An IEEE Societies and Technical Councils Publication
IEEE Data Descriptions

Received 22 August 2024; revised 15 October 2024; accepted 15 October 2024; Date of publication 18 October, 2024; date of current version 14 November, 2024.

*Digital Object Identifier 10.1109/IEEEDATA.2024.3483154*

# Descriptor: *A Norwegian Positive Energy Neighborhood Dataset of Electrical Measurements and Interviews on Energy Practices (NorPEN)*

**APOSTOLOS VAVOURIS [1], FERNANDA GUASSELLI [2], LINA STANKOVIC [1], VLADIMIR STANKOVIC [1], KIRSTEN GRAM-HANSSEN [2], AND SÉBASTIEN DIDIERJEAN [3]**

[1]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XQ, U.K.
[2]Department of the Built Environment, Aalborg University, Copenhagen 2450, Denmark
[3]VIBB AS, Drammen, 3045, Norway

CORRESPONDING AUTHOR: Apostolos Vavouris (e-mail: apostolos.vavouris@strath.ac.uk).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955422.

**ABSTRACT** Emerging concepts for sustainable urban development, such as positive energy districts/neighborhoods (PEDs/PENs), are being implemented in Europe in light of the goal of net-zero carbon emissions by 2050. From the building to the district/neighborhood scales, these projects are fostering the implementation of local renewable energy sources (RES), encouraging the adoption of electric vehicles (EVs) and smart home technologies (SHTs). In this context, a systemic analysis of energy prosumption in these new urban areas is of paramount importance to evaluate their social and technical impacts. This dataset from a Norwegian PEN enables such systemic analysis by bringing the following quantitative and qualitative data: 1) the aggregate loads of six households, 2) the estimated energy production based on the installed solar panel capacity, 3) the labelling of individual activities, 4) solar irradiance and weather data, 5) time of use of electric appliance surveys, and 6) in-depth semistructured interviews on household energy practices. Therefore, this dataset facilitates, multidisciplinary research on the sociotechnical aspects of PENs and net-positive buildings for their design and implementation, helping to answer the "what," "why," and "how" of energy prosumption.

**IEEE SOCIETY/COUNCIL** Signal Processing Society (SPS)

**DATA TYPE/LOCATION** Timeseries, Interviews; Eastern Norway (Østlandet), Norway

**DATA DOI/PID** 10.15129/7d3ac671-2b97-439b-92cf-ce4021e804d2

**INDEX TERMS** Energy efficiency, mixed-methods methodology, nonintrusive load monitoring (NILM), positive energy districts (PED), positive energy neighborhoods (PEN).

## BACKGROUND

According to the International Energy Agency (IEA) [1], the building sector is responsible for 26% of global energy sector emissions, 18% of which are emissions due to the production of electricity and heat used in buildings. To meet the net zero emissions (NZE) by 2050 scenario target [2], the building sector is under pressure to ensure 20% of the building stock is zero-carbon ready by then. Zero-carbon-ready buildings are highly energy-efficient and resilient buildings that either use renewable energy sources (RES) directly, or rely on a source of energy supply that can be fully decarbonized, such as electricity or district energy. Going beyond zero-carbon buildings, recent years have also seen an emergence of net-positive, also known as energy-positive or carbon negative, buildings that can produce more energy than they consume. The concept of net-positive buildings aligns with broader efforts to combat climate change and promote sustainable development by developing more sustainable and resilient communities. These buildings are designed not only to reduce their environmental impact, but also to actively contribute to the generation of clean energy. This is achieved by incorporating advanced energy-efficient technologies, RES, such as solar panels, wind turbines, or geothermal systems, energy storage solutions, and smart building management systems. Through these means, buildings are designed to optimize energy consumption, minimize waste, and use natural resources efficiently.

Within the European Union's strategic plans for energy transition [3], [4], emerging concepts for sustainable urban development are being implemented, which incorporate clusters of net positive buildings to create positive energy districts/neighborhoods (PEDs/PENs) that promote the implementation of local RES, encourage the adoption of electric vehicles (EVs) and smart home technologies (SHTs) [5], [6]. In these areas, buildings are designed not only to reduce their own environmental impact but also to actively contribute to the generation of clean energy at the district/neighborhood level or beyond these boundaries. Through these means, PEDs/PENs are expected to increase the energy independence of communities by reducing their reliance on external energy sources; hence, making them attractive housing solutions due to the significant reduction of their energy bills/costs and carbon footprint.

Although a number of load consumption datasets exist, these are usually collected from consumer-only households that are geographically distributed throughout a country without forming a single neighborhood. The result of a recent review of load consumption datasets is presented in Table I of [7]. Considering the expected increase in PED/PEN a systemic analysis of energy presumption in these new urban areas is of paramount importance to evaluate their social and technical impacts. In this article, we present a dataset from a Norwegian PEN that enables such systemic analysis by bringing quantitative and qualitative data together. To the best of our knowledge, the dataset described in this work stands apart from other available datasets as follows.

1) The first dataset of households that are located in a designed positive energy neighborhood (PEN).
2) Besides energy consumption and disaggregated load, estimated energy production, voltage, and current data are also provided in a sampling rate of 10-s.
3) The utility billing power consumption vector and the variable tariff vector are provided in a sampling rate of 60 min.
4) Weather parameters that can affect consumption and RES production, including, temperature, humidity, and solar insolation.
5) Qualitative data including in-depth semistructured interviews and time-of-use surveys with the homeowners, are included in the dataset.

The design and construction of net-positive buildings and neighborhoods require a multidisciplinary approach, involving architects, engineers, energy experts, sustainability design professionals, and social scientists. This dataset is expected to be valuable to stakeholders involved in the design process of net-zero and net-positive buildings and districts, by facilitating quantitative and qualitative energy analysis, perceived vs actual energy efficiency of designed PED, as well as to stakeholders involved in the study and deployment of smart microgrids with RES penetration. Through this dataset, the effects of hourly variable energy tariffs in end-users' consumption practices can be explored with the dataset already been used to explore the effects of smart energy technologies and how these affect household practices and demand shifting [6] as well as on a mixed-methods data-driven approach for energy-centric evaluation of net-positive households to answer the "what," "why," and "how" of energy prosumption in net-positive dwellings [8].

## COLLECTION METHODS AND DESIGN

The households participating in this study were recruited in a neighborhood in Eastern Norway (Østlandet), which is within the general concept of the PED/PEN, that houses approximately 70 middle-income families. The new urban area consists of several housing zones that are not yet fully developed. In this study, we targeted a zone built between 2018 and 2019 with buildings having a range of different typologies (including detached, semidetached, and flat-apartments) and different sizes (ranging from under 100 to approx. 200 sqm). The total area of the neighborhood of the aforementioned households is approximately 130 m wide by 325 m long. Sustainable technologies, such as solar panels, smart energy management systems, and smart charging for EVs, were installed at the building level, while a ground-source heat pump system was installed at the community level. In addition, passive house standards were also taken into account during the design of the houses. A door-to-door canvassing recruitment process [6] was conducted throughout 10 days in April 2022. Nine households were selected for this study based on two criteria: different demographic characteristics and housing typology. All selected households were equipped with three-phase installations, either fully electric or plug-in hybrid EVs, with some households also having a fast charge point installed. Adhering to GDPR guidelines, written consent declarations were obtained to collect, process, and publish data after anonymization. Out of the nine households, due to connectivity issues, data were collected from six households, more specifically houses 1,

TABLE I. **Summary of Households**

| ID | Type (Area) | PV (kW) | Azimuth (Tilt) | Occupants |
|---|---|---|---|---|
| 1 | 1 (148 $m^2$) | 4.8 | $65^o/-115^o$ ($15^o$) | 2 |
| 3 | 1 (193 $m^2$) | 6.4 | $35^o/-145^o$ ($15^o$) | 3 |
| 5 | 2 (90 $m^2$) | 14 | $40^o$ ($30^o$) | 4 |
| 6 | 1 (148 $m^2$) | 4.8 | $85^o/-95^o$ ($15^o$) | 3 |
| 7 | 2 (90 $m^2$) | 14 | $40^o$ ($30^o$) | 4 |
| 9 | 2 (130 $m^2$) | 9.6 | $-30^o/150^o$ ($15^o$) | 3 |

TABLE II. **Households Used for Training of the Models**

| Target Loads | REFIT [17] | ECO [13] |
|---|---|---|
| Heating | 1, 9, 16 | - |
| Washing machine | 1, 6, 8, 9, 18 | 1 |
| Tumble dryer | 1 | 1 |
| Washer-dryer | 9, 18 | - |
| Dishwasher | 1, 6, 8, 9, 18 | 2 |
| Electric hobs | - | 2 |
| Electric oven | - | 2 |
| Coffee machine | - | 1, 3, 5, 6 |
| Kettle | 6, 8, 9 | 1, 2, 3, 5, 6 |
| Microwave | 6, 9, 18 | 4, 5 |
| Fridge | 8, 18 | - |
| Freezer | 6, 8, 18 | 1, 2, 3 |
| Fridge-freezer | 18 | 1, 2, 3 |
| **EV** | **PECAN** [15] | **Dataset** [16] |
| Low-power (AU) | 661, 1642, 4373, 6139, 8156 | - |
| Low-power (NY) | 27 | - |
| High-power | - | 1 |

3, 5, 6, 7, and 9. Table I presents a summary of the selected households, their typology (detached with 4 rooms = type 1, semidetached with 2/3 rooms = type 2), and their characteristics.

### Household Aggregate Readings

The household aggregate readings were collected via the energy provider through smart meters installed within the households. The data were securely transmitted from the households to the utility provider and then through Azure transferred to the Server located in Glasgow, Scotland. Smart metering data were collected from six households for a period of two months [2022-02-09 23:00:00–2022-04-09 22:00:00 (UTC)]. Although it is challenging to accurately estimate the consumption and production profile of a household by a two-month sample, the period was selected to minimize the intrusiveness to the house occupiers and maximize the extracted information. More specifically the specified two month period was selected because: 1) the monitored period spans almost uniformly before and after the northward equinox and therefore the average solar irradiation is close to the yearly average; 2) the temperature range of the monitored period is close to the yearly average; and 3) the monitored period contained both normal working days and a week of school holidays [8]. The aggregated active/reactive import/export power and the current and voltage of each phase were collected. The voltage readings correspond to the potential difference between each phase and the neutral line.

The smart meter readings were transmitted every 10 s. Readings that failed to be transmitted were discarded from the smart meter and therefore subsequent readings do not contain information about the nontransmitted readings. Although all households have PV panels installed, these are wired in a separate circuit and therefore there is no solar interference in the collected data.

### Activities Decomposition

The major activities, including heating, cooking, laundry/cleaning, and EV charging, were disaggregated from the aggregated readings using WaveNet [9] and seq2subseq [10] neural networks designed for energy disaggregation and based on transfer learning as [11]. The task is simplified due to the availability of three-phase information [12]. Disaggregation results were validated through time-of-use surveys in two selected households and manually verified by a nonintrusive load monitoring (NILM) expert. The disaggregation models training was carried out using publicly available datasets based on the similarity of the load profiles to the six households in Norway (ECO [13], REFIT [14], PECAN [15], and the EV consumption dataset [16]). All datasets were resampled to 10 s by downsampling or upsampling. The training of the networks was performed using a mixture of different households with Table II containing a summary of the households and appliances used. Two different models were trained for the EV load disaggregation, one for high-level dedicated chargers (11 kW) and one for standard three-pin chargers (3 kW).

After disaggregation, the loads were then grouped into household routines based on known relationships between activities and the appliances used in those activities to connect quantifiable data on appliances with the range of activities that define daily life at home [18]. The following four energy intensive activities were considered: 1) EV charging; 2) heating; 3) laundry/cleaning: washing machine, tumble dryer, washer-dryer, and dishwasher activations; and 4) cooking: kettle, coffee machine, hobs, oven, and microwave activations.

Due to the relatively more complex nature and relatively lower power levels, the WaveNet model was used for domestic appliances (top section of Table II containing 13 appliances). On the other hand, the EV charging load, which is by nature a high-energy event, was disaggregated using

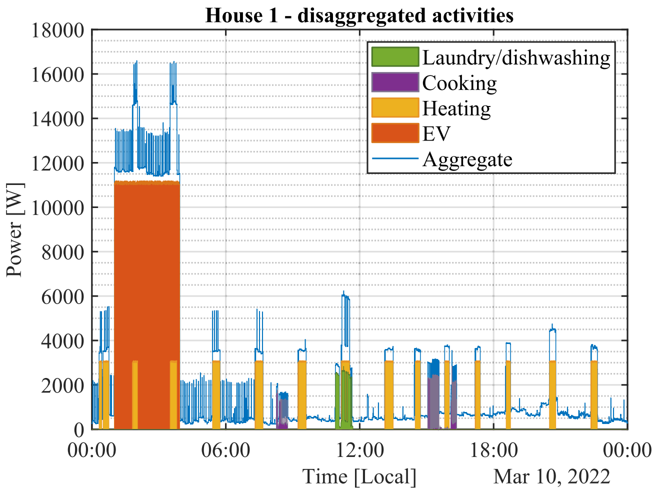


**FIG. 1. House 1: aggregated power and activities breakdown.**

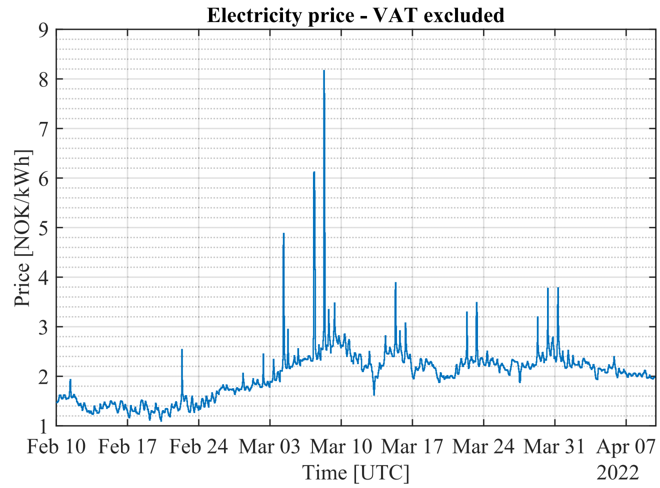


**FIG. 2. Hourly electricity price for the monitored period, VAT excluded.**

the seq2subseq model due to the higher convergence speed and lower computational costs. The default parameters for the WaveNet model were used, whereas for the seq2subseq model, the parameters used are as [11]. The validation of the energy disaggregation results was performed through soft labels obtained from time-of-use surveys that were handed over to the homeowners as well as through manual inspection of the signal by an energy signal expert. Fig. 1 illustrates a sample of the results of the disaggregation of activities as well as a sample of the aggregated active power signal.

### Utility Billing Vector and Hourly Tariff

Apart from low-frequency readings (10 s), the utility provider collected an hourly sample for billing purposes. The hourly sample included the cumulative active and reactive import and export energy, with the active energy samples measured in Watt-hours (Wh) and the reactive energy samples in volt-ampères-reactive-hours (VARh). These variables are used by the utility provider to bill the end-users. The transmission and collection of these readings are more robust as these are required to meet the utility requirements. Therefore, these readings can be used to estimate and interpolate missing values in the dataset. The hourly price vector of electricity in the region where the households that are located is also included in the dataset. The price vector (in Norske Krone, NOK) reflects the price of the energy consumed in the past 1 hour without the inclusion of the value-added tax (VAT). Fig. 2 shows the hourly price vector for the monitored period.

At the time of data being collected the VAT rate was 25%, the grid fees for importing energy from the grid were approx. 0.4 NOK/KWh and the compensation for providing energy to the grid when exporting was approx. 0.1 NOK/kWh. Note that VAT is charged only when importing energy from the grid and not when energy is exported to the grid.

### Weather Data

The weather data were collected from The Norwegian Meteorological Institute [19], including the following variables: air temperature ($^{o}C$), relative humidity ($\%$), surface pressure (hPa), precipitation [in mm/h], wind speed (m/s') and wind direction (deg) sampled at 5-min intervals.

The dew point $T_d$ was calculated through the vapor pressure and saturation vapor expression of the relative humidity as

$$RH = 100\% \times \frac{E}{E_s} \tag{1}$$

where, based on the Clausius-Clapeyron [20] relation, the vapor pressure is given by

$$E = E_0 \times e^{((L/R_v)\times(1/T_0 - 1/T_d))} \tag{2}$$

and the saturation pressure by

$$E_s = E_0 \times e^{((L/R_v)\times(1/T_0 - 1/T))} \tag{3}$$

with the saturation vapor pressure $E_0 = 0.611$ kPa, the latent heat of vaporization $L = 2.453 \times 10^6$ J/kg, the gas constant for moist air $R_v = 461$ J/(kg×K), $L/R_v = 5423$ K, $T_0 = 273.15$ K and $T$ being the air temperature. By solving for the dew point, $T_d$, it will be given in Kelvin by

$$T_d = \frac{1}{\frac{1}{T} - \frac{L}{R_V} \times \ln\left(\frac{RH}{100\%}\right)}. \tag{4}$$

The dew point was converted to Celsius by subtracting the constant 273.15 from the Kelvin temperature. Temperature, dew point, and relative humidity for the entire monitored period are presented in Fig. 3.

The solar data were generated using Copernicus Climate Change Service information 2024 [21] in 1-min intervals[1].

[1]Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

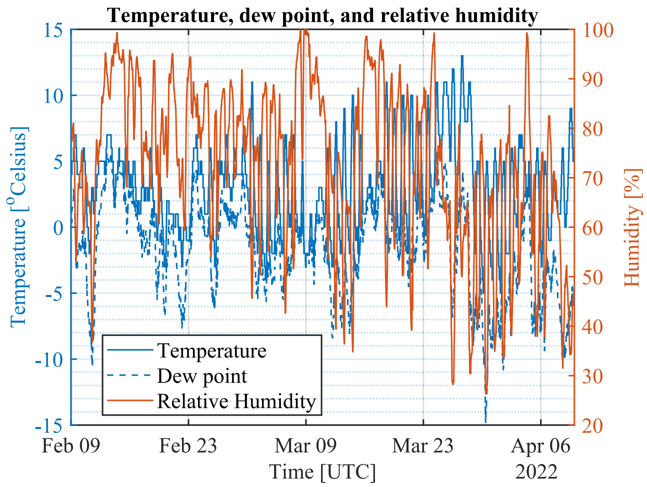


**FIG. 3.** **Weather data for the monitored period.**

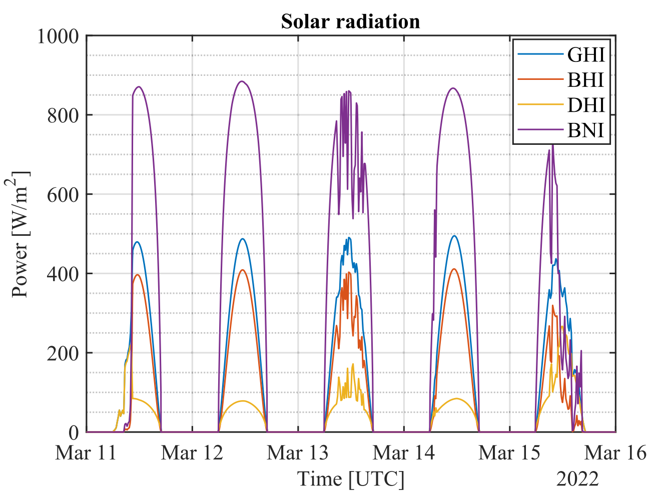


**FIG. 4.** **Solar radiation data sample.**

The collected data were the global (GHI), the direct (BHI), and the diffuse (DHI) solar irradiance, the direct normal irradiance (BNI), the cloud coverage, the cloud type and the albedo. A sample of the collected data is presented in Fig. 4. The zenith angle of the solar disc was calculated as

$$\theta_Z = \cos^{-1}\left(\cos(\phi)\cos(\delta)\cos(\omega) + \sin(\delta)\sin(\phi)\right) \quad (5)$$

where $\phi$ is the latitude, $\delta$ is the declination of the Sun and $\omega$ is the hour angle. The declination [22] of the Sun ($\delta$), with a range $-23.5^{\text{o}} \leq \delta \leq 23.5^{\text{o}}$, is given by

$$\delta = \Phi \times \cos\left(\frac{C\,(d - d_r)}{d_y}\right) \quad (6)$$

where $\Phi$ is the tilt angle and equal to $23.5^{\text{o}}$, $C = 360^{\text{o}}$, $d$ is the Julian Day, $d_r$ is the Julian Day for summer solstice (equal to 172 for nonleap years), $d_y$ is the number of days per calendar year (i.e., 365 days or 366 days for leap years).

The hour angle, $\omega$, is given by

$$\omega = 15^{\text{o}} \times (t - 12) \quad (7)$$

where $t$ is given by

$$t = \text{hours} + \text{minutes}/60 + \text{seconds}/3600. \quad (8)$$

### Solar Production Estimation

As solar production was not monitored, PV production was estimated based on the installed solar capacity, roof tilt, and azimuth angle (Table I) and the local weather and solar data. PV panels are either installed in a fixed tilted rooftop ($30^{\text{o}}$) or on flat rooftops with dual-tilt system ($15^{\text{o}}$). The azimuth angle is measured from South with positive values toward the West and negative toward the East. The PV production estimation was performed using the Global Solar Energy Estimator [23] with a granularity level of one hour. The power output from the PV panels was calculated through the direct and diffuse plane irradiance with an average temperature-dependent solar panel efficiency of 93% [24] and an inverter efficiency of 90%.

### Qualitative Data

The qualitative data consists of in-depth semistructured interviews on household energy practices and a time of use of electric appliances survey. Interviews were conducted face-to-face over 10 days in April 2022, simultaneously with the recruitment process. They lasted an average of 67.5 min and were audio-recorded and transcribed *ad verbum*. Minor language editing was performed after the transcriptions, considering that the interviews were conducted in English, though neither the researcher nor the participants were native English speakers. Due to the semi-structure character of the interviews, the questions were arranged into six themes:

1) "walking through" the smart homes and smart apps to understand how technologies mediate energy practices;
2) motivations for buying a smart home in the new neighborhood;
3) motivation for buying an EV, both to uncover meanings and ways of engagement;
4) understanding changes in energy practices due to new materialities—old house versus new house;
5) heating and cooling practices;
6) sociodemographics.

The themes of the interview guideline were drawn from previous studies on household energy practices within the theoretical framework of social practice theories. In this sense, the interviews aimed to go beyond the traditional occupants' behavior and lifestyle approaches and focus on variations of energy practices (individual energy-consuming habits and routines) that are rooted in collective socio-material structures [25]. Interviews were conducted with all nine households; however, only four interviews are included in this dataset paper, namely households 1, 3, 5, and 9. As interviews with households 6 and 7 included extensive sensitive information, anonymization of the interviews so

**TABLE III. Energy Data Availability as the Number of Available (Actual) and Expected Samples**

| ID | Metering Data Actual/Expected | Metering Data Ratio | Billing Data Actual/Expected | Billing Data Ratio |
|---|---|---|---|---|
| 1 | 505 708/509 400 | 99.28% | 1415/1416 | 99.93% |
| 3 | 505 073/509 400 | 99.15% | 1415/1416 | 99.93% |
| 5 | 486 485/509 400 | 95.50% | 1416/1416 | 100% |
| 6 | 505 835/509 400 | 99.30% | 1415/1416 | 99.93% |
| 7 | 490 564/509 400 | 96.30% | 1414/1416 | 99.86% |
| 9 | 499 672/509 400 | 98.09% | 1415/1416 | 99.93% |
| **Total** | 3 481 320/3 565 800 | 97.63% | 9904/9912 | 99.92% |

**TABLE IV. Quality of Energy Data: Length Data Gaps as % of Samples With Data Gap Lengths Less Than the Given Time Period**

| House ID | 10-s | 1-min | 30-min | 1-h | 6-h |
|---|---|---|---|---|---|
| 1 | 99.18% | 99.93% | 99.93% | 99.93% | 100% |
| 3 | 99.00% | 99.88% | 99.88% | 99.88% | 99.88% |
| 5 | 95.32% | 99.74% | 99.84% | 99.84% | 99.84% |
| 6 | 99.10% | 99.83% | 99.84% | 99.84% | 99.84% |
| 7 | 96.13% | 99.79% | 99.84% | 99.84% | 99.84% |
| 9 | 97.97% | 99.90% | 99.91% | 99.91% | 99.91% |

that they could be understandable was not feasible. In the period following the interviews, a survey on the time of use of electric appliances was sent by phone message or email to the participants; two responses were obtained. A time-of-use survey was developed based on [26] and consisted of type of appliances, frequency of use, time of use during the week, weekend use, weekday use, and appliances for long-term illness.

## VALIDATION AND QUALITY

In this study, energy data from six different households with a total of 34 852 816 samples were collected from a total expected energy-related samples of 35 697 648, which corresponds to a missing data rate of 2.36%. A summary of the actual versus the expected timestamps during which data were collected is presented in Table III. Each timestamp corresponds to 10 readings for the metering data and four readings for the billing data. As expected, the hourly billing vector has statistically less missing values, partly due to the fact that billing data are collected every hour (compared with 10 s for the metering data) and partly due to the prioritization of the collection of billing data to preserve utility and accurate billing.

In addition to data availability, the quality of the collected data was explored by estimating the length of missing data gaps. Table IV includes the percentage of the data that the maximum gap interval does not exceed a period spanning 10 s to 6 h. The majority of the gaps within the dataset are in the range of 10 s to 1 min, with very few gaps having a duration greater than 1 min per household. There has been no gap interval of more than 6 h for any of the six households. Small gaps can be filled through an interpolation method, whereas longer gaps can be filled from average average historical consumption data. All the collected and generated disaggregated streams were manually inspected (visual inspection) by an energy expert to assess the validity of the data. No erroneous spikes were identified in the collected energy readings. The expected accuracy of the disaggregated data is expected to be similar to the demonstrated accuracy of the used disaggregation algorithms as already demonstrated in the literature [9], [10], [11], [12]. As electricity tariff data were obtained through the energy provider, the tariff data do not suffer from missing values, and all 1415 hourly pricing readings are available. PV data were cross-validated through the Prediction of Worldwide Energy Resource (POWER) [27] portal with an average deviation of the solar irradiance data of less than 1%.

The technical validation of qualitative data, such as interviews, is not as objective as the quantitative data. The results that may be obtained from this data rely on the content analysis techniques that will be deployed as well as the theoretical framework chosen by researchers. Nonetheless, since the interview guideline was created based on a theoretical framework of social practices, we can highlight relevant connections between materials, skills, and meanings (i.e., the elements of practices according to [28]) that can be obtained from the interviews. The data reveal several meanings ascribed to RES, EVs, and SHTS, as well as different ways of engagement with such technologies for energy management. Competencies and skills to handle such technologies can also be found throughout the interviews. We summarise a few themes that can be potentially explored in the data below: *Materials:* PVs, ground source heat pumps, smart home technologies, electric vehicles, and smart apps in general. *Skills/Competencies:* Basic tech skills are needed to run the smart home. Some households enjoy acquiring tech skills through interaction with technology, while others prefer/need digital or in-person technical assistance. As a community, the relationship with neighbors in the process of acquiring knowledge on energy technologies was also uncovered. The smart system's complexity and load of information can exclude certain households, such as elderlies and others who do not have time to learn how to handle such devices and mobile apps. The systems can be complex even for households with previous knowledge of energy and IT. *Meanings/engagement:* Affordability, energy efficiency, and convenience are some of the meanings that may be found in the interviews as drivers of households' engagement with their smart homes, EVs and neighborhood. Exclusionary design, unmet expectations, technical issues, time demanded to set up/learn how to set up features and automation, as well as gender issues in handling smart technologies, can be found as some of the reasons for households' disengagement with energy demand.

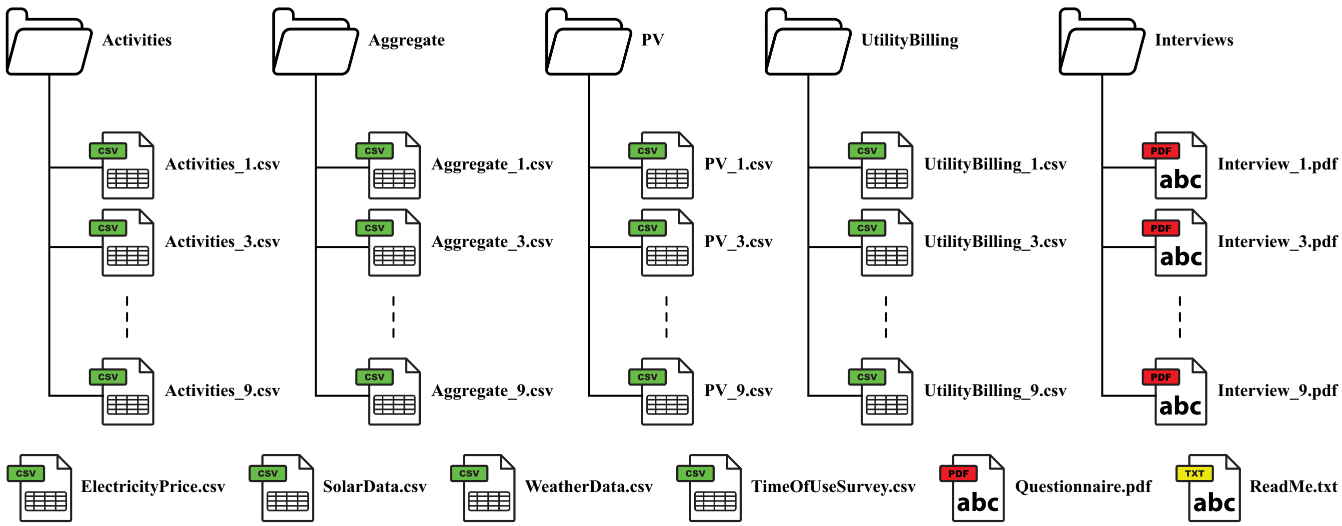


**FIG. 5. Dataset structure.**

## RECORDS AND STORAGE

Adhering to the FAIR principles [29], the recommended file formats by the UK data service for data sharing, reuse, and preservation [30], and the practices in NILM literature [31] the data are made available in the form of CSV and TXT files. There are four CSV files for each household, one containing the household total energy consumed data, one containing the disaggregated activities, one containing the solar production data, and one containing the utility billing vector data. Fig. 5 represents the structure of the dataset. All timestamps are in Coordinated Universal Time (UTC) [YYYY-MM-DD HH:mm:ss] format. UTC format was selected as it is the primary global standard to regulate time. The CSV files ("Aggregate_#.csv") containing the aggregate data have the following columns.

1) $Timestamp$: timestamp of the data point.
2) $ActivePowerPositive$: positive aggregated active power [W].
3) $ActivePowerNegative$: negative aggregated active power [W].
4) $ReactivePowerPositive$: positive aggregated reactive power [VAR].
5) $ReactivePowerNegative$: negative aggregated reactive power [VAR].
6) $PhaseOneCurrent$: phase 1 current [A].
7) $PhaseTwoCurrent$: phase 2 current [A].
8) $PhaseThreeCurrent$: phase 3 current [A].
9) $PhaseOneVoltage$: phase 1 voltage [V].
10) $PhaseTwoVoltage$: phase 2 voltage [V].
11) $PhaseThreeVoltage$: phase 3 voltage [V].

The CSV files ("Activities_#.csv") containing the disaggregated activities data have the following columns.

1) $Timestamp$: timestamp of the data point.
2) $Heating$: estimated heating power [W].
3) $Cooking$: estimated cooking power [W].
4) $LaundryDishwashing$: estimated laundry/dishwashing power [W].
5) $EV$: estimated EV charging power [W].

The CSV files ("PV_#.csv") containing the solar production data have the following columns.

1) $Timestamp$: timestamp of the collected data point.
2) $PV$: estimated PV power production [W].

The CSV files ("UtilityBilling_#.csv") containing the utility billing vector energy data have the following columns.

1) $Timestamp$: timestamp of the data point.
2) $CumulativeActiveImportEnergy$: aggregated active energy imported from the grid from the date of the installation of the meter [Wh].
3) $CumulativeActiveExportEnergy$: aggregated active energy exported to the grid from the date of the installation of the meter [Wh].
4) $CumulativeReactiveImportEnergy$: aggregated reactive energy imported from the grid from the date of the installation of the meter [VARh].
5) $CumulativeReactiveExportEnergy$: aggregated reactive energy exported to the grid from the date of the installation of the meter [VARh].

Further to these, a single CSV file containing the hourly billing vector, a single CSV file that includes the weather variables and a single CSV file containing the solar data are provided. The CSV file ("ElectricityPrice.csv") containing the hourly billing vector has the following columns.

1) $Timestamp$: end timestamp of the billing period.
2) $Price$: hourly market price [NOK] excluding VAT.

The CSV file ("WeatherData.csv") containing the weather data has the following columns.

1) $Timestamp$: end timestamp of the data point.
2) $Temperature$: air temperature in Celsius [°C].

3) $RelativeHumidity$: atmospheric relative humidity [%].
4) $Dewpoint$: dewpoint obtained from (4) [ºC].
5) $SurfacePressure$: surface pressure [hPa].
6) $Precipitation$: rain precipitation [mm/h].
7) $WindSpeed$: wind speed at 10 meters from the ground [m/s].
8) $WindDirection$: wind direction at 10 meters from the ground [º].

The CSV file ("SolarData.csv") containing the solar data has the following columns.

1) $Timestamp$: end timestamp of the data point.
2) $GHI$: global solar irradiance, i.e., the total irradiance on a horizontal surface at ground level [W/sqm].
3) $BHI$: direct solar irradiance, i.e., the beam irradiance on a horizontal surface at ground level [W/sqm].
4) $DHI$: diffuse solar irradiance, i.e., the diffuse irradiance on a horizontal surface at ground level [W/sqm].
5) $BNI$: direct solar irradiance on a mobile plane at normal incidence that follows the sun [W/sqm].
6) $Zenith$: solar zenith angle [º].
7) $Albedo$: reflective coefficient on ground [%].
8) $CloudCoverage$: cloud coverage [%].
9) $CloudType$: cloud type, $-1 =$ no value, $0 =$ no clouds, $5 =$ low-level cloud, $6 =$ medium-level cloud, $7 =$ high-level cloud, and $8 =$ thin cloud.

Further to the quantitative data, the qualitative data are organized in interviews and questionnaire data as follows: 1) the semistructured interviews are organized in a single folder containing the four interviews in PDF format; and 2) the questionnaire template and the replies to the time of use survey are made available in PDF and CSV file accordingly. Last, there is a single TXT read-me file that summarises the content of the dataset.

## INSIGHTS AND NOTES

The dataset is made available in CSV format, which can be easily accessed by the majority of the scientific computing packages, including MATLAB, SPSS, R, and Python.

## SOURCE CODE AND SCRIPTS

The code was developed using MATLAB and Python 3.8 and deployed on a Windows machine. Code from public repositories that have been used in this dataset can be accessed at: https://github.com/DLZRMR/seq2subseq [10], https://github.com/jiejiang-jojo/fast-seq2point [9] and https://github.com/renewables-ninja/gsee [23].

## ACKNOWLEDGMENT

Quantitative methodology: A.V.; qualitative methodology: F.G.; software development: A.V.; quantitative data collection: A.V., S.D.; qualitative data collection: F.G.; quantitative data preparation, curation and analysis: A.V.; qualitative data preparation, curation and analysis: F.G.; visualizations: A.V.; writing original draft: A.V., F.G.; editing: L.S, V.S.; paper review: VS; supervision, funding: L.S., V.S., K.G. All authors reviewed the manuscript.

The authors have declared no conflicts of interest.

The study was registered on 30/03/2022 in the GDPR office of Aalborg University (ref. no AAU031-1005380).

## REFERENCES

[1] IEA, "Tracking clean energy progress 2023," Paris, 2023. [Online]. Available: https://www.iea.org/reports/tracking-clean-energy-progress-2023
[2] IEA, "Global energy and climate model," Paris, 2023. [Online]. Available: https://www.iea.org/reports/global-energy-and-climate-model
[3] European Commission, "The strategic energy technology plan—at the heart of energy research and innovation in Europe," Luxembourg, 2018. [Online]. Available: https://www.iea.org/reports/net-zero-by-2050
[4] IEA, "Net zero by 2050 - a roadmap for the global energy sector," Paris, 2021. [Online]. Available: https://www.iea.org/reports/net-zero-by-2050
[5] J. Brozovsky, A. Gustavsen, and N. Gaitani, "Zero emission neighbourhoods and positive energy districts–a state-of-the-art review," *Sustain. Cities Soc.*, vol. 72, 2021, Art. no. 103013.
[6] F. Guasselli, A. Vavouris, L. Stankovic, V. Stankovic, S. Didierjean, and K. Gram-Hanssen, "Smart energy technologies for the collective: Time-shifting, demand reduction and household practices in a positive energy neighbourhood in Norway," *Energy Res. Social Sci.*, vol. 110, 2024, Art. no. 103436.
[7] S. Athanasoulias et al., "The Plegma dataset: Domestic appliance-level and aggregate electricity demand with metadata from Greece," *Sci. Data*, vol. 11, no. 1, p. 376, 2024.
[8] A. Vavouris, F. Guasselli, L. Stankovic, V. Stankovic, K. Gram-Hanssen, and S. Didierjean, "A complex mixed-methods data-driven energy-centric evaluation of net-positive households," *Appl. Energy*, vol. 367, 2024, Art. no. 123404.
[9] J. Jiang, Q. Kong, M. D. Plumbley, N. Gilbert, M. Hoogendoorn, and D. M. Roijers, "Deep learning-based energy disaggregation and on/off detection of household appliances," *ACM Trans. Knowl. Discovery Data (TKDD)*, vol. 15, no. 3, pp. 1–21, 2021.
[10] Y. Pan, K. Liu, Z. Shen, X. Cai, and Z. Jia, "Sequence-to-subsequence learning with conditional gan for power disaggregation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2020, pp. 3202–3206.
[11] A. Vavouris, B. Garside, L. Stankovic, and V. Stankovic, "Low-frequency non-intrusive load monitoring of electric vehicles in houses with solar generation: Generalisability and transferability," *Energies*, vol. 15, no. 6, p. 2200, 2022.
[12] A. Vavouris, L. Stankovic, V. Stankovic, and J. Shi, "Benefits of three-phase metering for load disaggregation," in *Proc. 9th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities Transp., (BuildSys)*, New York, NY, USA: ACM, 2022, pp. 393–397.
[13] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The eco data set and the performance of non-intrusive load monitoring algorithms," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings (BuildSys)*, New York, NY, USA: ACM, 2014, pp. 80–89.
[14] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study," *Scientific Data*, vol. 4, no. 1, 2017, Art. no. 160122.
[15] Pecan Street Inc., "Dataport," 2022. [Online]. Available: https://www.pecanstreet.org/dataport/
[16] A. Vavouris, L. Stankovic, and V. Stankovic, "Smart meter electricity of a household in Germany with electric vehicle charging annotation," 2022.
[17] D. Murray and L. Stankovic, "Refit: Electrical load measurements (cleaned)," 2016. [Online]. Available: https://doi.org/10.15129/31da3ece-f902-4e95-a093-e0a9536983c4

[18] L. Stankovic, V. Stankovic, J. Liao, and C. Wilson, "Measuring the energy intensity of domestic activities from smart meter data," *Appl. Energy*, vol. 183, pp. 1565–1580, 2016.

[19] Meteorologisk Institutt, "Norsk klima service senter," 2023. [Online]. Available: https://seklima.met.no

[20] O. L. Brown, "The clausius-clapeyron equation," *J. Chem. Educ.*, vol. 28, no. 8, p. 428, 1951.

[21] Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store (ADS), "Cams Solar Radiation time-Series," 2023. [Online]. Available: https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-solar-radiation-timeseries?tab=overview

[22] W. Schüepp, "Die bestimmung der komponenten der atmosphärischen trübung aus aktinometermessungen," *Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B*, vol. 1, pp. 257–346, 1949.

[23] S. Pfenninger and I. Staffell, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, pp. 1251–1265, Nov. 2016.

[24] T. Huld, R. Gottschalg, H. G. Beyer, and M. Topič, "Mapping the performance of PV modules, effects of module type and data averaging," *Sol. Energy*, vol. 84, no. 2, pp. 324–338, 2010.

[25] K. Gram-Hanssen, "New needs for better understanding of household's energy consumption–behaviour, lifestyle or practices?" *Architect. Eng. Des. Manage.*, vol. 10, nos. 1–2, pp. 91–107, 2014.

[26] G. Trotta, A. Hansen, L. Aagaard, and K. Gram-Hanssen, *Survey Questionnaire on Households' Use of Smart Home Technology and Their Time of Use of Electric Appliances*, 1st ed. Institut for Byggeri, By og Miljø (BUILD), Aalborg Universitet, Feb. 2023. [Online]. Available: https://vbn.aau.dk/ws/portalfiles/portal/554031181/Survey_questionnaire_on_households_use_of_smart_home_technology_and_their_time_of_use_of_electric_appliances.pdf

[27] NASA, "Power data access viewer," 2023. [Online]. Available: https://power.larc.nasa.gov/data-access-viewer/

[28] E. Shove, M. Pantzar, and M. Watson, *The Dynamics of Social Practice: Everyday Life and How It Changes,* Newbury Park, CA, USA: Sage, 2012.

[29] M. D. Wilkinson et al., "The fair guiding principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, pp. 1–9, 2016.

[30] UK Data Service, "Recommended formats: File formats recommended by the UK data service," 2024. [Online]. Available: https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/

[31] C. Klemenjak, A. Reinhardt, L. Pereira, S. Makonin, M. Bergés, and W. Elmenreich, "Electricity consumption data sets: Pitfalls and opportunities," in *Proc. 6th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities, Transp.*, 2019, pp. 159–162.