PROCEEDINGS OF SCIENCE

Machine Learning in the O-CEI Swiss Energy Prosumer Pilot

J. White,[a,*] **Raoul Dupuis,**[a] **Mohsen Salimi Khanghah**[a] **and Konstantin Skaburskas**[a,b]

[a] *HEPIA,*
4, Rue de la Prairie, 1202 Geneva, Switzerland
[b] *Digital for Planet,*
111 Überlandstrasse, 8600 Dübendorf, Switzerland
E-mail: john.white@hesge.ch, raoul.dupuis@hesge.ch, mohsen.salimikhanghah@hesge.ch, konstantin.skaburskas@hesge.ch

This paper presents a data-driven analysis of residential electricity consumption in the Swiss pilot of the O-CEI project, using high-frequency (4 Hz) measurements from a household electricity load monitor, including active and reactive power. We describe a practical event-labelling pipeline based on sliding-window Student's t-tests and appliance-specific power/time constraints, and use the resulting labels to train a supervised classifier for appliance detection. A random-forest model is evaluated on test data, showing high recall for kettle events (0.9478) but moderate precision (0.5001), highlighting the impact of class imbalance and noted imperfect automated labels. Beyond foreground appliance detection, we discuss the identification of recurring background loads via spike–steady-state–drop signatures to improve the classifier's performance and enable a more comprehensive understanding of household energy consumption patterns. The paper concludes with a discussion of the implications of these results for flexibility-oriented analysis.

International Symposium on Grids and Clouds (ISGC2026)
15-20, March, 2026
Academia Sinica, Taipei, Taiwan

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0) All rights for text and data mining, AI training, and similar technologies for commercial purposes, are reserved. ISSN 1824-8039 . Published by SISSA Medialab.

PoS(ISGC2026)030

1. Introduction

The O-CEI project [1] is a forward-thinking initiative aimed at addressing key challenges in the European supply chain by creating an open and cross-sector Cloud-Edge-IoT (CEI) continuum platform. O-CEI leverages and upgrades technological innovations from previous successful projects (e.g., existing cloud technologies, including meta-operating systems, cognitive cloud technologies, and decentralized swarm intelligence). It will provide a comprehensive framework for integration, deployment, and validation across eight multidimensional pilots in five European countries, in key strategic sectors including electromobility, software-defined vehicles, logistics, smart agriculture and agrifood, smart urban environments, and multiple electricity grid management. Uptake of the proposed technologies will pave the way for a more sustainable and resilient CEI ecosystem, fostering a smoother transition towards a cleaner energy future.

Energy Transmission System Operators [2] (TSOs) have expressed that chieving flexibility in the energy market may quickly overburden their computing infrastructures. Hence, the concept of the computing continuum as a single, manageable entity that breaks down the silos of edge and cloud computing has gained momentum. Thus, several cross-disciplinary actions are needed to minimize the associated ecological impacts (e.g., in renewable energy and social education). Being able to exploit those resources and have them cooperate in matters of anticipation and prediction in energy flexibility is pivotal for O-CEI and Europe.

2. Swiss Pilot

The goal of the Swiss pilot is to foster prosumers [1] into the increasingly liberalized energy market pursuing high social acceptability, enabling them to reap new forms of electricity production and consumption. The pilot is deployed at four locations in Switzerland, that engage users and energy flexibility in four types of buildings or installations: urban apartments (Geneva); mixed residential/industrial (Fribourg); ski villages (Valais); electric vehicle (EV) charging stations across Switzerland. In general, each deployment consists of energy measurment devices, outlined in Section 3, installed at a building level or at a more granular level (e.g., at the level of different rooms) and a communication infrastructure to transmit the energy data to the O-CEI platform.

This pilot will demonstrate how innovative Smart IoT solutions (already) in place can become part of an integral (O-CEI) platform, increasing such acceptance and user engagement while deploying Smart Energy Services (SES) promoted by advanced O-CEI utilities and technologies. It aims to democratise energy access and support flexible, community-driven ecosystems through secure, AI-enhanced "energy flexibility applications".

Flexibility is defined as the ability to shift/shed the peak load from the demand and/or supply side. This definition applies to different levels: national or regional grids, Local Energy Communities [3] (LECs) and buildings. Flexibility applications can target end-users such as Distribution Systems Operators (DSOs) [2], LECs, businesses and prosumers.

[1] A prosumer is a person (or entity) that both produces and consumes energy.

PoS(ISGC2026)030

3. Energy Measuring Devices

The equipment that is used in the Swiss pilot to measure energy consumption and production is provided by a Swiss company, CLEMAP [4]. CLEMAP is an expert in energy measurement and monitoring solutions, and a technology partner in O-CEI project. In general, the devices are mounted in a fuse box or switch cabinet on a DIN-standard rail, connected to the three-phase electrical installation, and accurately measure voltage and current, including active and reactive power, using the `ade9k` chip [5]. This allows energy measurement at rates up to 12 Hz. By default, energy data is sent to the CLEMAP Cloud in encrypted form via Wi-Fi or Ethernet. The CLEMAP Energy Monitor (CEM) is an energy meter, and the CLEMAP Load Management (CLM) is an energy meter with load management capabilities. The CLM can control the power supply to connected loads: locally via Modbus TCP; or remotely via API (REST API, MQTT, or SDAT/EBIX). Thus, allowing for load shedding or shifting based on predefined rules or real-time conditions.

4. Deployments

The Swiss pilot in O-CEI consists of four deployments across Switzerland, each with distinct characteristics and energy needs. These locations include: urban apartments; tourism infrastructure in a mountain village; a mixed residential/industrial area; EV charging stations at various locations across Switzerland.

4.1 Polygones

The deployment at the Polygones [6] cooperative is located in the city of Meyrin, a suburb of Geneva. The building consists of 28 apartments, with apartment sizes ranging from 3 to 7 rooms [7]. The apartments are equipped with smart meters, smart plugs, and a home energy management system (HEMS) that allows residents to monitor and control their energy consumption. The deployment also includes a photovoltaic (PV) system that generates energy for the entire building.

4.2 Val d'Anniviers

The Val d'Anniviers [8] deployment is located in the mountain villages of Val d'Anniviers in the Valais canton of the Swiss Alps. It addresses the energy needs of the tourism infrastructure, including hotels, restaurants, and ski facilities. The system utilizes CEM devices to monitor energy consumption related to seasonal and variable tourism activities.

4.3 Groupe E

The Groupe E deployment will occur at the Marly Innovation Centre in Fribourg, within the canton of Fribourg. This mixed-use area combines residential, industrial, and technological sectors, each with different energy needs. The deployment will use smart meters and CEM devices, focusing on energy consumption from both residential and industrial activities. The expected electrical loads will be diverse, including high-load heat pumps, electric vehicles, and industrial machinery.

PoS(ISGC2026)030

4.4 CLEMAP EV charging stations

CLEMAP have deployed their CLM devices at various EV charging stations across Switzerland. The CLM devices are connected to multiple EV charging stations which, in turn, are equipped with multiple EV charging plugs. Therefore, there is not a one-to-one correspondence between the CLM devices and the EV charging plugs and the energy consumption captured by the CLM devices is an aggregate of the energy consumption of multiple EV charging plugs. The monitored EV charging loads are then logically grouped into a pool that represents a Virtual Power Plant (VPP), i.e., an aggregated set of distributed flexible loads that can be coordinated as a single resource for energy flexibility services.

5. Scenarios

In the Swiss pilot, each scenario is a set of flexibility applications that rely on the "energy consumption/production forecasting service". So far, three scenarios have been identified. Each of these three scenarios is represented by one flexibility application.

5.1 Flexibility applications for DSOs/LECs

This scenario deals with flexibility applications that enable: DSOs to ensure their power grid stability; LECs to optimise their consumption costs and minimise their dependence on DSOs. Practical examples of flexibility applications for DSOs include:

- Demand response: This application enables DSOs to send signals to consumers to reduce or shift their energy consumption during peak hours or when the grid is under stress. This can be done through price signals, incentives, or direct control of certain appliances.
- Grid balancing: This application allows DSOs to balance the supply and demand of electricity in real-time by adjusting the energy consumption of connected devices or by activating energy storage systems.

Practical examples of flexibility applications for LECs include:

- Time-of-use pricing: This application allows LECs to take advantage of time-of-use pricing by shifting energy consumption to cheaper electricity times, thus reducing their energy costs.
- Energy storage management: This application helps LECs to manage their energy storage systems, by charging them during low energy price times and discharging them during high energy price times, optimising their energy costs and reducing their dependence on DSOs.
- Peer-to-peer energy trading: This application allows LECs to trade energy within, enabling them to sell excess energy generated from renewable sources to other members of the community, thus reducing their dependence on DSOs and optimising their energy costs.

The energy consumption/production forecasting service must be able to provide an accurate forecast of the energy consumption and production of the prosumers and LECs. It must also be able to distinguish the different types of loads and generation sources, and identify which loads can be shifted in time and which cannot be shifted in time.

PoS(ISGC2026)030

5.1.1 Heat pumps

A practical example of flexibility applications for DSOs and LECs is the case of heat pumps. This scenario focuses on heat pumps, which are a key technology for decarbonising heating and cooling in buildings. These appliances typically consume large amounts of electricity, and their energy consumption can be shifted in time without significantly affecting the comfort of the users. The heat pump application can be shown as in Figure 1. In this application, the data collected by the

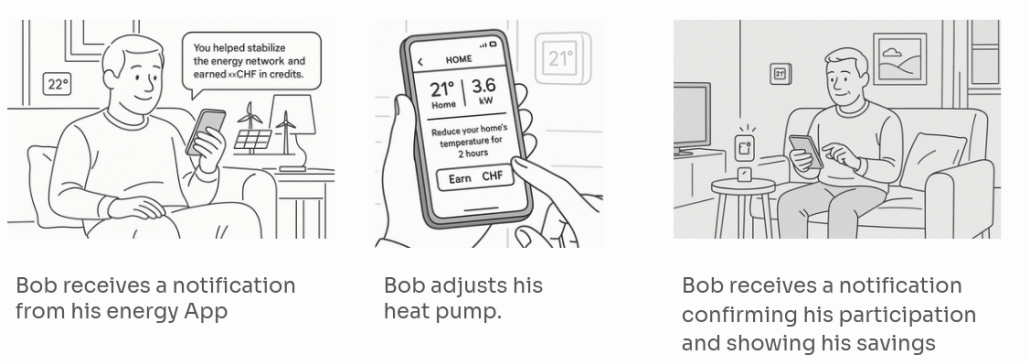


Figure 1: Flexibility application for heat pumps. The prosumer is encouraged through savings.

energy load monitoring devices (e.g. smart meters, CLEMAP devices) is used to forecast the energy consumption of the heat pump. In order to do this, the energy signature, the characteristic pattern of energy consumption of a heat pump, needs to be identified via Non-Intrusive Load Monitoring [9] (NILM) techniques. The flexibility application in this scenario will enable heat pump owners to optimise their energy consumption and costs, while also providing grid services to DSOs.

5.1.2 PhotoVoltaic self-consumption

In principle, the flexibility applications for heat pumps can be extended to other types of flexible loads, such as EVs and industrial machinery. Also, within a household, the loads that can be shifted in time (e.g. washing machine, dishwasher) can be optimised in a similar way, by shifting their energy consumption to cheaper electricity times or when there is excess renewable energy available.

In the case of the Polygones deployment (Section 4.1), a flexibility application can be used to optimize the building's energy consumption by shifting electrical loads to times when excess photovoltaic (PV) energy is available. The energy data collected by the CLEMAP devices needs to be analyzed to identify the energy signatures of the appliances and determine which loads can be shifted in time and which cannot using NILM techniques. For example, the electrical load of a washing machine can be moved, but the fridge load cannot.

5.2 Optimising charging costs of EVs according to grid services

This scenario focuses on charging stations for EVs. The owners of the EV charging stations can use the flexibility applications in this scenario to optimise their energy consumption and costs, while also providing grid services to DSOs. As the energy market in Switzerland is liberalised, the flexibility applications in this scenario will enable EV charging station owners to take advantage of time-of-use pricing and peer-to-peer energy trading. Energy pricing is variable under the DSO

PoS(ISGC2026)030

contract and is posted at least 1 day in advance, so the flexibility applications in this scenario will enable EV charging station owners to shift their energy consumption to reduce their energy costs. The CLEMAP devices deployed at the EV charging stations can communicate with the EV chargers and limit current via a Modbus interface, allowing EV charging station owners to monitor and control their energy consumption in real time.

5.2.1 Peak shaving for EV charging stations

Some energy contracts with DSOs include a monthly peak, which is the maximum energy consumption during a month. Once the monthly peak is reached, the energy price increases significantly for the rest of the month after the peak is reached. Therefore the EV charging station owners have a strong incentive to avoid reaching the monthly peak, and the prediction of whether and when the monthly peak will be reached is important to minimize costs i.e. the peak is “shaved" by reducing the energy consumption of the EV charging stations before the peak is reached. The data collected by the CLEMAP devices at the EV charging stations are used to forecast the EV charging stations 'energy consumption and to predict whether and when the monthly peak will be reached. In order to do this, the energy signature, the characteristic pattern of energy consumption of the EVs at the EV charging stations, needs to be identified via NILM techniques.

5.3 Tourism scenario

This scenario is related to touristic resorts. It will enable hotels, restaurants and ski resorts to optimise their energy consumption. E.g. reducing energy consumption, increasing self consumption, load management, energy storage. This scenario will be defined in more detail in the near future.

6. Social Acceptability

Local Energy Communities (LECs) that are formed primarily in residential areas to ensure electrical autonomy based on the installation of solar panels on the roofs of buildings correspond to one of the major strategies of the European Union [3] to achieve the energy transition. This transition to a renewable energy mix requires strong citizen involvement to promote the stability of electricity networks. In the case of micro smart grids, the consumer becomes both a producer and consumer, a prosumer. To predict and plan the output of local consumption and LECs, data needs to be collected from participants, in addition to monitoring the power cycles of appliances.

On the social sustainability side, a study on the program's acceptance will enable the development of a communication campaign to equip end-users with the means to consciously engage in the transition towards renewable energies. This dimension of social acceptance is crucial for the development of a coherent and coordinated micro smart grids in Europe. The traditional business-to-consumer economic model will be completely challenged in the coming years and replaced by consumer-to-consumer models, as consumers will also become producers.

Social acceptability [10] is a crucial factor in the successful deployment of energy flexibility applications. It encompasses individuals' and communities' willingness to adopt and engage with new technologies and practices that affect their energy consumption and production.

Practical strategies to improve social acceptability include:

PoS(ISGC2026)030

- User engagement: Actively involving users in the development and deployment of SES to ensure their needs and preferences are considered.
- Transparency: Providing clear information about how SES operate, their benefits, and potential impacts on users' daily lives.
- Education: Raising awareness about the benefits of energy flexibility and how users can contribute to a more sustainable energy system.

7. Data Analysis

Here, we present the data analysis of the Swiss pilot, focusing on the management of micro smart grids in residential areas. The data analysis will be based on data collected from households in the pilot, as well as on monitoring the power cycles of appliances. The data will be analyzed using AI algorithms and cooperative, distributed models to create context-aware, self-adaptive energy applications. The analysis will also consider the social acceptance of the smart grid and renewable energy applications, as well as the economic models that will be challenged by the transition towards renewable energies. The results of the data analysis will inform the design and implementation of the energy applications, as well as support the communication campaign aimed at giving end-users the means to consciously engage in the transition towards renewable energies. The data analysis will also be used to evaluate the performance of the smart grid and renewable energy applications, as well as to identify any potential barriers to their implementation and adoption. Overall, the data analysis will play a crucial role in the success of the Swiss pilot and the broader O-CEI project, as it will provide valuable insights into the technical, social, and economic aspects of energy flexibility and the transition towards renewable energies.

7.1 Data Collection

The data used in this analysis comes from a "pioneering" household in the Swiss pilot, which is equipped with a CEM (see Section 3). Care is taken to preserve the privacy of the household, and all data is anonymized to ensure that individual identities or locations cannot be discerned.

The data is collected at a frequency of 4 Hz in order to be sure to recognize appliance transitions of less than 1 second and includes measurements of voltage, current, active and reactive power. The CEM collects data and stores it locally on the device in a SQLite database by default. However, in this case, the data are accumulated locally into batches and streamed to a remote PostgreSQL TimescaleDB [11] database, hosted within the local household network, for further analysis. The measurements are stored with the columns `time`, `l1_u`, `l1_i`, `l1_s`, `l1_p`, `l1_q`, and `l1_energy`, where `time` is the timestamp of the measurement, `l1_u` is the voltage, `l1_s` is the apparent power, `l1_p` is the active power, `l1_q` is the reactive power, and `l1_energy` is the cumulative energy consumption. Note the "`l1`" indicates that the measurements are taken from the first phase of the household's electrical system, the CLEMAP can also measure from all three phases as described in Section 3, but in this household has a single-phase electrical installation.

The important point is that the CEM provides both the active and reactive power measurements, which are crucial for accurate appliance detection and classification. Typically, resistive loads such as kettles and toasters will show more signal in the active power signal compared to the reactive

PoS(ISGC2026)030

power signal, while inductive loads such as fridges and washing machines have a significant reactive component in addition to their active draw. The reactive power measurements can help to distinguish between different types of appliances, as well as to identify the operating state of an appliance (e.g., whether a fridge is in the cooling cycle or the defrost cycle).

The power signal for a typical whole day is shown in Figure 2, where we can see the active and reactive power consumption of the household over a 24 hour period. The power signal shows

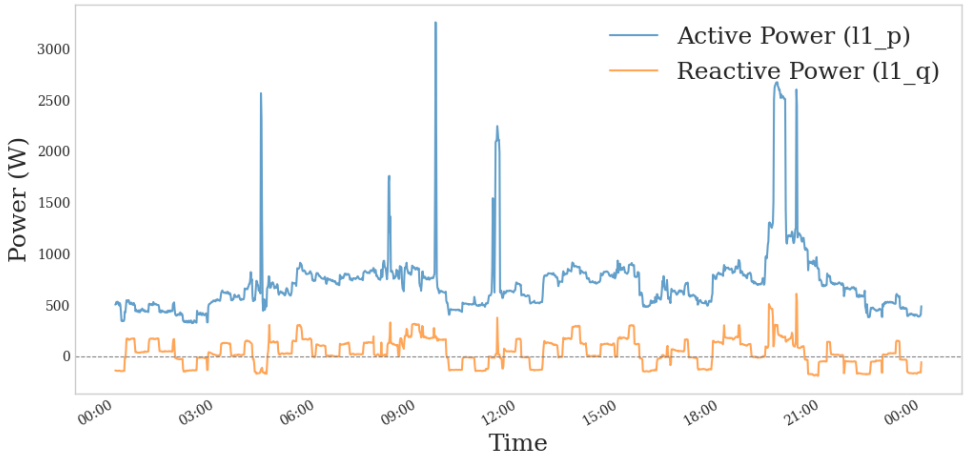


Figure 2: Active and reactive power consumption of the household over a 24 hour period. The data is downsampled to 1 minute intervals for visualization purposes, but the original data is collected at 4 Hz.

the typical daily patterns of energy consumption, with peaks in the morning and evening when the household is most active, and lower consumption during the night when the household is less active. The appliances that are used during the active day time can be considered as power loads that can be moved by the users. The quieter night time can be considered as a baseline for the household's energy consumption, which can be used to identify the appliances that are used during the day. The power loads that are seen during the night time can be considered as background noise, which can be filtered out to focus on the appliances that are used during the day. The data collected from the household will be used to train a classifier to detect the appliances based on their power consumption patterns, as well as to analyze the flexibility of the household's energy usage.

7.2 Appliance Detection and Labelling

The data set is split into a training set and a test set, where the training set (80% of data) is used to train the classifier and the test set (20% of data) is used to evaluate the performance of the classifier. Within the training set, we need to label the data with the corresponding appliances in order to train the classifier. The data split is performed for a single household, so that the training and test sets are not independent in the sense that they come from the same household, with the same appliances and wiring. This means that the classifier may be able to learn the specific characteristics of this household's appliances and wiring, which may not generalize to other households. However,

PoS(ISGC2026)030

the training and test sets are independent in the sense that they are split chronologically, so that the classifier is trained on one part of the timeline and evaluated on a different part of the timeline. This means that the classifier is not able to learn from future data, which is important for evaluating its performance in a real-world scenario where it will be deployed on new data. The techniques used on this household can be applied to other households, but the classifier will need to be retrained for each household, as the appliances and wiring may be different.

A simple labelling process is used to determine whether an appliance is on or off, based on the power consumption. The labelling process is done by comparing the power consumption patterns

Name	Power Level (W)	Tolerance (W)	Min Time (s)	Max Time (s)
Kettle	2500	125	30	300
Fridge	100	20	60	3600
Dishwasher	2000	50	60	3600
Laundry	1500	50	900	7200
TV	200	20	30	3600
Oven	2500	100	60	3600

Table 1: Appliance power levels and tolerances for labelling.

in the training set with the power levels and tolerances as given in Table 1. For example, if the power consumption is around 2500 W with a tolerance of 125 W for a duration between 30 and 300 seconds, we can label a sequence of power measurements as a kettle. Power steps up of the power levels, within the defined tolerances, in Table 1 are used to identify the start of an appliance event, while power steps down of the power levels are used to identify the end of an appliance event.

The power steps up and down are identified using a sliding-window two-sample Student's t-test [12], a statistical test that can be used to determine whether there is a significant difference between the means of two samples. The t-test is applied to the power measurements in a sliding window, where the first window is used as the reference sample and the second window is used as the test sample. Here we use a window size of 2 seconds (8 samples at 4 Hz), an exclude window of 0 seconds and an alpha value for the t-test of 0.001, which means that we are looking for a significant difference between the two samples with a confidence level of 99.9%. The detected up and down power steps are then matched to each other uniquely. This is done by assigning a score to each possible match based on the closeness of the time between the steps (70%), the power level (25%), and the significance penalty (5%). Where the power level score is the absolute difference between the up and down power steps; the significance is calculated at the RMS of the standard deviations of the up and down measurements and the significance penalty is the inverse of significance. Therefore we favour matches that have a close time between the up and down steps, a close power level between the up and down steps, and a high significance of the power steps.

Following the same process for the other appliances, we can label the data with the corresponding appliances, which will be used to train a classifier for appliance detection. Figure 3 shows an example of the labelled data, where we can see the power consumption patterns for different appliances over a period of time. As can be seen in Figure 3, the labelling process is not perfect, and there are some other appliance events ($\approx$ 500 W for 2 min and 1000 W for 1 min) that are not labelled, which can be considered as background noise.

PoS(ISGC2026)030

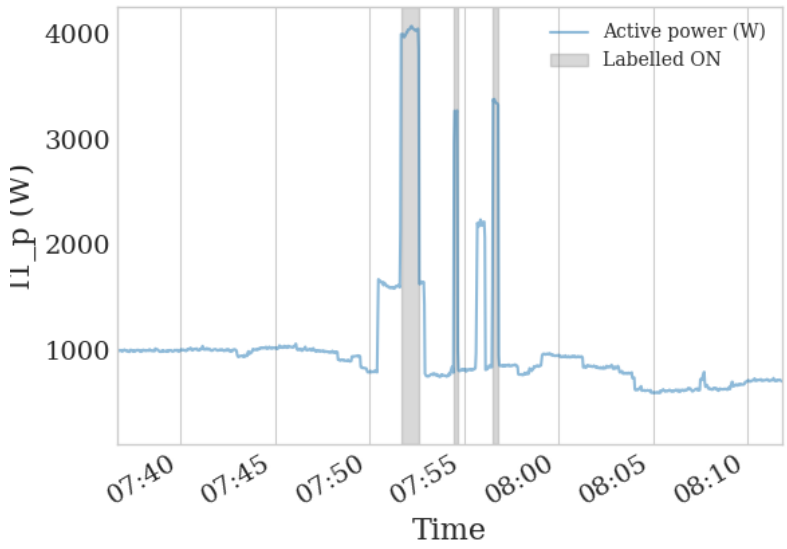


Figure 3: The event labelling is overlaid on the active power signal for three kettle "ON" events.

A complication arises when multiple appliances are simultaneously on, which can lead to overlapping power signatures. In such cases, the labelling process becomes more complex, as we need to consider the combined power consumption of multiple appliances. For example, if a kettle and an oven are on at the same time, the power consumption will be the sum of the power levels of both appliances, which can make it difficult to distinguish between them based on the power consumption alone. In such cases, additional features such as the reactive power measurements, or to consider the temporal patterns of the power consumption help distinguish appliances.

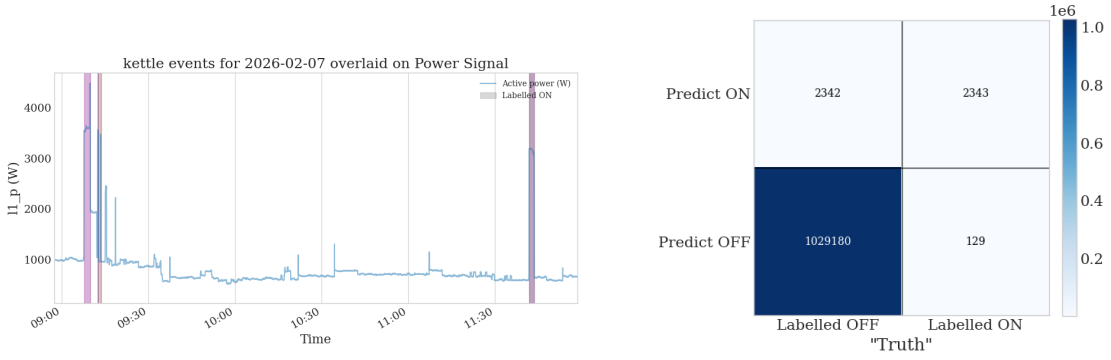
7.3 Classifier Training and Evaluation

The classifier is trained on the labelled training data using a supervised learning algorithm, such as a Random Forest classifier [13]. The model takes the power consumption patterns in the active and reactive power measurements as input features and the corresponding appliance labels as target outputs. The trained classifier is then evaluated on the test set, where we can compare the predicted appliance labels with the true labels. At the present time we do not have an independent method for labelling the test set, see Section 8.1, so we are using the same labelling process as for the training set. This is a limitation, because any systematic errors or assumptions in the labelling procedure may be reproduced in both datasets, leading to a biased estimate of classification performance.

Figure 4a shows an example of the predicted appliance labels from the classifier overlaid in the active power signal for four kettle "ON" events. As can be seen in Figure 4a, the classifier is broadly able to correctly identify the kettle events, but this plot is misleading as it does not show the classification in detail.

As can be seen in Figure 4a, there is a large number of negative samples (no kettle events) compared to a small number of positive samples (kettle events), which can lead to an imbalance in the dataset and can affect the performance of the classifier. The class imbalance is addressed by using

PoS(ISGC2026)030



(a) Predicted appliance labels from the classifier overlaid on the active power signal for four kettle "ON" events.

(b) Confusion matrix for the predicted appliance labels from the classifier.

Figure 4: Prediction results from classifier.

`class_weight='balanced_subsample'` in the RandomForest classifiers plus stratified train/test splitting and precision/recall-based threshold selection, rather than data-level resampling, since our positive class (kettle on events) corresponds to short transient events in continuous time-series data where resampling risks /distorting the temporal structure. We report 0.21% training-segment positive rate (the first 80 % of the timeline) and 0.084% positive rate in the labeled dataset and precision/recall (not accuracy) as our primary metrics to make the class imbalance and its handling transparent.

Another issue that should be recognized in this scheme is that the internal train/validation split used to fit the model and pick the decision threshold is a random row-level shuffle over $\approx$ 4 Hz data. A kettle event spans approximately between 8 and 125 s (32–500 consecutive rows in the time series), so a random shuffle scatters same-event rows across train and "validation", which are not independent, these are near-duplicates of the same physical event a 0.25 s apart. This inflates the validation-based metrics, including the auto-selected probability threshold.

The time of day is already included as an additional feature in the classification model, and is calculated as the hour of the day, which is then cyclically encoded using sine and cosine functions to avoid the discontinuity between 23:59 and 00:00. Also the day of the week is included as an additional feature in the classification model. In the training stage, these and other features are evaluated for their importance in the classification model. A second training stage is then performed where features that are not important, as selected by a cut threshold (10%), are pruned from the model. A modest, but real, signal like time-of-day almost never clears this cut threshold when competing against the raw power signal the label itself is thresholded on. Currently, the time of day and day of week features are not calculated to be important enough to be included in the final model, but they are still considered in the training stage and evaluated for their importance.

A better way to evaluate the performance of the classifier is to show the confusion matrix for the predicted labels, which can be used to evaluate the performance of the classifier and to identify any potential issues with the labelling process or the classifier's ability to distinguish between different appliances. The confusion matrix, as shown in Figure 4b, shows the true positive, false positive, true negative, and false negative rates for each appliance, which can help to identify any potential

PoS(ISGC2026)030

issues with the classifier's performance. For example, if the classifier is correctly identifying the kettle events but is also misclassifying some of the other appliance events as kettle events, this would indicate that the classifier is not able to distinguish between the kettle events and some of the other appliance events, which can be considered as false positives. This is likely due to the fact that the labelling process is not perfect, as it relies on annotations made by an algorithm, and there are some other appliance events that have similar power levels to the kettle events, which can lead to confusion for the classifier.

The performance of the classifier can be evaluated using metrics such as precision, recall, and F1-score. For this classifier, we have a precision of 0.5, a recall of 0.9478, and an F1-score of 0.6547, which indicates that the classifier is performing well in terms of correctly identifying the kettle events, but there are some false negatives where the classifier is missing some of the kettle events. It must be noted though, that this data set is highly imbalanced, with a large number of negative samples (no kettle events) compared to a small number of positive samples (kettle events), which can lead to an overestimation of the classifier's performance. Therefore, it is important to consider the context of the data and the potential issues with the labelling process when evaluating the performance of the classifier.

7.4 Background Identification

The appliance detection described above focuses on identifying events for specific appliances based on their power consumption patterns. In general, this process can be considered a way to identify the "foreground" events in the power signal, i.e., those of interest for analyzing the household's energy usage and flexibility. However, there are also "background" events in the power signal that are not of interest for flexibility, as they represent loads that cannot be moved or stopped. For example, a refrigerator or freezer cannot be unplugged or paused to reduce load or avoid a peak [2], but can still affect overall power consumption and classifier performance. A general characteristic of these types of loads is that they typically include an electric motor, which can also generate a reactive power signal that can be used to help identify them.

If the background loads can be identified and filtered out, this can help to improve the performance of the classifier by reducing the noise in the power signal and allowing the classifier to focus on the foreground events that are of interest for flexibility. For example, there may be other appliances that are not included in the labelling process, or there may be other sources of power consumption such as lighting or electronics that are not directly related to the appliances of interest. In Figure 2, we can see the active and reactive power signal for a typical day, where it can be seen that there are repeating patterens in the reactive power signal that correspond to the background loads.

The feature noted for these types of loads that are both inductive and resistive is that they start with a high "inrush" [14] current. This can be seen as a "spike" in both the active and reactive power signals, followed by lower steady-state power consumption. This spike-steady-state-drop pattern, denoted as a spike-load, can be used to identify the background loads, as well as to distinguish them from the foreground loads that are of interest for flexibility.

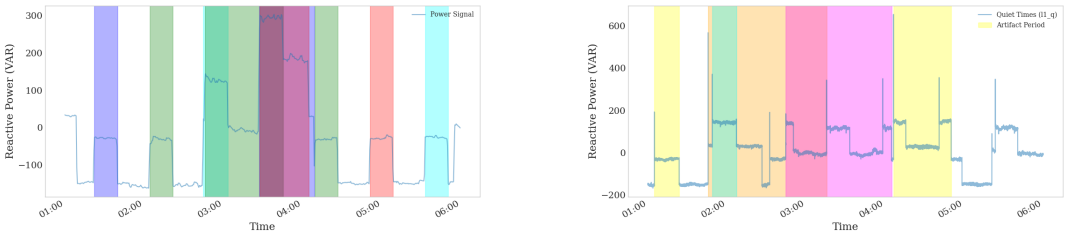
[2] This may not be entirely accurate, as some advanced refrigerators or freezers may have limited flexibility based on their autonomy rating or defrosting cycles.

PoS(ISGC2026)030

The task of identifying background loads can be considered a separate classification problem, where the goal is to classify them based on their power consumption patterns and filter them from the power signal to improve the performance of the foreground appliance classifier. In order to do this the spike-load pattern can be identified using a peak detection algorithm, and then the power consumption up and down patterns following the spike-load can be analyzed to determine whether they correspond to a background load or a foreground load. The spikes are detected using a Median Absolute Deviation [15] (MAD)-based peak detection algorithm, which is a robust method for identifying peaks in a signal while being less sensitive to outliers compared to other methods.

The spike detection uses a hybrid multi-stage spike detector: it first finds candidate anomalies using MAD-based outlier scoring; then independently detects statistically significant step-up transitions using the same t-test method as described in Section 7.2 to find candidate step changes in the power signal; intersects/compares both timestamp sets to flag suspicious MAD-only events and prunes likely false positives by checking nearby power-difference behaviour; unions the surviving MAD and t-test candidates; then refines event timing by snapping each candidate to a nearby local peak and de-duplicates clustered times; recalculates event magnitudes by subtracting pre-event baseline power in a lookback window (excluding a small pre-spike region).

This strategy is employed to attempt to categorize devices that run as background loads, which have a characteristic spike pattern and power on/off pattern, as described above. We use a window between 01h00 and 06h00 ("quiet times") to identify the background loads and search in the reactive power signal so as not to be disturbed by randomness in the active power signal (e.g., from lighting or electronics). As we collect data at 4 Hz we can look for power spikes greater than 100 VAR and less than 1000 VAR that last for 1 s (multiple sample for the spike rise and fall time) and are followed by a steady state power consumption in the reactive domain followed by a power drop. The power drop is matched to the power step before and after the spike using the same scoring method as described in Section 7.2. The "quiet time" windows are split into training and test sets with a split of 50%. Figure 5a shows an example of the labelled spike-power step sequences in a training set quiet time window. As can be seen in Figure 5a, there are clearly repeating signals with ≈ 20 minutes



(a) The labelled spike-power step sequences for background load identification.

(b) The background events predicted by the classifier, overlaid on the reactive power signal for a quiet time window.

Figure 5: Background load identification results. Colour is used to separate the sequences and does not correspond to different appliances.

duration repeating every ≈ 40 minutes. Between 03h00 and 04h00 there are overlapping signals. We used a supervised binary random forest classifier trained on labeled spike-drop candidate pairs with engineered temporal and power-matching features, where the target indicates whether a candidate

PoS(ISGC2026)030

drop is the true matched drop for a given spike.

The model is then evaluated on the test set quiet time windows, where we can compare the classifier-predicted background load events with the reference labels. Figure 5b shows an example of the classifier-predicted background load events, overlaid in the reactive power signal for a quiet time window. As can be seen in Figure 5b, the classifier is broadly able to correctly identify the background load events, but this plot is misleading as it does not show the classification in detail. The main problem seen in this plot is that there are some false negatives where the classifier is missing some of the background load events. This is likely due to the fact that the labelling process for the background loads is not perfect. This is addressed in Section 8, where we discuss the need for a more accurate labelling process for both the foreground and background loads, which can be achieved by using an independent method to identify when appliances are on, such as manual recording or smart plugs.

8. Further work

There are some shortcomings in the current analysis that need to be addressed in future (and current) work, such as the fact that the labelling process for both the foreground and background loads is not perfect, which can lead to issues with the training and evaluation of the classifiers.

8.1 Labelling

To produce a more accurate classifier, we need a more accurate labeling process, which can be achieved by using an independent method to identify when appliances are on. This could be achieved by asking the inhabitants to manually record, at worst on paper and at best via an application on a smartphone or tablet, when appliances are used, noting the start and end times of each use.

A more sophisticated approach could involve equipping the appliances of interest with sensors or smart plugs to automatically record their usage based on their power consumption patterns. In the case of the Swiss pilot, data security and privacy concerns make it difficult to use smart plugs for labeling if they are connected to a manufacturer's cloud service, as is the case with many commercially available smart plugs.

However, there are some smart plugs that can be used for labelling without connecting to a cloud service, such as the SONOFF smart plug [16], which uses the Zigbee protocol. The smart plug is paired with a Zigbee coordinator, which can be connected to the local network and can stream the data to the local PostgreSQL TimescaleDB database via MQTT. There are some limitations to using smart plugs for labelling, such as the fact that they can only be used for appliances that are plugged into them, and they may not be able to capture the usage of appliances that are hardwired into the electrical system.

Some work to equip the appliances of interest in the household has started. These appliances fall into two categories:

- The appliances that are of interest for flexibility, which are typically the appliances that can be moved or stopped, such as the kettle, dishwasher, laundry machine, TV and oven;
- The appliances that are not of interest for flexibility, which are typically the background loads such as the refrigerator and freezer.

PoS(ISGC2026)030

Two figures, one of a foreground load (kettle) and one of a background load (dehumidifier) are shown in Figures 6. In the left pair of plots in Figure 6, the bottom plot shows the total power

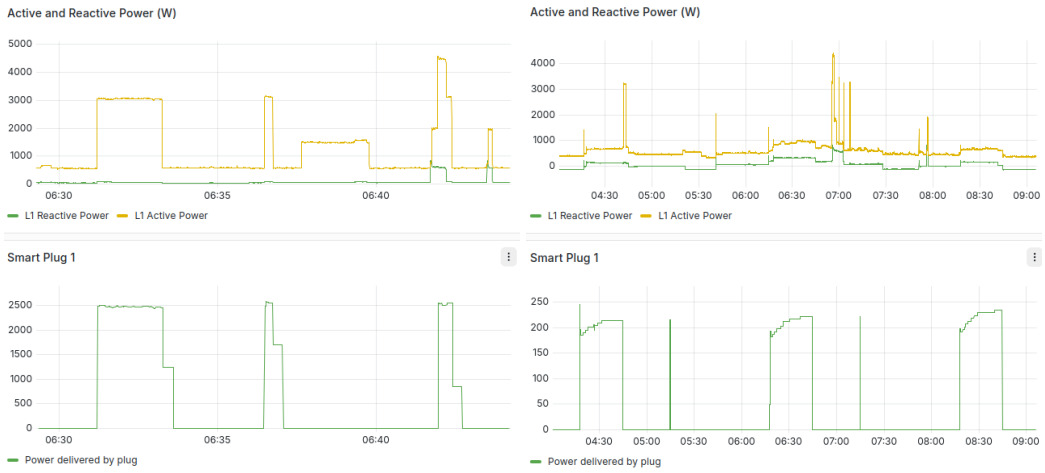


Figure 6: The power delivered by the smart plug is shown on the bottom plots and the active and reactive load as seen by the CLEMAP device on the top plots.

delivered through the smart plug for the kettle, and we can see three events where the kettle is switched on and off. In the top plot, we can see the corresponding active and reactive power measurements from the CLEMAP device, with the active power dominating for these three events.

In the right pair of plots in Figure 6, the bottom plot shows the power delivered through the smart plug for the dehumidifier, and we can see three events where the dehumidifier is switched on and off. In the top plot, we can see the corresponding active and reactive power measurements from the CLEMAP device, showing the power load in both domains, with the reactive power load clearly visible in sync with the active power load as compared to the kettle case.

With knowledge of the appliances attached to the smart plugs, we can use this independent data stream to label (timestamp) the CLEMAP data to identify which and when the appliances are on, which can be used to train a more accurate classifier for appliance detection and background load identification. This work is in the early stages and ongoing, but it is expected to soon lead to a significant improvement in the classifiers' performance.

8.2 Cloud-Edge IT Architecture

The findings here can relate to the design and optimization of smart energy and Cloud-Edge IT infrastructure. The 4 Hz data collected by the CLEMAP device provides high-resolution insights into appliance usage patterns, which can inform both local and cloud-based decision-making and preserve data privacy. The choice of using SONOFF/Zigbee smart plugs (or other local edge devices), which do not rely on a manufacturer's cloud service, aligns with the edge-focused architecture and ensures that sensitive appliance usage data remains within the local network.

Depending on how the classifier is used in the future, the classifier's model could be run on the CLEMAP device directly, and its output could be used to inform local control decisions. If the classifier is re-trained on subsequent smaller datasets, the model could be updated and deployed to the CLEMAP device which enables continuous improvement of the local inference

PoS(ISGC2026)030

and control capabilities. If the classifier is retrained on larger datasets, the model must be updated and deployed to a cloud-based infrastructure, as the physical limitations of the CLEMAP device prevent it from handling larger datasets locally. The cloud-based infrastructure can provide more computational resources for training and updating the model, enabling more powerful analytics and decision-making capabilities.

Given current precision (0.50) and recall (0.9478), the classifier is not yet reliable enough to drive autonomous local actuation (e.g., an edge gateway independently triggering CLM load-shedding on "kettle detected"). At this precision level,a false-positive rate this high would misfire real control action too often. At this stage, the classifier is better suited to cloud/batch-side analytics (periodic flexibility assessment and the app-notification pattern shown in Figure 1).

The RandomForest training itself (multi-million-row feature matrices, needing careful memory management) is a batch/cloud-side workload. The inference workload is much lighter, and the current results suggest that the trained, pruned model (small feature set, capped depth) is cheap enough to run inference on constrained edge hardware near the CLEMAP device.

8.3 Broader smart energy infrastructure applications

The results presented here are drawn from a single household, but the labelling pipeline, feature engineering and RandomForest classifier are not specific to this household, and the same methodology can be applied to the other deployments described in Section 5. The Polygones cooperative (Section 4.1), for example, comprises 28 individually metered apartments (only 5 so far are metered via a CLEMAP device); running the same per-appliance detection pipeline across all of them would move the analysis from a single load profile to a distribution of load profiles for a building-level micro smart grid, enabling aggregated flexibility signals (e.g. how many kettles or washing machines could plausibly be shifted at a given time) to be exposed to a LEC or DSO rather than to a single prosumer.

This could also be applied to the mixed residential/industrial site at Groupe E (Section 4.3), where the spike to steady-state and drop signature used here for background loads could be adapted to identify the duty cycles of heat pumps and industrial machinery (Section 5.1.1), feeding directly into the heat-pump and PhotoVoltaic self-consumption flexibility applications. This is expected to be more challenging than this single-household case, due to the high variability in occupancy and machine usage throughout the year.

The seasonal, tourism-driven loads at Val d'Anniviers (Section 4.2) present a harder case, since occupancy and appliance mix vary strongly over the year; there, the classifier would likely need periodic retraining, or a transfer-learning approach that reuses features learned on one site to bootstrap labelling on another, rather than a single fixed model.

Beyond individual buildings, the CLEMAP-equipped EV charging stations (Section 4.4) already aggregate many charging plugs into pools coordinated as a Virtual Power Plant (VPP). The same appliance-detection logic applied here to a kettle or dehumidifier could, in that context, be applied to individual charging sessions to distinguish, for instance, opportunistic top-up charging from full overnight charging. In turn, this gives the VPP operator a finer-grained, per-session estimate of how much load in the pool is actually available for peak shaving (Section 5.2.1) or grid balancing, rather than treating the pool as a single aggregate load.

PoS(ISGC2026)030

More generally, once such classifiers are deployed at scale across many households, buildings and EV charging stations, the Cloud-Edge architecture discussed in Section 8.2 becomes more important. The local, per-site inference keeps raw appliance-level data on the edge device, while only the resulting flexibility estimates, not the underlying power traces, are aggregated centrally to support LEC or DSO-level demand response, grid balancing and time-of-use optimisation across the micro smart grid as a whole.

This points to a natural next step for the pilot: evaluating whether models trained on one deployment transfer, with little or no retraining, to another, which would determine how much of the labelling effort described in Section 8.1 must be repeated at each new site as the pilot scales from a single household to the full portfolio of O-CEI pilot deployments.

9. Conclusions

This article presents the work in progress of the Swiss pilot in O-CEI, focusing on NILM-based identification of foreground and background loads. The work shows promise in providing load identification and forecasting services. It is realised that the performance of the classifier is likely to be highly dependent on the quality of the labelled dataset, and therefore more work is needed to ensure that the labelled dataset is of high quality. There is more practical work needed to produce a labelled dataset of foreground and background load events, and to evaluate the performance of the classifier in more detail. This will involve deploying independent load-monitoring devices to provide reference labels for foreground and background load events, and using more advanced evaluation metrics to assess the classifier's performance in greater detail.

References

[1] O-CEI project. https://o-cei.eu/.

[2] Energy operators. https://www.gridx.ai/knowledge/what-is-a-grid-operator.

[3] Georgia Asimakopoulou and Nikos Hatziargyriou. Models for the integration of local energy communities in energy markets. June 2018.

[4] CLEMAP. https://www.clemap.com/.

[5] ADE9000 energy metering datasheet. https://www.scribd.com/document/735679082/8816508-ade9000.

[6] Polygones cooperative. https://www.polygones.ch/.

[7] Counting rooms in switzerland. https://www.swisscaution.ch/en/blog/post/count-rooms-switzerland/.

[8] Val d'Anniviers municipality. https://www.valdanniviers.ch/fr/.

[9] G.W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

PoS(ISGC2026)030

[10] E. Fragniere, S. Sandoz, N. Abdennadher, M. Moussa, G. D. M. Serugendo, and P. Glass. Improving the social acceptability of microgrids. In *2023 IEEE Renewable Energy and Sustainable E-Mobility Conference (RESEM)*, pages 1–5. IEEE, 2023.

[11] PostgreSQL TimescaleDB. https://www.tigerdata.com/.

[12] Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

[13] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[14] Inrush current. https://en.wikipedia.org/wiki/Inrush_current.

[15] P. J. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993.

[16] SONOFF iplug Wi-Fi smart plug S60 series. https://sonoff.tech/products/sonoff-iplug-wi-fi-smart-plug-s60-series.

PoS(ISGC2026)030