data from both meters are included separately, with their sum at each timestamp representing total consumption. For households without separate HP measurements, established non-intrusive load monitoring techniques can be applied to disaggregate the HP load from the overall household consumption [6]. Table 1 summarizes the pre-processed SM measurements provided in the dataset files, along with the corresponding OBIS codes [11, 26], the standard used by utilities in German-speaking countries to categorize different types of measurements (see Additional Explanations of OBIS Codes).

2.1.3 Weather Data. Heating and cooling activity often correlates with outdoor temperature conditions [53]. To account for this temperature dependence in SMD analysis, the dataset includes weather data from 8 nearby weather stations. Using each household's address, we matched it to the nearest weather station based on haversine distance and provided a mapping between household identifiers and their corresponding weather station identifiers. The weather data is sourced from MeteoSwiss [10], a highly reputable nationwide weather service, where data is available for download upon registration. It includes daily and hourly measurements of key meteorological variables, such as maximum, minimum, and average temperatures, humidity, precipitation, sunshine duration, wind speed, pressure, and dew point temperature. Table 6 in the Appendix presents a comprehensive description of the available weather data variables. Note that MeteoSwiss derives heating degree day (HDD) and cooling degree day (CDD) numbers from the average daily temperature (T_{avg}). The estimated heating energy demand for a building at a given location is assumed to be directly proportional to its HDD value. MeteoSwiss follows standards

Table 1: Description of smart meter data variables. Active electrical energy received from the grid (in kWh) is available in both 15-minute and daily resolution, while active electrical energy returned to the grid (in kWh) and reactive electrical energy received (in kvarh) are only available in daily resolution.

SMD Variable Name	Variable Description	OBIS Code
Household_ID	unique household identifier	
AffectsTimePoint	timestamp category relative to energy consultants' visit date (before, after, during, or unknown)	
Timestamp	UTC timestamp of data measurement (*YYYY-MM-DD HH:MM:SS*)	
kWh_received_Total	household's total active energy consumed	1-1:1.8.0 / 1-1:1.29.0
kWh_received_HeatPump	when separate SM: active energy consumed by HP	1-1:1.8.0 / 1-1:1.29.0
kWh_received_Other	when separate SM: active energy consumed by appliances other than HP	1-1:1.8.0 / 1-1:1.29.0
kWh_returned_Total	household's total active energy returned (if PV system available)	1-1:2.8.0
kvarh_received_capacitive_Total	household's total reactive capacitive energy consumed	1-1:8.8.0
kvarh_received_capacitive_HeatPump	when separate SM: reactive capacitive energy consumed by HP	1-1:8.8.0
kvarh_received_capacitive_Other	when separate SM: reactive capacitive energy consumed by appliances other than HP	1-1:8.8.0
kvarh_received_inductive_Total	household's total reactive inductive energy consumed	1-1:5.8.0
kvarh_received_inductive_HeatPump	when separate SM: reactive inductive energy consumed by HP	1-1:5.8.0
kvarh_received_inductive_Other	when separate SM: reactive inductive energy consumed by appliances other than HP	1-1:5.8.0

defined for both the United States and Switzerland (SIA 381/3 [49]), as given by the formulas below.

$$\text{HDD (Switzerland)} = \begin{cases} 20\ ^\circ\text{C} - T_{\text{avg}} & \text{if } T_{\text{avg}} > 12 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$\text{HDD (United States)} = \max(0, 18.3\ ^\circ\text{C} - T_{\text{avg}}) \tag{2}$$

$$\text{CDD (United States)} = \max(0, T_{\text{avg}} - 18.3\ ^\circ\text{C}) \tag{3}$$

2.1.4 Metadata. All households can access their electricity consumption data from SMs free of charge via an online customer portal provided by their utility company. Within this portal, they can voluntarily enter metadata to receive personalized energy efficiency recommendations. Examples of collected metadata variables include the number of residents and the type of heat distribution system. A complete list of available variables is provided in Table 4 in the Appendix. The metadata is available for all households, regardless of group assignment. However, it is included in the *HEAPO* dataset to offer contextual insights that enhance advanced SMD analytics, particularly for households in the control group where protocol data is not available.

2.2 Dataset Structure

The dataset is organized in an intuitive folder and file structure (see illustration in Figure 5 in the Appendix) to minimize barriers to data exploration. This structure allows users to quickly grasp the context without needing to reference this paper in detail, making it especially useful for those working in programming languages other than Python, for which we provide code for data loading and exploration (see the following section). The data is divided into four main subfolders, corresponding to different sources: *meta_data*, *reports*, *smart_meter_data*, and *weather_data*. Each subfolder is accompanied by descriptive files that clarify data availability and explain the meaning of each column or variable. Metadata and protocol data are provided in single CSV files for ease of access. In contrast, weather data is split into separate CSV files per weather station, organized into subfolders based on temporal resolution (daily and hourly). Similarly, each household's SMD is stored in a separate CSV file, grouped into subfolders by resolution (15-minute, cumulative counters, daily, and monthly).

2.3 Python-Based Dataloader

The dataset is accompanied by a Python script that provides a data loading class with built-in functionalities for intuitive data exploration, visualization, and filtering. For example, a single function call can load SMD and weather data into a unified pandas DataFrame, where weather data is automatically interpolated from hourly to 15-minute resolution if the SMD has the latter resolution. Additionally, time series data for treatment households can be easily segmented into periods before, during, and after an energy consultant's on-site visit or filtered to include only a specific number of months before and after consultation. Another function enables adjusting the temporal resolution of the data, allowing users to aggregate SMD into hourly or weekly intervals, even if these were not originally measured. Figure 2 showcases a heatmap visualization of a treatment household's SMD before and after consultation, generated using our script. The x-axis represents dates, the y-axis represents time of day, and energy consumption (in kWh) is color-encoded as pixels. With the provided Python script, generating such a figure requires only a few lines of code, as demonstrated in Figure 3.

2.4 Technical Validation

For the technical validation of the SMD, we assess the consistency of total active energy consumption measurements (kWh_received_Total) between daily data and 15-minute load profiles. First, we aggregate the 15-minute load profiles to a daily resolution by summing the values for each household and date. Next, we match the measurements by household and date, retaining only observations where at least 10 days of overlapping data exist between the summed 15-minute data and the daily data. This results in 437,592 days of observations from 866 households included in the evaluation. To account for variations in data availability and potential noise levels among different SMs, we compute error levels individually for each household, reporting the mean and standard deviation across all households. The errors

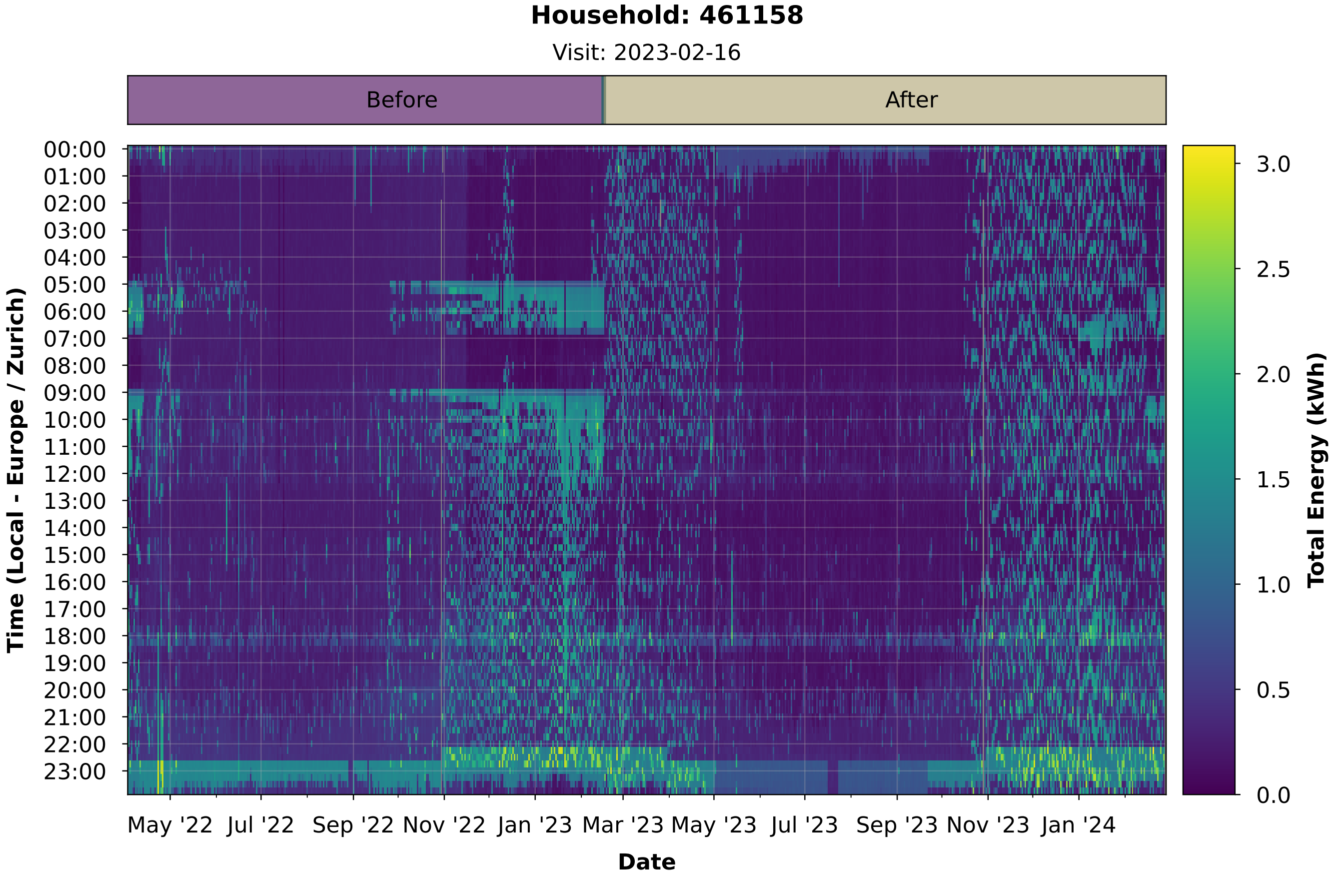

Figure 2: Example of 15-minute resolution SMD for a single household (Household_ID: 461158) visualized as a heat map. The graph shows total active electrical energy consumption (in kWh) using colored pixels, with the date on the x-axis and the time of day on the y-axis. A shift in consumption patterns is noticeable after the energy consultant's visit, where the night setback was deactivated as part of the HP optimization.

```python
# import code and load data
from heapo import *
heapo=HEAPO(data_path='/path/to/datafolder/', use_local_time=False)
household_id=461158 # define household of interest
df=heapo.load_smart_meter_data(household_id,resolution='15min')
report_date=heapo.load_protocol_data(household_id)['Visit_Date'].values[0]

# create figure for heatmap, colorbar and data range
fig,ax=plt.subplots(2,2,figsize=(12,8),gridspec_kw={'height_ratios':[0.3,4],'width_ratios':[1,0.05]})

# plot heatmap, color bar, and consultation range
hm = heapo.plot_heatmap(df,'kWh_received_Total',ax=ax[1,0],hour_interval=1,fontsize=14)
heapo.plot_cbar_on_ax(hm['mesh'],cax=ax[1,1],fontsize=14,cbarlabel='Total Energy (kWh)')
heapo.plot_consultation_range_on_ax(ax[0,0],df,report_date,fontsize=14)

# finalize figure formatting
ax[0,1].set_visible(False)
ax[0,0].set_title('Visit: {}'.format(report_date),fontsize=14,pad=10)
fig.suptitle('Household: {}'.format(household_id),fontsize=16,fontweight='bold')
fig.subplots_adjust(top=1.2)
fig.tight_layout()
plt.show()
```

Figure 3: Example of Python code demonstrating the use of the dataloader and its data visualization features. This code snippet produces the graph shown in Figure 2.

are calculated by treating one dataset as the ground truth and the other as the predicted values. Considering that daily cumulative sums are used for billing, they can be regarded as high-quality data. The validation yields an R^2 score of 0.99 ± 0.01 per household and day, a mean absolute error of 0.48 ± 0.12 kWh, and a symmetric mean absolute percentage error of 2.46 ± 0.69 %. Additionally, we validate the weather data by analyzing variations in daily average temperature (*Temperature_avg_daily*) across different weather stations. We compute the standard deviation of daily temperature measurements among the eight stations and evaluate 1,844 days where data is available for all stations. The analysis reveals an average inter-station difference of $1.06 \pm 0.31°C$. The complete implementation details and code for reproducing these results are available in the GitHub repository of the data loader.

3 Descriptive Analyses

This section provides an overview of the study population, analyzes household energy consumption, and highlights common HP issues identified during on-site visits.

3.1 Study Population Statistics

Of all households in the dataset, 1,358 (96.45%) have submitted some Metadata through their utility's online tool. Table 2 summarizes key statistics for all households, as well as for the treatment and control groups separately. Approximately 55.26% use an air-source HP, 40.91% rely on a ground-source HP, while the remaining households have unspecified HP types. The majority of households (60.44%) use their HP also for domestic hot water (DHW) production, while 30.11% rely on an electric water heater (EWH). Additionally, 24.36% have indicated to own at least one electric vehicle (EV), and 34.65% have a PV system, as inferred from SMD measurements indicating energy returned to the grid. The studied households have an average of 2.87 residents (standard deviation: 1.1) and a mean living area of 193.28 m^2 (standard deviation: 71.65).

Table 2: Overview of population based on collected metadata.

Description	All Households	Treatment Group	Control Group
Households - Total	1,408	214	1194
Households - With 15-min SMD	1,407 (99.93 %)	214 (100.0 %)	1,193 (99.92 %)
Households - With daily SMD	1,298 (92.19 %)	156 (72.9 %)	1,142 (95.64 %)
Households - With air-source HP	778 (55.26 %)	109 (50.93 %)	669 (56.03 %)
Households - With ground-source HP	576 (40.91 %)	52 (24.3 %)	524 (43.89 %)
Households - With PV system	516 (36.65 %)	49 (22.9 %)	467 (39.11 %)
Households - With EV	343 (24.36 %)	33 (15.42 %)	310 (25.96 %)
Households - With dryer	1081 (76.78 %)	119 (55.61 %)	962 (80.57 %)
Households - With freezer	1233 (87.57 %)	139 (64.95 %)	1094 (91.62 %)
Households - With radiators	394 (27.98 %)	48 (22.43 %)	346 (28.98 %)
Households - With floor heating	1062 (75.43 %)	117 (54.67 %)	945 (79.15 %)
Households - With DHW by HP	851 (60.44 %)	112 (52.34 %)	739 (61.89 %)
Households - With DHW by EWH	424 (30.11 %)	44 (20.56 %)	380 (31.83 %)
Households - With DHW by solar	70 (4.97 %)	7 (3.27 %)	63 (5.28 %)
Number of residents (Mean ± Std)	2.87 ± 1.1	2.78 ± 1.11	2.88 ± 1.1
Living area in m^2 (Mean ± Std)	193.28 ± 71.65	187.27 ± 60.39	194.11 ± 73.04

3.2 Household Electricity Consumption

Daily household electricity consumption in the *HEAPO* dataset is influenced by outdoor temperature, as heat demand supplied by the HPs varies with ambient conditions. Additionally, factors such as building insulation, number of residents, living area, and personal habits further impact energy use [4]. While the metadata accompanying the SMD enables a detailed analysis of these factors, this paper focuses solely on introducing the dataset and does not explore them in depth. However, to provide an overview of electricity consumption in this dataset, Figure 6 presents different visualizations of daily household energy intensity (kWh per square meter), calculated as the total active electrical energy consumed divided by the reported living area. This metric is used to normalize electricity consumption relative to building size. Figure 6a displays boxplots of daily energy intensity across all households as a function of mean outdoor temperature, rounded to integer values. In contrast, Figures 6b and 6c differentiate by HP type, showing average energy intensity per household for temperature ranges of 0-5°C and 5-10°C. Only households with a minimum of 10 days of data in each temperature range are included. Each scatter point represents a household, with horizontal and vertical error bars indicating individual standard deviations. All graphs reveal a largely linear relationship between daily energy intensity and outdoor temperature, consistent with previous findings in [7]. Additionally, the lower two graphs reveal significant variations in individual household consumption, indicating that factors beyond building size should be considered when developing algorithms to assess HP energy efficiency using SMD.

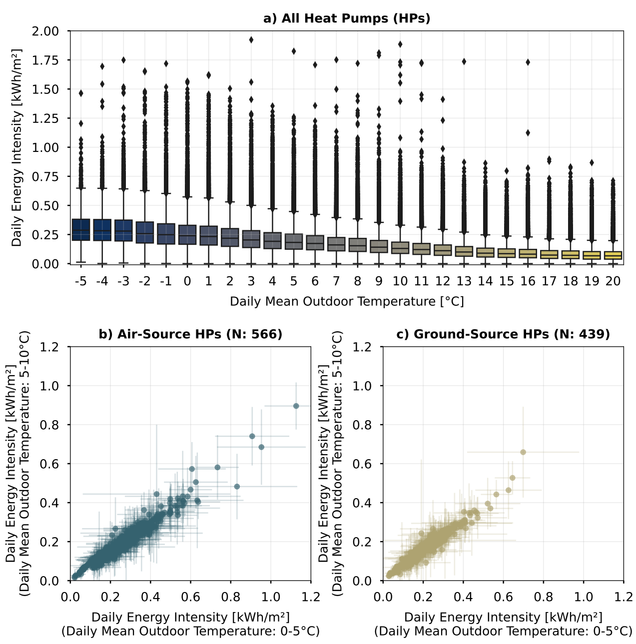

Figure 4: Daily household energy intensity (kWh per square meter of living area) as a function of outdoor temperature for households in the dataset. Each scatter point represents a household mean, with the horizontal and vertical error bars indicating the respective standard deviations of that household's values.

3.3 Common Issues Identified in Protocols

Building on the analysis by Weigert [51], which categorizes common field issues with HPs, we summarize the most frequently reported problems from energy consultants' on-site visits in Table 3. The absolute and relative frequencies are based on the 410 reports included in the *HEAPO* dataset. Several of the most common issues directly impact energy efficiency. The heating curve, which determines the relationship between outdoor temperature and the flow temperature of the heating system, is often set too high. This results in excessive energy consumption, as the HP generates more heat than necessary. Additionally, night setback is frequently activated, causing floor and wall surfaces to cool overnight and requiring extra energy for reheating in the morning. Another common issue is that the heating limit setting, which defines the outdoor temperature below which heating is activated, is often set too high, leading to unnecessary heating. These findings highlight the need for algorithms that assess HP settings for optimal efficiency. Moreover, educating users and installers on proper configuration is essential, as familiarity with the technology has been shown to translate to better HP performance [12, 13, 20, 39].

Table 3: Common issues identified by energy consultants in protocol data from on-site visits.

Reported Issue	Relative Frequency	Absolute Frequency
Heating curve setting chosen too high	40,98 %	168
Night setback mistakenly activated	36,10 %	148
Heating limit setting chosen too high	25,61 %	105
Descaling of DHW necessary	17,80 %	73
Expansion system configured inappropriately	13,41 %	55
Problems with air ducting of air-source HPs	11,25 %	27
HP inappropriately sized	10,00 %	41
Pipes not insulated well enough	9,02 %	37
DHW temperature misconfigured	7,80 %	32
Brine pressure not alright for ground-source HPs	7,19 %	12
Circulation pump regulation set incorrectly	6,34 %	26
Geothermal probe temperature too low for ground-source HPs	5,99 %	10
Install thermostatic valves for individual room regulation	4,88 %	20
HP is dirty	3,17 %	13
HP technically not in order	1,46 %	6
Basic functions of HP are not in order	1,46 %	6

4 Conclusion

The *HEAPO* dataset provides real-world measurements of energy consumption and grid feedback, collected from 1,408 households in the canton of Zurich, Switzerland, between 2018 and 2024. All households utilize HPs for space and water heating. In addition to smart meter data data, the dataset includes metadata, weather data, and protocol data from on-site visits, where energy consultants optimized HP systems. To support research and analysis, the dataset is accompanied by a Python-based dataloader that facilitates data exploration. While the dataset's geographic focus on Switzerland may limit its direct applicability to regions with different climates and building standards, we believe it provides a valuable foundation for advancing HP optimization. Potential applications include estimating energy savings pre-intervention, assessing achieved treatment effects post-intervention, predicting optimal HP configurations for installers and users, analyzing operational HP flexibility, and enhancing load disaggregation methods. By providing this dataset, we aim to support researchers and practitioners in developing innovative methods to optimize HP performance in the field and create scalable digital services.

Acknowledgments

The research project is funded by the Swiss Federal Office of Energy under the grant number SI/502257. We thank Lorenz Deppeler, Hardy Schröder, Michael Peter, and Nina Bichsel for their help on the data collection. Further, we acknowledge the support of other people involved: Jan Marckhoff, Tobias Graml, Dominik Bilgeri, Stephan Renz, Carina Alles, Rita Kobler, and Roland Brüniger.

Code and Data Availability

Available on GitHub: https://github.com/tbrumue/heapo.

Author Contributions

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

References

[1] Philip Agee, Leila Nikdel, and Sydney Roberts. 2021. A measured energy use, solar production, and building air leakage dataset for a zero energy commercial building. *Scientific Data* 8, 1 (2021), 299. doi:10.1038/s41597-021-01082-8

[2] Klaas Bauermann. 2016. German Energiewende and the heating market–Impact and limits of policy. *Energy Policy* 94 (2016), 235–246. doi:10.1016/j.enpol.2016.03.041

[3] Christian Baumann and Peter Kepplinger. 2023. Application of a flexibility estimation method for domestic heat pumps with reduced system information and data. *Cleaner Energy Systems* 6 (2023), 100081. doi:10.1016/j.cles.2023.100081

[4] Christian Beckel, Leyna Sadamori, Thorsten Staake, and Silvia Santini. 2014. Revealing household characteristics from smart meter data. *Energy* 78 (2014), 397–410. doi:10.1016/j.energy.2014.10.025

[5] Zoe Branford and Jennifer Roberts. 2022. The Installer Skills Gap in the UK Heat Pump Sector and the Impacts on a Just Transition to Net-Zero. https://pureportal.strath.ac.uk/en/publications/the-installer-skills-gap-in-the-uk-heat-pump-sector-and-the-impac Last Accessed: 2025-03-19.

[6] Tobias Brudermueller, Fabian Breer, and Thorsten Staake. 2023. Disaggregation of Heat Pump Load Profiles From Low-Resolution Smart Meter Data. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (Istanbul, Turkey) *(BuildSys '23)*. Association for Computing Machinery, New York, NY, USA, 228–231. doi:10.1145/3600100.3623731

[7] Tobias Brudermueller, Markus Kreft, Elgar Fleisch, and Thorsten Staake. 2023. Large-scale Monitoring of Residential Heat Pump Cycling Using Smart Meter Data. *Applied Energy* 350 (2023), 121734. doi:10.1016/j.apenergy.2023.121734

[8] Tobias Brudermueller, Ugne Potthoff, Elgar Fleisch, Felix Wortmann, and Thorsten Staake. 2025. Estimation of energy efficiency of heat pumps in residential buildings using real operation data. *Nature Communications* 16, 1 (2025), 2834. doi:10.1038/s41467-025-58014-y

[9] Tobias Brudermueller, Florian Wirth, Andreas Weigert, and Thorsten Staake. 2022. Automatic Differentiation of Variable and Fixed Speed Heat Pumps With Smart Meter Data. In *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 412–418. doi:10.1109/SmartGridComm52983.2022.9961055

[10] Bundesamt für Meteorologie und Klimatologie MeteoSchweiz. 2024. Messwerte und Messnetze. https://www.meteoschweiz.admin.ch/service-und-publikationen/applikationen/messwerte-und-messnetze.html#param=messwerte-lufttemperatur-10min&table=false Last Accessed: 2025-02-17.

[11] Bundesverband der Energie- und Wasserwirtschaft BDEW e. V. 2009. EDI@Energy OBIS-Kennzahlen-System. https://www.bundesnetzagentur.de/DE/Beschlusskammern/BK06/BK6_81_GPKE_GeLi/Mitteilung_Nr_20/Anlagen/Obis-Kennzahlen-System_2.0.pdf?__blob=publicationFile&v=2 Last Accessed: 2025-02-12.

[12] Sally Caird, Robin Roy, and Stephen Potter. 2012. Domestic heat pumps in the UK: user behaviour, satisfaction and performance. *Energy Efficiency* 5, 3 (2012), 283–301. doi:10.1007/s12053-012-9146-x

[13] Mandy de Wilde. 2021. A Heat Pump Needs a Bit of Care: On Maintainability and Repairing Gender–Technology Relations. *Science, Technology, & Human Values* 46, 6 (2021), 1261–1285. doi:10.1177/0162243920978301

[14] Thomas Decker and Klaus Menrad. 2015. House owners' perceptions and factors influencing their choice of specific heating systems in Germany. *Energy Policy* 85 (2015), 150–161. doi:10.1016/j.enpol.2015.06.004

[15] Robbe Decuypere, Ben Robaeyst, Liselot Hudders, Bastiaan Baccarne, and Dieneke Van de Sompel. 2022. Transitioning to energy efficient housing: Drivers and barriers of intermediaries in heat pump technology. *Energy Policy* 161 (2022), 112709. doi:10.1016/j.enpol.2021.112709

[16] Matteo Dongellini, Claudia Naldi, and Gian Luca Morini. 2021. Influence of sizing strategy and control rules on the energy saving potential of heat pump hybrid systems in a residential building. *Energy Conversion and Management* 235 (2021), 114022. doi:10.1016/j.enconman.2021.114022

[17] Klaus C Engelen. 2023. Heat Pump Fiasco. *The International Economy* 37, 2 (2023), 10–13.

[18] European Commission. 2024. Heat Pumps: Heat pumps are key to enabling the clean energy transition and achieving the EU's carbon neutrality goal by 2050. https://energy.ec.europa.eu/topics/energy-efficiency/heat-pumps_en Last Accessed: 2025-03-19.

[19] David Fischer, Thomas Erge, Marc-André Triebel, and Raphael Hollinger. 2017. Business models using the flexibility of heat pumps-a discourse. In *12th IEA Heat Pump Conference*. 1–12.

[20] C Gleeson, P Biddulph, R Lowe, J Love, A Summerfield, E Oikonomou, J Wingfield, and C Martin. 2017. Analysis of heat pump data from the renewable heat premium payment scheme (RHPP) to the Department of Business, Energy and Industrial Strategy: compliance with MCS Installation Standards. (2017).

[21] Joachim Globisch, Matthias Kühnbach, Elisabeth Dütschke, and Anke Bekk. 2020. The stranger in the German energy system? How energy system requirements misalign with household preferences for flexible heat pumps. *Energy Research & Social Science* 67 (2020), 101604. doi:10.1016/j.erss.2020.101604

[22] Konstantin Hopf, Mariya Sodenkamp, and Thorsten Staake. 2018. Enhancing energy efficiency in the residential sector with smart meter data analytics. *Electronic Markets* 28 (2018), 453–473. doi:10.1007/s12525-018-0290-9

[23] Konstantin Hopf, Andreas Weigert, and Thorsten Staake. 2022. Value creation from analytics with limited data: a case study on the retailing of durable consumer goods. *Journal of Decision Systems* (2022), 1–37. doi:10.1080/12460125.2022.2059172

[24] Yifeng Hu, David P. Yuill, Seyed Ali Rooholghodos, Amir Ebrahimifakhar, and Yuxuan Chen. 2021. Impacts of simultaneous operating faults on cooling performance of a high efficiency residential heat pump. *Energy and Buildings* 242 (2021), 110975. doi:10.1016/j.enbuild.2021.110975

[25] International Energy Agency - Technology Collaboration Programme on Heat Pumping Technology (IEA - HPT TCP). 2023. Annex 56 on Digitalization and IoT for Heat Pumps. https://heatpumpingtechnologies.org/annex56/ Last Accessed: 2025-03-19.

[26] Shobhit Jain, M. Pradish, A. Paventhan, M. Saravanan, and Arindam Das. 2018. Smart Energy Metering Using LPWAN IoT Technology. In *ISGW 2017: Compendium of Technical Papers*. Springer Singapore, Singapore, 19–28. doi:10.1007/978-981-10-8249-8_2

[27] Jack Kelly and William Knottenbelt. 2015. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific data* 2, 1 (2015), 1–14. doi:10.1038/sdata.2015.7

[28] Christoph Klemenjak, Christoph Kovatsch, Manuel Herold, and Wilfried Elmenreich. 2020. A synthetic energy dataset for non-intrusive load monitoring in households. *Scientific data* 7, 1 (2020), 108. doi:10.1038/s41597-020-0434-6

[29] J Zico Kolter and Matthew J Johnson. 2011. REDD: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, Vol. 25. Citeseer, 59–62. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d85a51e2978f4563ee74bf9a09d3219e03799819 Last Accessed: 2025-03-19.

[30] Fabian Krug, Laura Maier, and Dirk Müller. 2025. Quantifying the demand response potential of heat pumps and electric vehicles considering communication protocol constraints. *Sustainable Energy, Grids and Networks* 42 (2025), 101662. doi:10.1016/j.segan.2025.101662

[31] Zachary E. Lee, Qingxuan Sun, Zhao Ma, Jiangfeng Wang, Jason S. MacDonald, and K. Max Zhang. 2020. Providing Grid Services With Heat Pumps: A Review. *ASME Journal of Engineering for Sustainable Buildings and Cities* 1, 1 (01 2020), 011007. doi:10.1115/1.4045819

[32] Jesus Lizana, Claire E. Halloran, Scot Wheeler, Nabil Amghar, Renaldi Renaldi, Markus Killendahl, Luis A. Perez-Maqueda, Malcolm McCulloch, and Ricardo Chacartegui. 2023. A national data-based energy modelling to identify optimal heat storage capacity to support heating electrification. *Energy* 262 (2023), 125298. doi:10.1016/j.energy.2022.125298

[33] Jenny Love, Andrew Z.P. Smith, Stephen Watson, Eleni Oikonomou, Alex Summerfield, Colin Gleeson, Phillip Biddulph, Lai Fong Chiu, Jez Wingfield, Chris Martin, Andy Stone, and Robert Lowe. 2017. The addition of heat pump electricity load profiles to GB electricity demand: Evidence from a heat pump field trial. *Applied Energy* 204 (2017), 332–342. doi:10.1016/j.apenergy.2017.07.026

[34] Hatef Madani and Erica Roccatello. 2014. A comprehensive study on the important faults in heat pump system during the warranty period. *International Journal of Refrigeration* 48 (2014), 19–25. doi:10.1016/j.ijrefrig.2014.08.007

[35] Stephen Makonin, Bradley Ellert, Ivan V Bajić, and Fred Popowich. 2016. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific data* 3, 1 (2016), 1–12. doi:10.1038/sdata.2016.37

[36] Clayton Miller, Anjukan Kathirgamanathan, Bianca Picchetti, Pandarasamy Arjunan, June Young Park, Zoltan Nagy, Paul Raftery, Brodie W Hobson, Zixiao Shi, and Forrest Meggers. 2020. The building data genome project 2, energy meter data from the ASHRAE great energy predictor III competition. *Scientific Data* 7, 1 (2020), 368. doi:10.1038/s41597-020-00712-x

[37] Andrea Monacchi, Dominik Egarter, Wilfried Elmenreich, Salvatore D'Alessandro, and Andrea M Tonello. 2014. GREEND: An energy consumption dataset of households in Italy and Austria. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 511–516. doi:10.1109/SmartGridComm.2014.7007698

[38] Fabian L Müller and B Jansen. 2019. Large-scale demonstration of precise demand response provided by residential heat pumps. *Applied Energy* 239 (2019), 836–845. doi:10.1016/j.apenergy.2019.01.202

[39] Balakrishnan Narayanaswamy, Bharathan Balaji, Rajesh Gupta, and Yuvraj Agarwal. 2014. 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* (Memphis, Tennessee) *(BuildSys '14)*. Association for Computing Machinery, New York, NY, USA, 50–59. doi:10.1145/2674061.2674067

[40] Michael Nast, Ole Langniß, and Uwe Leprich. 2007. Instruments to promote renewable energy in the German heat market – Renewable Heat Sources Act. *Renewable energy* 32, 7 (2007), 1127–1135. doi:10.1016/j.renene.2006.04.014

[41] Lars Nolting, Simone Steiger, and Aaron Praktiknjo. 2018. Assessing the validity of European labels for energy efficiency of heat pumps. *Journal of Building Engineering* 18 (2018), 476–486. doi:10.1016/j.jobe.2018.02.013

[42] Lucas Pereira, Donovan Costa, and Miguel Ribeiro. 2022. A residential labeled dataset for smart meter data analytics. *Scientific Data* 9, 1 (2022), 134. doi:10.1038/s41597-022-01252-2

[43] Carlos Quesada, Leire Astigarraga, Chris Merveille, and Cruz E Borges. 2024. An electricity smart meter dataset of Spanish households: insights into consumption patterns. *Scientific Data* 11, 1 (2024), 59. doi:10.1038/s41597-023-02846-0

[44] R. Renaldi, A. Kiprakis, and D. Friedrich. 2017. An optimisation framework for thermal energy storage integration in a residential heat pump heating system. *Applied Energy* 186 (2017), 520–529. doi:10.1016/j.apenergy.2016.02.067 Sustainable Thermal Energy Management (SusTEM2015).

[45] Arthur Rinaldi, Selin Yilmaz, Martin K Patel, and David Parra. 2022. What adds more flexibility? An energy system analysis of storage, demand-side response, heating electrification, and distribution reinforcement. *Renewable and Sustainable Energy Reviews* 167 (2022), 112696. doi:10.1016/j.rser.2022.112696

[46] Oliver Ruhnau, Lion Hirth, and Aaron Praktiknjo. 2019. Time series of heat demand and heat pump efficiency for energy system modeling. *Scientific data* 6, 1 (2019), 1–10. doi:10.1038/s41597-019-0199-y

[47] Marlon Schlemminger, Tobias Ohrdes, Elisabeth Schneider, and Michael Knoop. 2022. Dataset on electrical single-family house and heat pump load profiles in Germany. *Scientific data* 9, 1 (2022), 56. doi:10.1038/s41597-022-01156-1

[48] Leo Semmelmann, Daniel Schmid, Sarah Henni, Anya Heider, Birgit Schachler, and Christof Weinhardt. 2023. On the impact of heat pump installations and peak blocking strategies on grid expansion costs. In *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*. 1–6. doi:10.1109/ISGTEUROPE56780.2023.10407931

[49] SIA Standard. 1982. 381/3: Heating degree-days in Switzerland. *Swiss Association of Engineers and Architects (in German), Zurich, Switzerland* (1982). https://files.designer.hoststar.ch/5f/6e/5f6e7891-a311-445a-92aa-05a327d8819c.pdf Last Accessed: 2025-03-04.

[50] Alberto Tejeda, Anamaria Milu, Philippe Riviere, and Dominique Marchio. 2014. Energy Consequences of Non-optimal Heat Pump Parameterization. In *International High Performance Buildings Conference*. https://docs.lib.purdue.edu/ihpbc/110/ Last Accessed: 2025-03-19.

[51] Andreas Weigert. 2022. Identification and classification of heat pump problems in the field and their implication for a user-centric problem recognition. *Energy Informatics* 5, 1 (2022), 70. doi:10.1186/s42162-022-00250-3

[52] Andreas Weigert, Konstantin Hopf, Sebastian A Günther, and Thorsten Staake. 2022. Heat pump inspections result in large energy savings when a pre-selection of households is performed: A promising use case of smart meter data. *Energy Policy* 169 (2022), 113156. doi:10.1016/j.enpol.2022.113156

[53] Paul Westermann, Chirag Deb, Arno Schlueter, and Ralph Evins. 2020. Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. *Applied Energy* 264 (2020), 114715. doi:10.1016/j.apenergy.2020.114715

[54] Bochao Zhao, Minxiang Ye, Lina Stankovic, and Vladimir Stankovic. 2020. Non-intrusive load disaggregation solutions for very low-rate smart meter data. *Applied Energy* 268 (2020), 114949. doi:10.1016/j.apenergy.2020.114949

A Appendix

A.1 Acronyms

AMI	Advanced metering infrastructure
ASHP	Air-source heat pump
CDD	Cooling degree day
COP	Coefficient of performance
DHW	Domestic hot water
DR	Demand response
EKZ	Elektrizitätswerke des Kantons Zürich
EU	European Union
EV	Electric vehicle
EWH	Electric water heater
GSHP	Ground-source heat pump
HDD	Heating degree day
HP	Heat pump
IEA	International Energy Agency
OBIS	Object Identification System
PV	Photovoltaic
SFOE	Swiss Federal Office of Energy
SM	Smart electricity meter
UK	United Kingdom
US	United States
UTC	Coordinated universal time

A.2 Additional Explanations of OBIS Codes

OBIS codes are standardized identification codes commonly used for measuring devices and data transmission, enabling the clear identification of measured values such as energy quantities or meter readings. This standard is widely adopted by utilities in German-speaking countries. An OBIS code consists of five value groups in the format **A-B:C.D.E**:

- **A** - Defines the medium (e.g., electricity; gas; water; heat).
- **B** - Specifies internal or external channels.
- **C** - Indicates the measured variable (e.g., active, reactive, or apparent power; current; voltage).
- **D** - Represents the measurement type (e.g., maximum value; current value; time integral).
- **E** - Identifies the tariff category (e.g., total; tariff 1; tariff 2).

The SMD measurements in the *HEAPO* dataset correspond to the following OBIS codes (excluding tariff-specific information):

- **1-1:1.8.0** - Daily meter reading of active energy received from the grid (kWh).
- **1-1:1.29.0** - 15-minute load profile of active energy received from the grid (kWh).
- **1-1:2.8.0** - Daily meter reading of active energy returned to the grid (kWh); available only for households with a PV system and return delivery capability.
- **1-1:5.8.0** - Daily meter reading of reactive inductive energy received from the grid (kvarh); refers to the first quadrant component.
- **1-1:8.8.0** - Daily meter reading of reactive capacitive energy received from the grid (kvarh); refers to the fourth quadrant component.

For further details on Object Identification System (OBIS) codes, see [11, 26].

A.3 Research Project

The public release of this dataset is part of the collaborative research project *KIWO – Künstliche Intelligenz für die Wärmepumpenoptimierung (Artificial Intelligence for Heat Pump Optimization)*. Funded by the Swiss Federal Office of Energy (SFOE) under grant SI/502257[1], the project investigates the intersection of machine learning and SMD to provide households with personalized feedback on HP electricity consumption and optimization strategies. The project is a collaboration between:

(1) **Bits to Energy Lab**: A research group leveraging technology for sustainability, with teams in Switzerland and Germany at ETH Zurich, the University of St. Gallen, the University of Bamberg, and FAU Erlangen-Nuremberg. The research group specializes in machine learning for energy applications and behavioral interventions to enhance user engagement.
(2) **Swiss Federal Office of Energy (SFOE)**: Switzerland's leading authority on energy supply and policy, promoting renewable energy and efficiency. In 2024, the SFOE planned CHF 480 million in investment and managed a CHF 1.3 billion grid surcharge fund.
(3) **Elektrizitätswerke des Kantons Zürich (EKZ)**: The utility for the canton of Zurich, supplying power to 125 municipalities directly and 35 through local distributors. With a 16,000 km grid, EKZ provides around 10% of Switzerland's electricity to 1 million residents and businesses.
(4) **Enerlytica (formerly BEN Energy)**: A leading data analytics company in the energy sector across Germany, Austria, and Switzerland, with over 150 projects and partnerships with 50+ companies.
(5) **Hoval**: A global provider of heating and indoor climate solutions, specializing in HPs, solar systems, and multi-energy heating technologies.

A.4 Detailed Data Variable Descriptions

The following tables provide detailed descriptions of the variables included in the metadata, protocol data, and weather data.

Table 4: Description of metadata variables.

Variable Name	Variable Description
Household_ID	unique household identifier
Survey_Building_Type	type of building (house, appartment)
Survey_Building_LivingArea	living area of building (m^2)
Survey_Building_Residents	number of household residents
Survey_HeatPump_Installation_Type	HP type (air-source, ground-source)
Survey_HeatDistribution_System_FloorHeating	heat distributed by floor heating
Survey_HeatDistribution_System_Radiator	heat distributed by radiator
Survey_DHW_Production_ByHeatPump	DHW produced by HP
Survey_DHW_Production_ByElectricWaterHeater	DHW produced by EWH
Survey_DHW_Production_BySolar	DHW produced by solar collector
Survey_Installation_HasDryer	household has a dryer/tumbler
Survey_Installation_HasFreezer	household has a fridge-freezer
Survey_Installation_HasElectricVehicle	household has an EV

[1]More details: https://www.aramis.admin.ch/Grunddaten/?ProjectID=48862 (Last accessed: 2025 March 14).

Table 5: Description of protocol data variables.

VariableName	Description
Report_ID	unique report identifier
Household_ID	unique household identifier
Visit_Year	year of visit by energy consultant
Visit_Date	date of visit by energy consultant
Building_Type	type of building (single family house, multi family house, other)
Building_HousingUnits	housing units of building
Building_ConstructionYear	year of building construction
Building_ConstructionYear_Interval	interval of building construction year
Building_Renovated_Windows	indication if windows of building are renovated
Building_Renovated_Roof	indication if roof of building is renovated
Building_Renovated_Walls	indication if walls of building are renovated
Building_Renovated_Floor	indication if floor of building is renovated
Building_FloorAreaHeated_Total	total heated floor area of the building (in m^2)
Building_FloorAreaHeated_Basement	heated floor area of the basement of the building (in m^2)
Building_FloorAreaHeated_GroundFloor	heated floor area of the ground floor of the building (in m^2)
Building_FloorAreaHeated_FirstFloor	heated floor area of the first floor of the building (in m^2)
Building_FloorAreaHeated_SecondFloor	heated floor area of the second floor of the building (in m^2)
Building_FloorAreaHeated_TopFloor	heated floor area of the top floor of the building (in m^2)
Building_FloorAreaHeated_AdditionalAreasPlanned	indication if additional areas are planned to be heated any time soon
Building_FloorAreaHeated_AdditionalAreasPlannedSize	size of additional areas that are planned to be heated any time soon (in m^2)
Building_Residents	number of household residents
Building_PVSystem_Available	indication if building has a PV system
Building_PVSystem_Size	size of PV system installed if available (in kWp)
Building_ElectricVehicle_Available	indication if household has one or more EVs or plugin hybrid vehicles
HeatPump_Installation_Type	HP type (air-source, ground-source, water-source)
HeatPump_Installation_Year	year of installation of HP
HeatPump_Installation_Manufacturer	manufacturer of installed HP
HeatPump_Installation_Model	model of installed HP
HeatPump_Installation_HeatingCapacity	heating capacity of installed HP (in kW)
HeatPump_Installation_Refrigerant_Type	type of refrigerant used by HP
HeatPump_Installation_Refrigerant_Content	weight of refrigerant used by HP (in kg)
HeatPump_Installation_Normpoint	operating point of HP - according to product certificate
HeatPump_Installation_Normpoint_COP	COP of HP at standardized operating point - according to product certificate
HeatPump_Installation_Normpoint_ElectricPower	electric power of HP at standardized operating point (in kW) - according to product certificate
HeatPump_Installation_Normpoint_HeatingPower	thermal power of HP at standardized operating point (in kW) - according to product certificate
HeatPump_Installation_Location	location of installed HP(inside, outside, split)
HeatPump_Installation_InternetConnection	indication if HP has a communication module, i.e., can deliver monitoring data via internet
HeatPump_Installation_ControllerNotAccessible	indication if HP controller is not accessible due to password protection
HeatDistribution_System_Radiators	indication if heat is distributed by radiators
HeatDistribution_System_FloorHeating	indication if heat is distributed by floor heating
HeatDistribution_System_ThermostaticValve	indication if heat distribution uses thermostatic valves
HeatDistribution_System_BufferTankAvailable	indication if buffer tank is used for heat distribution
DHW_Production_ByHeatPump	indication if DHW is produced by the HP
DHW_Production_ByElectricWaterHeater	indication if DHW is produced by EWH
DHW_Production_BySolar	indication if DHW is produced by solar collector
DHW_Production_ByHeatPumpBoiler	indication if DHW is produced by HP boiler
DHW_ByHeatPump_TimeInterval	time period that DHW is responsible for DHW production (whole year, half year, independent)
DHW_Production_TypeOfHeating	type of heating for DHW production (central, decentral)
DHW_Production_Residents	amount of residents for which DHW is produced
DHW_Circulation_NotInUse	indication if building does not use circulation for DHW distribution
DHW_Circulation_ByTraceHeating	indication if circulation for DHW distribution is done by trace heating
DHW_Circulation_ByCirculationPump	indication if circulation for DHW distribution is done by circulation pump
DHW_Circulation_SwitchedByTimer	indication if circulation for DHW distribution is switched by timer
DHW_Sterilization_Available	indication if sterilization system for DHW production is available
DHW_Sterilization_Active	indication if sterilization system for DHW production is not only available but also in use (i.e., active)
HeatPump_Clean	indication if HP is clean according to the impression of energy consultant during inspection
HeatPump_BasicFunctionsOkay	indication if basic functionalities of HP are okay or not as assessed by energy consultant during inspection
HeatPump_TechnicallyOkay	indication if HP is technically okay according to the impression of energy consultant during inspection
HeatPump_ElectricityConsumption_YearlyEstimated	estimation of yearly electricity consumption of HP by energy consultant (in kWh)
HeatPump_ElectricityConsumption_Categorization	categorization of HP energy consumption according to energy consultant (normal, rather high, rather low)
HeatPump_Installation_CorrectlyPlanned	indication if HP is sized appropriately according to the energy consultant during inspection
HeatPump_Installation_IncorrectlyPlanned_Categorization	categorization if HP is oversized or undersized (if inappropriately sized according to the energy consultant)
HeatPump_AirSource_AirDuctsDistanceOkay	indication if distance between air ducts of ASHP is okay according to inspection by energy consultant
HeatPump_AirSource_AirDuctsFree	indication if air ducts of ASHP are free according to inspection by energy consultant
HeatPump_AirSource_AirDuctsCleaningRequired	indication if cleaning of air ducts of ASHP is necessary according to inspection by energy consultant
HeatPump_AirSource_AirDuctsDrainOkay	indication if drain of air ducts of ASHP is okay according to inspection by energy consultant
HeatPump_AirSource_EvaporatorClean	indication if evaporator of ASHP is clean according to inspection by energy consultant
HeatPump_GroundSource_BrineCircuit_Length	length of brine circuit of GSHP
HeatPump_GroundSource_BrineCircuit_Depth	depth of brine circuit of GSHP
HeatPump_GroundSource_BrineCircuit_NumberOfHoles	number of drilled holes of brine circuit of GSHP
HeatPump_GroundSource_BrineCircuit_CoolingCapacity	thermal capacity of ground probe of GSHP (in kW)
HeatPump_GroundSource_BrineCircuit_AntiFreezeExists	indication if GSHP uses antifreeze
HeatPump_GroundSource_CurrentPressure	brine pressure of GSHP at time point of inspection by energy consultant
HeatPump_GroundSource_CurrentPressure_Okay	indication if brine pressure of a GSHP is okay according to inspection by energy consultant
HeatPump_GroundSource_CurrentTemperature	brine temperature of GSHP at time point of inspection by energy consultant
HeatPump_GroundSource_CurrentTemperature_Okay	indication if brine temperature of GSHP is okay according to inspection by energy consultant
HeatPump_HeatingCurveSetting_TooHigh_BeforeVisit	indication if HP heating curve setting was chosen too high before the visit of the energy consultant
HeatPump_HeatingCurveSetting_Changed	indication if energy consultant changed the heating curve setting (can be considered as reduction only)
HeatPump_HeatingCurveSetting_Outside20_BeforeVisit	heating curve before visit of energy consultant: supply temperature (in °C) at outdoor temperature of 20 °C
HeatPump_HeatingCurveSetting_Outside0_BeforeVisit	heating curve before visit of energy consultant: supply temperature (in °C) at outdoor temperature of 0 °C
HeatPump_HeatingCurveSetting_OutsideMinus8_BeforeVisit	heating curve before visit of energy consultant: supply temperature (in °C) at outdoor temperature of -8 °C
HeatPump_HeatingCurveSetting_Outside20_AfterVisit	heating curve after visit of energy consultant: supply temperature (in °C) at outdoor temperature of 20 °C
HeatPump_HeatingCurveSetting_Outside0_AfterVisit	heating curve after visit of energy consultant: supply temperature (in °C) at outdoor temperature of 0 °C

Table 5 continued from previous page

HeatPump_HeatingCurveSetting_OutsideMinus8_AfterVisit	heating curve after visit of energy consultant: supply temperature (in °C) at outdoor temperature of -8 °C
HeatPump_HeatingLimitSetting_TooHigh_BeforeVisit	indication if HP heating limit setting was chosen too high before the visit of the energy consultant
HeatPump_HeatingLimitSetting_Changed	indication if energy consultant changed the heating curve setting (can be considered as reduction only)
HeatPump_HeatingLimitSetting_BeforeVisit	setting of HP heating limit before the visit of the energy consultant
HeatPump_HeatingLimitSetting_AfterVisit	setting of HP heating limit after the visit of the energy consultant
HeatPump_NightSetbackSetting_Activated_BeforeVisit	indication if HP night setback was activated before visit(i.e., reduction of supply temperature at nighttime)
HeatPump_NightSetbackSetting_Activated_AfterVisit	indication if HP night setback is activated after visit (i.e., reduction of supply temperature at nighttime)
DHW_TemperatureSetting_Categorization	categorization of DHW setpoint temperature value according to energy consultant (normal, too low, too high)
DHW_TemperatureSetting_Changed	indication if energy consultant changed setpoint temperature for DHW during visit
DHW_TemperatureSetting_BeforeVisit	setpoint temperature for DHW before visit of the energy consultant (in °C)
DHW_TemperatureSetting_AfterVisit	setpoint temperature for DHW after visit of the energy consultant (in °C)
DHW_Storage_LastDescaling_TooLongAgo	indication if storage system of DHW was not descaled for a long time according to energy consultant
DHW_Storage_LastDescaling_Year	year of last descaling of DHW storage system
HeatDistribution_ExpansionTank_Pressure_Categorization	categorization of preset pressure of expansion tank according to energy consultant (too high, too low, okay)
HeatDistribution_ExpansionTank_Pressure_Actual	actual value of preset pressure of expansion tank (in bar)
HeatDistribution_ExpansionTank_Pressure_Target	target value of expansion tank's preset pressure computed by energy consultant ((plant height + 3 meters) / 10) (in bar)
HeatDistribution_ExpansionTank_SystemHeight	plant height (used for calculating target value of preset pressure of expansion tank) (in meters)
HeatDistribution_Circulation_PumpStagePosition_Changed	indication if energy consultant reduced the positional setting of the circulation pump for heat distribution
HeatDistribution_Circulation_PumpStagePosition_BeforeVisit	positional setting of the circulation pump for heat distribution before visit of energy consultant
HeatDistribution_Circulation_PumpStagePosition_AfterVisit	positional setting of the circulation pump for heat distribution after visit of energy consultant
HeatDistribution_Recommendation_InsulatePipes	indication if energy consultant recommends to better insulate pipes in the building
HeatDistribution_Recommendation_InstallThermostaticValve	indication if energy consultant recommends to install thermostatic valves for heat distribution
HeatDistribution_Recommendation_InstallRPMValve	indication if energy consultant recommends to install RPM-regulated circulation pump for heat distribution

Table 6: Description of weather data variables

Variable Name	Parameter	Resolution	Variable Description
Weather_ID			unique weather station identifier
Timestamp			UTC timestamp of measurement (*YYYY-MM-DD HH:MM:SS*)
Temperature_avg_hourly	temperature [°C]	hourly	average air temperature 2 m above ground
Temperature_max_daily		daily	maximum temperature 2 m above ground
Temperature_min_daily		daily	minimum air temperature 2 m above ground
Temperature_avg_daily		daily	average air temperature 2 m above ground
HeatingDegree_SIA_daily		daily	heating degree day number (Swiss standard SIA 381/3 [49])
HeatingDegree_US_daily		daily	heating degree day number (US standard)
CoolingDegree_US_daily		daily	cooling degree day number (US standard)
DewPoint_hourly		hourly	Dew point 2 m above ground (instantaneous value)
Humidity_avg_hourly	humidity [%]	hourly	average relative humidity 2 m above ground
Humidity_avg_daily		daily	average relative humidity 2 m above ground
Precipitation_total_hourly	precipitation [mm]	hourly	total precipitation
Precipitation_total_daily		daily	total precipitation
Pressure_BarometricHeight_avg_hourly	air pressure [hPa]	hourly	average air pressure at barometric height (QFE)
Pressure_SeaLevelStandardAtmosphere_avg_hourly		hourly	average air pressure at sea level with standard atmosphere (QNH)
Pressure_SeaLevel_avg_hourly		hourly	average air pressure at sea level (QFF)
Sunshine_duration_hourly	sunshine duration [h]	hourly	total sunshine duration
Sunshine_duration_daily		daily	Sunshine duration - daily total
WindSpeed_hourly	wind [m/s]	hourly	average wind speed

A.5 Folder and File Structure

Figure 5 presents the dataset's folder and file structure, as detailed in Section 2.2.

```
HEAPO Data
├── meta_data
│   ├── households.csv ........ overview of households and their general data availability.
│   ├── meta_data_variables.csv ........ description of metadata variables.
│   └── meta_data.csv ........ actual metadata per household.
├── reports
│   ├── protocols_variables.csv ........ description of inspection protocol variables.
│   └── protocols.csv ........ actual inspection protocol data (one row per on-site visit).
├── smart_meter_data
│   ├── 15min ........ 15-minute load profile readings transmitted by SMs.
│   │   ├── <household_id>.csv ........ one CSV-file for each household with available data.
│   │   └── ⋮
│   ├── cumulative_counters ........ daily cumulative meter readings transmitted by SMs.
│   │   ├── <household_id>.csv ........ one CSV-file for each household with available data.
│   │   └── ⋮
│   ├── daily ........ SMD in daily resolution derived from cumulative counters.
│   │   ├── <household_id>.csv ........ one CSV-file for each household with available data.
│   │   └── ⋮
│   ├── monthly ........ SMD in monthly resolution derived from cumulative counters.
│   │   ├── <household_id>.csv ........ one CSV-file for each household with available data.
│   │   └── ⋮
│   └── overview ........ overview of each household's SMD availability by resolution.
│       ├── smart_meter_data_15min_overview.csv
│       ├── smart_meter_data_cumulative_counters_overview.csv
│       ├── smart_meter_data_daily_overview.csv
│       └── smart_meter_data_monthly_overview.csv
└── weather_data
    ├── daily ........ daily weather data measurements.
    │   ├── <weather_id>.csv ........ one CSV-file per weather station.
    │   └── ⋮
    ├── hourly ........ hourly weather data measurements.
    │   ├── <weather_id>.csv ........ one CSV-file per weather station.
    │   └── ⋮
    └── overview
        ├── weather_variables_availability.csv
        └── weather_variables.csv
```

Figure 5: File and folder structure of the *HEAPO* dataset.

A.6 Data Availability per Household

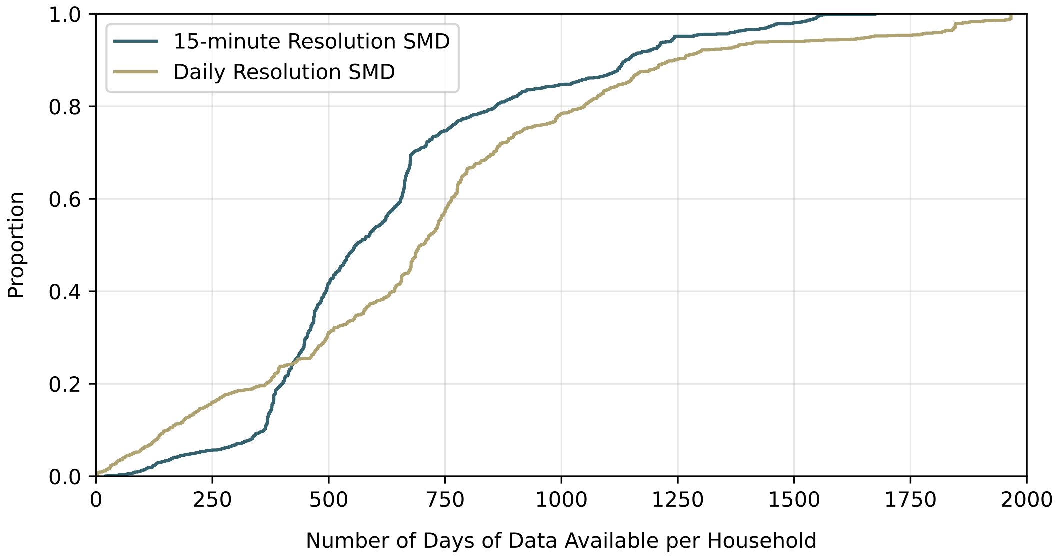

Figure 6: Empirical cumulative distribution function (ECDF) of the number of days of available SMD per household, shown for both 15-minute and daily resolutions. The x-axis indicates the number of available days, while the y-axis represents the proportion of households in the dataset.