Vasile Mihaela Alina
131 IEC

Project

-DSP-

1. Introduction

This project is based on the **CCAI tutorial** "Smart Meter Data Analytics: Practical Use-Cases and Best Practices of Machine Learning Applications for Energy Data in the Residential Sector." Its main goal is to **explore how smart meter data** — time-series measurements of electricity consumption — can be analyzed using machine learning techniques to gain insights into **residential energy use**.

Smart meters provide **detailed consumption data at regular intervals** (in this case, 15-minute resolution), enabling the monitoring, forecasting, and optimization of energy demand. These capabilities are essential for **supporting the transition** to **sustainable energy systems.**

2. Preparations

The initial steps of this project involved preparing the environment by importing relevant Python libraries for:

- **Data manipulation**: pandas, numpy
- **Visualization**: matplotlib, seaborn
- **Change point detection**: ruptures
- **Symbolic Aggregate Approximation**: saxpy
- **Feature extraction**: tsfresh
- **Machine learning**: scikit-learn

These tools support various stages of data processing, analysis, and modeling.

2.1 Importing packages

The project also implements several helper functions to:

- **Handle timezone** conversions to align data timestamps correctly
- **Plot raw and aggregated energy consumption data** to visualize patterns over time
- **Calculate daily energy consumption** and mean daily profiles to summarize usage trends

Together, these components form the foundation for more advanced analyses, including anomaly detection, load disaggregation, and predictive modeling. The ultimate goal is to improve energy efficiency and enable smarter energy management in residential buildings.

2.2 Importing data

Next, we import real-world smart meter data sampled at 15-minute intervals. The dataset comes from four Swiss households, each equipped with two smart meters: one for the heat pump and another for all other appliances. Additionally, we import temperature data from a nearby weather station, which will be useful for subsequent analysis.

3. Best practices for visualizing smart meter data

Once the smart meter data has been successfully loaded and cleaned, the next crucial step is to explore and understand it. Simply printing raw numbers or tables is rarely effective, as the sheer volume of data can be overwhelming and meaningful patterns may be obscured by noise. Instead, graphical visualizations provide a much more intuitive way to quickly identify trends, anomalies, and behavioral patterns over time.

Below, we outline several best practices and visualization techniques that are particularly useful when working with time-series energy consumption data from smart meters.

3.1 Time-series visualizations of energy data

The most straightforward way to visualize smart meter data is through a **line chart** that plots **energy consumption over time**. For example, plotting power readings at regular intervals across multiple days provides a comprehensive view of the dataset and reveals patterns such as **value ranges** and **periods of missing data**.

However, there are some challenges:

- The raw data is often **very dense**, making it difficult to zoom in and detect finer daily behaviors or relate spikes to specific times.
- **Daily cycles**, **seasonal changes**, and short-term fluctuations can overlap and clutter the plot.

To address these challenges, a common approach is **data aggregation** or averaging:

- Calculate average power consumption over defined intervals, such as **15 minutes** or **1 hour**. This smooths out fluctuations and highlights meaningful patterns.
- By reducing data resolution, it's easier to observe **daily routines**, compare **weekdays to weekends**, and detect seasonal trends.

For instance, averaging data at 15-minute intervals reveals clearer daily profiles showing how power consumption rises and falls throughout the day. **Hourly averaging** further simplifies the data and is especially useful when visualizing trends over **weeks or months**.

Vasile Mihaela Alina
131 IEC

A)plot all days in 1 graph , one after the other - 1s resolution

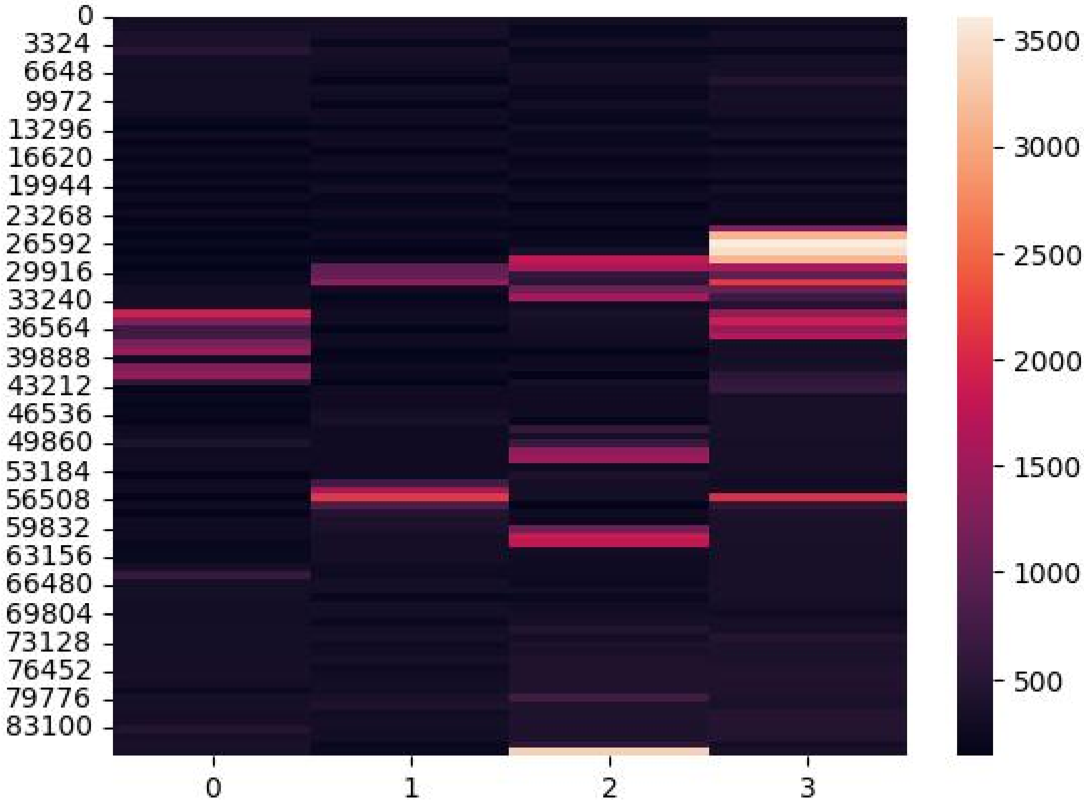


Fig 1a) - plot all days (1s resolution)

This figure presents the total electricity consumption for multiple days at a very high resolution (1 second). It shows all the raw, detailed fluctuations but can appear visually dense, making it challenging to discern overall patterns due to the granularity. These would represent the instantaneous power draw changes from individual household events, like a light bulb being switched on/off, or the slight power adjustment of an electronic device in real-time.

Raw electricity consumption data with extreme detail:

- The **dense, noisy appearance** reflects every tiny power fluctuation
- **Each spike** could be someone turning on a light switch
- Continuous variations show appliances cycling on/off (like refrigerator compressors)
- **The "fuzzy" appearance is normal** - homes constantly have small power changes

Vasile Mihaela Alina
131 IEC

B) Plot all days in 1 graph, one after the other - 1h resolution

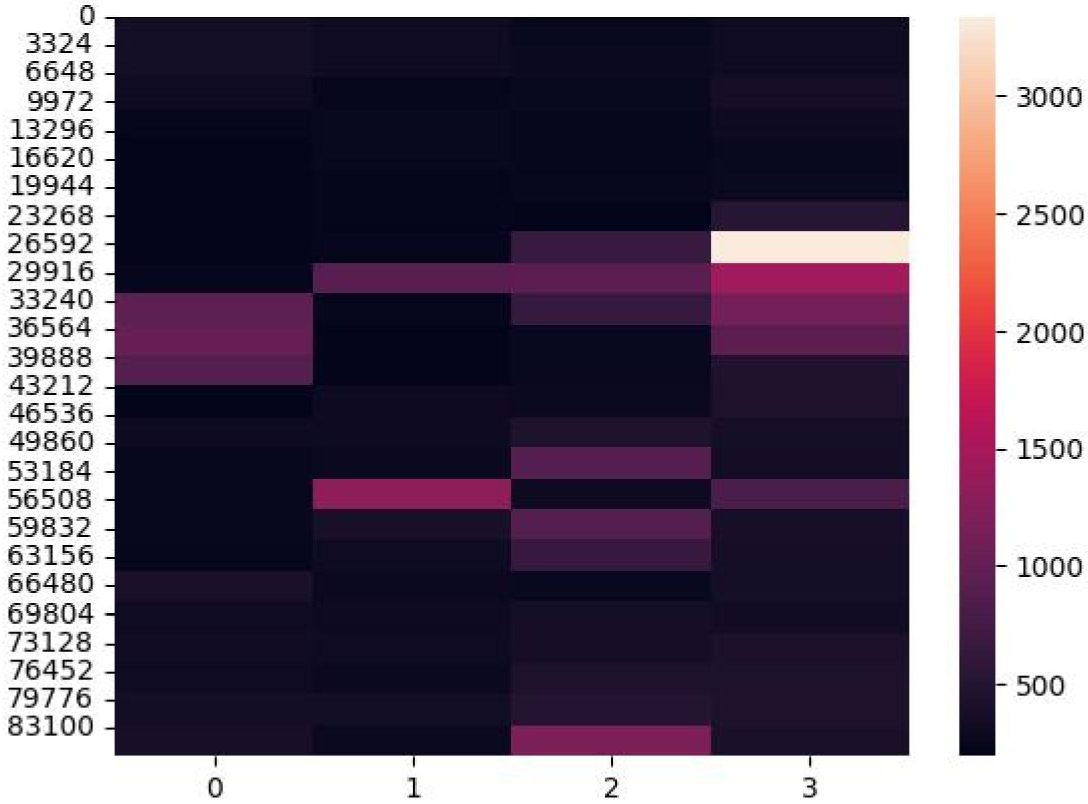


Fig 1b) - plot all days (1h resolution)

This figure displays the total electricity consumption for multiple days, aggregated to an hourly resolution. This smoother representation helps in observing general daily and weekly consumption trends and larger shifts in energy demand. For example, a consistent rise in power during the evening hours (e.g., 6 PM - 9 PM) would typically indicate increased household activity like cooking, lighting use, and entertainment systems being active.

So same data but averaged over each hour:

- **Clear daily patterns emerge** - low at night, higher during day
- **Morning peaks around 7-9 AM**: people getting ready (coffee makers, hair dryers, lights)
- **Evening peaks around 6-9 PM**: cooking dinner, TV, lighting
- **Overnight valleys**: only essential appliances running (fridge, router, alarm clocks)
- **Weekend vs weekday differences**: Different wake-up times and activity patterns

3.2 Visualizing distributions of power profile

While time-series plots provide insight into when energy is used, distribution plots (like histograms) help us understand how much energy is consumed and how frequently.

When plotting a histogram of power values:

- We often see that **most energy consumption readings** are **low**, with a long tail extending toward higher power usage.
- **Peaks in the histogram indicate recurring**, stable power levels likely caused by **specific appliances** running at regular intervals (like a refrigerator or heater cycling on and off).

Using a **logarithmic scale on the y-axis** enhances the **visibility of smaller peaks** that might be hidden on a linear scale, especially at higher energy values where counts are low but still meaningful.

Histograms also **highlight any anomalies or unusual values** — for example, unexpected spikes or constant high usage that could indicate faulty devices or errors in the data.

Before plotting, it's essential to clean the data by converting non-numeric entries (like 'N/A' or '-') to NaN and removing them to ensure an accurate distribution.

P_smx1

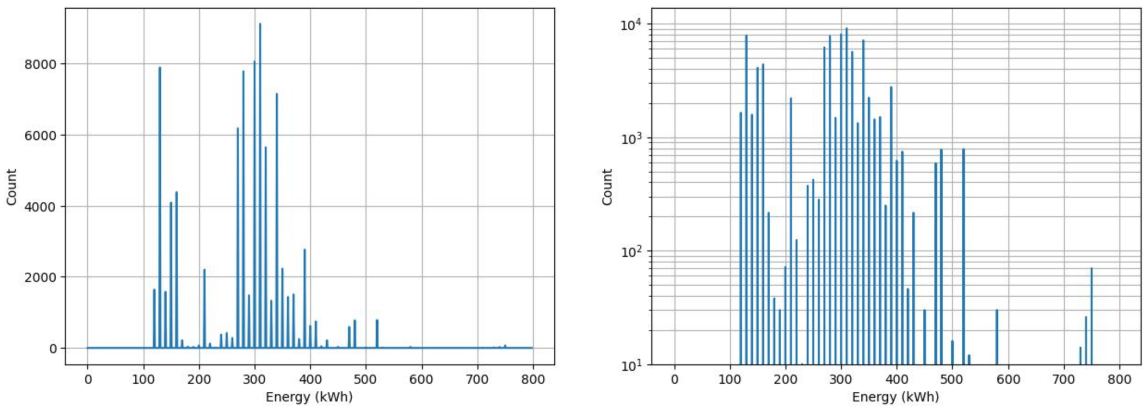


Fig 2.1 a) - P_smx1

- **Dominant peak around 100-200W**: This household has a consistent baseload
Likely includes: refrigerator running ~150W, router ~20W, cable box ~30W, standby electronics ~50W
- **Secondary peaks at 400-600W**: Regular use of medium-power appliances
Could be: washing machine cycles, dishwasher, or electric water heater cycling
- **Small peak at 1500W+**: Occasional high-power appliance use
Probably: hair dryer, electric kettle, or microwave during meal prep
- **Log scale shows**: Even rare high-power events (like briefly using a space heater) are visible

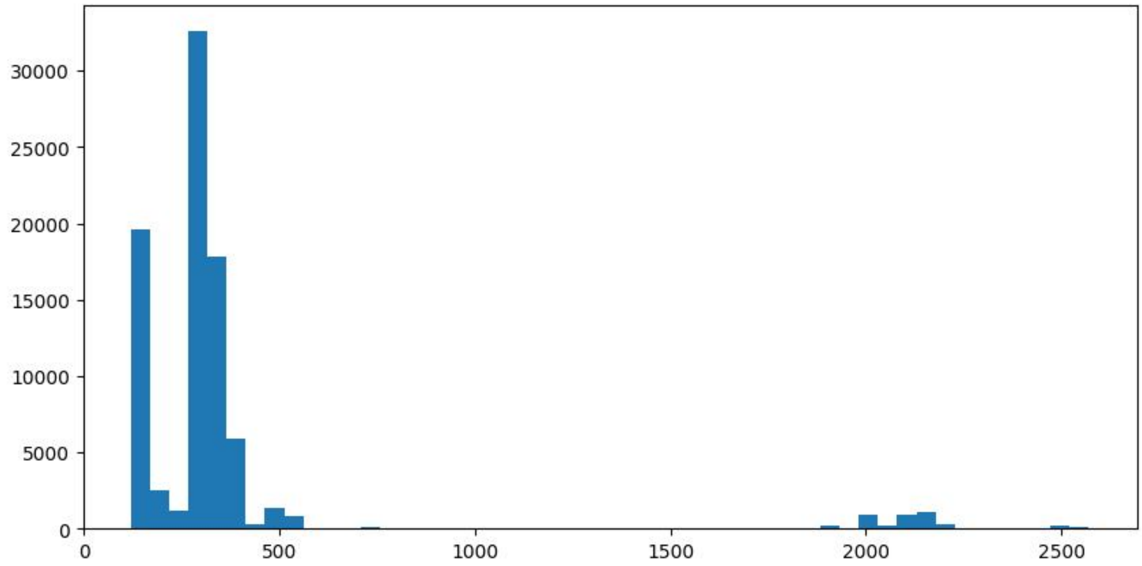


Fig 2.1 b) - P_smx1

- **Sharp, tall peak around 150W**: This household spends most time at this power level

Suggests efficient appliances and good energy habits
The refrigerator's compressor cycling creates this consistent power draw

- **Gradual tail to higher powers**: Less frequent but normal usage of larger appliances
- **Narrow peak suggests**: Predictable daily routine with consistent appliance usage

P_smx2

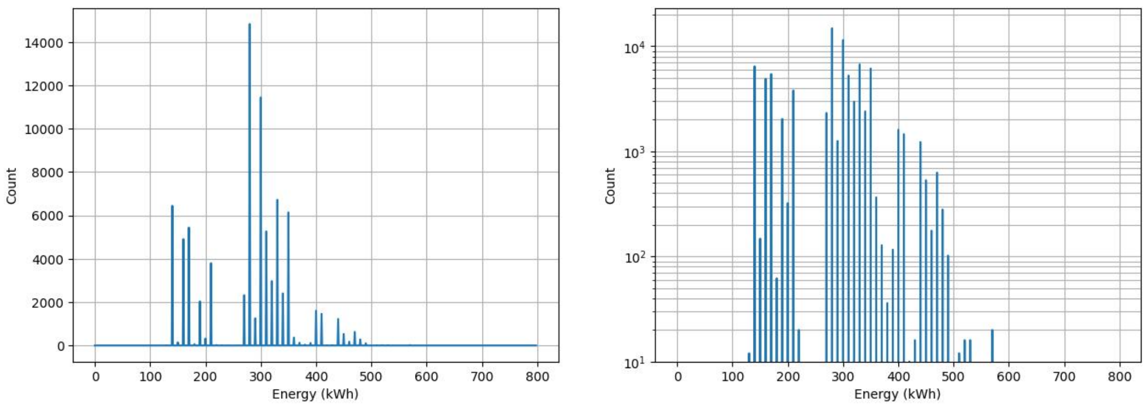


Fig 2.2 a) - P_smx2

- **Multiple distinct peaks**: This day had more varied appliance usage

Peak 1 (~80W): Baseload period, possibly during sleeping hours
Peak 2 (~250W): Standard daytime consumption with lights and electronics
Peak 3 (~800W): Cooking or laundry activities

- **Higher variability**: Suggests a more active day with different appliances used
- **Broader distribution**: Less routine, more spontaneous appliance usage

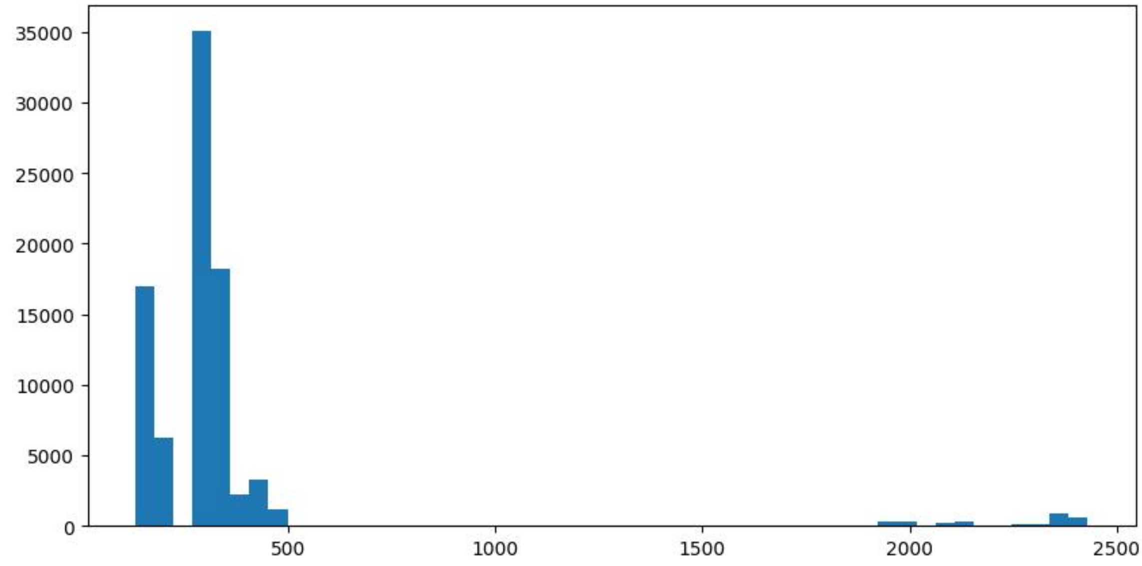


Fig 2.2 b) - P_smx2

- **Broader, flatter peak**: More varied energy usage throughout the day

Could indicate: weekend day with flexible schedule
Or: multiple family members with different routines

- **Less defined single peak**: Suggests less predictable appliance usage patterns
- **Wide base**: Significant time spent at various power levels, indicating active household

P_smx3

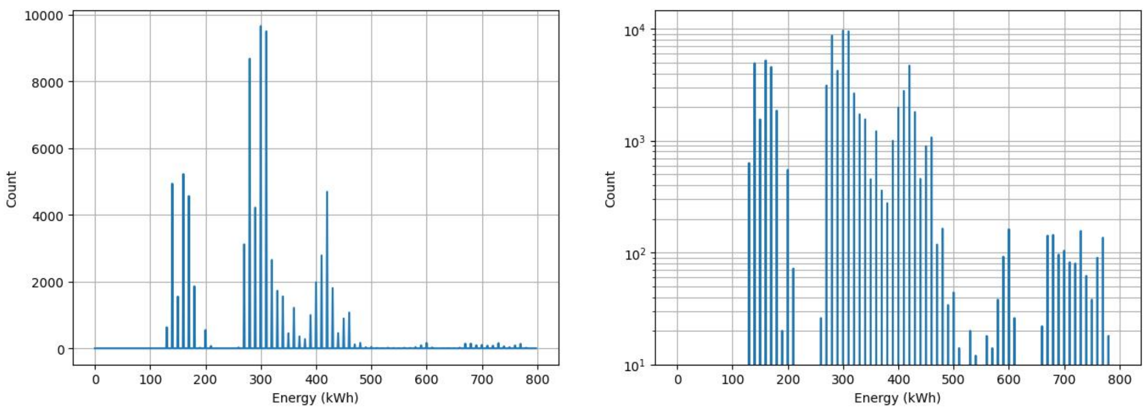


Fig 2.3 a) - P_smx3

- **Very high peak at extreme right (4000W+)**: Major appliance usage

Likely: electric oven for baking/roasting, electric car charging, or multiple high-power appliances simultaneously
Could be: holiday cooking with oven + stovetop + multiple devices

- **Multiple intermediate peaks**: Complex appliance usage pattern

Suggests: busy household day with cooking, cleaning, and entertainment

- **Extreme values visible**: Log scale reveals even brief, intense power events

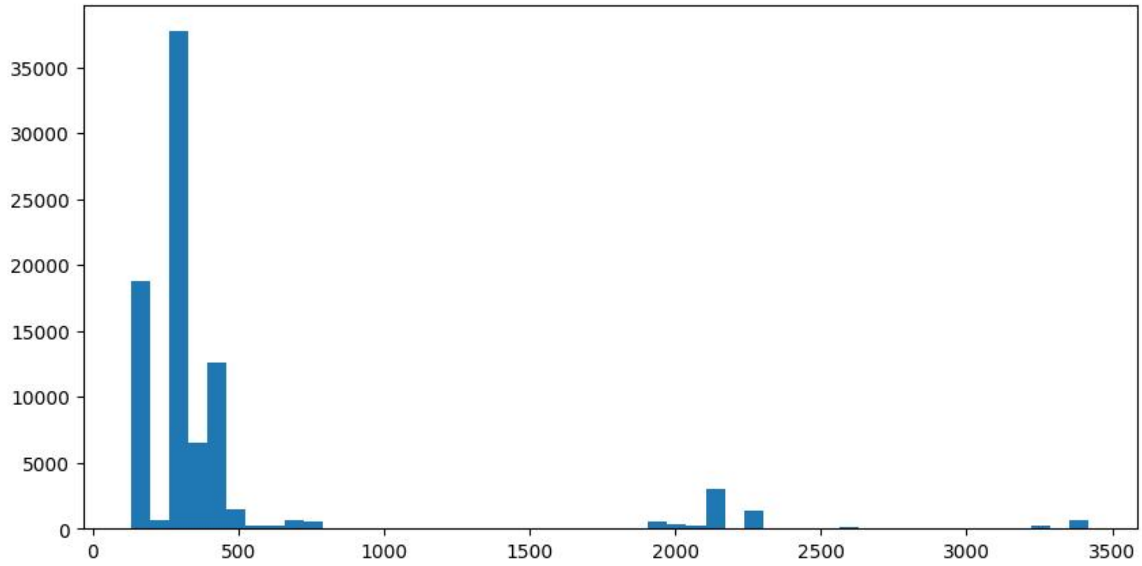


Fig 2.3 b) - P_smx3

- **Multiple smaller peaks instead of one dominant peak**: Varied activity day

Peak 1: Baseload during quiet periods
Peak 2: Normal daytime activities
Peak 3: Cooking/entertainment period
Peak 4: High-activity period (cooking dinner with multiple appliances)

- **No single dominant pattern**: Suggests flexible schedule or special occasion day

P_smx4

The result of this error is excluding day no.4 from the study. When encountering a power outage, the first line of code rewrites itself. Even after erasing and running the code, the error still remains.

Day 4 data was corrupted or had measurement errors:

- Power outage: Meter lost power and data logging failed
- Communication error: Smart meter couldn't transmit data properly
- Meter malfunction: Hardware issue causing invalid readings
- Data corruption: File transfer or storage problem
- This is normal: Smart meters occasionally have data gaps due to technical issues

Vasile Mihaela Alina
131 IEC

P_smx5

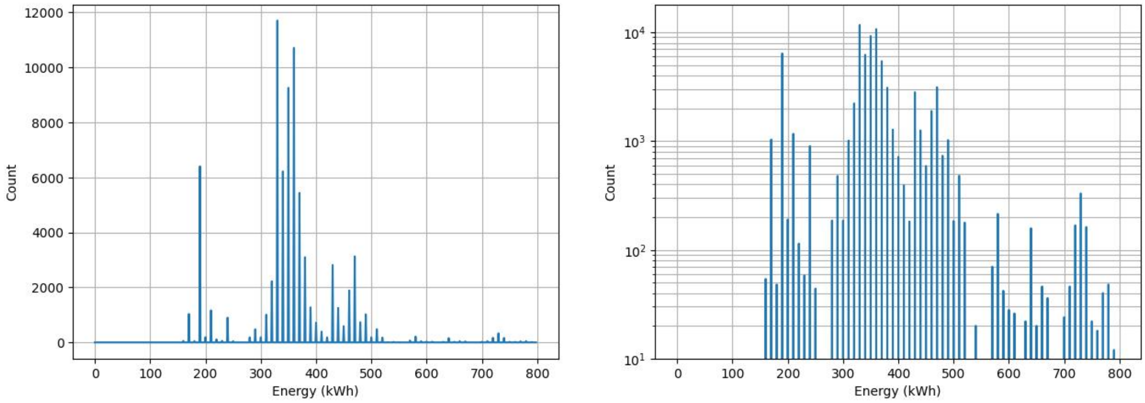


Fig 2.4 a) - P_smx5

- **Cluster of very low power readings**: Extended period with minimal usage

Could indicate: residents were away for part of the day
Only "vampire loads" running: router, cable box, alarm clock, refrigerator in efficient mode

- **Typical baseload peak**: Normal standby power consumption when present
- **Moderate high-power usage**: Normal appliance usage when residents returned

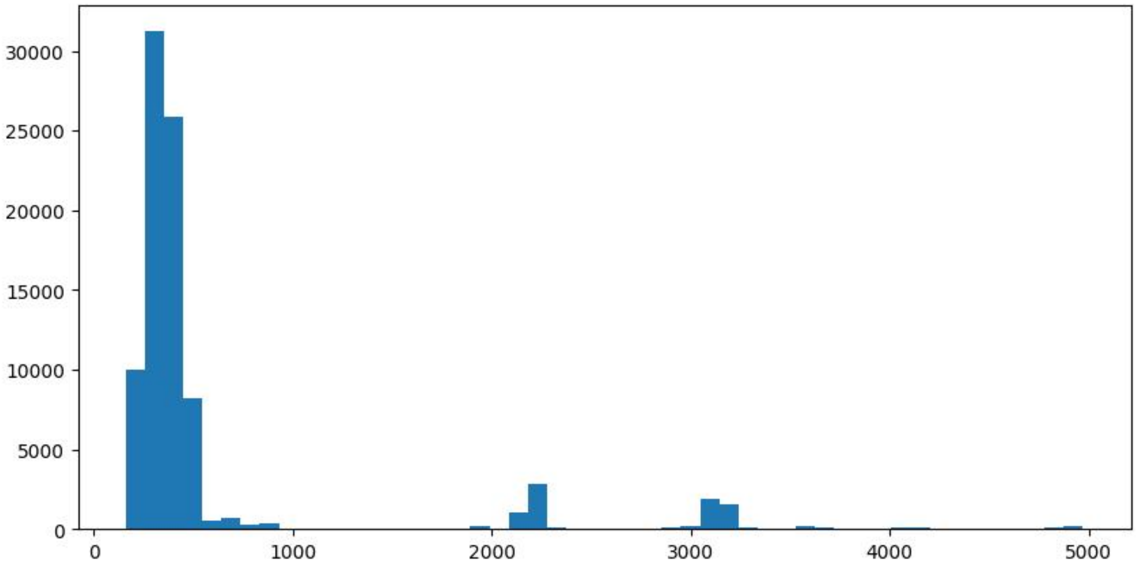


Fig 2.4 b) - P_smx5

- **Clean, defined peak**: Predictable usage pattern

Suggests: regular weekday routine
Consistent appliance cycling (refrigerator, heating/cooling system

- **Symmetrical distribution**: Balanced energy usage throughout active periods
- **Sharp cutoff at low end**: Clear distinction between "nobody home" and "normal activity" power levels

Generally, in these graphs we see:

For a)

- **Large peak at low power** (~50-200W): Baseload from always-on devices
Refrigerator compressor, WiFi router, cable box, LED lights

- **Smaller peaks at medium power** (~500-1000W): Common appliances
Dishwasher, washing machine, microwave

- **Tiny peaks at high power** (~2000W+): Occasional high-power use
Electric kettle, hair dryer, space heater, oven

For b)

- **Dominant single peak**: Suggests one major appliance runs frequently
Could be a heat pump cycling on/off regularly

- **Multiple peaks**: Different appliances with distinct power signatures
Each "hump" represents a different device's typical power draw

- **Wide, flat distribution**: Highly variable usage throughout the day

3.2 Multi-dimensional visualizations

To explore **longer periods** and more **complex patterns**, simple line graphs become less practical. Instead, **heatmaps** provide a powerful way to visualize smart meter data as a **two-dimensional image** where color encodes power values.

The key idea:

- The **x-axis** represents **days** (or dates) in chronological order.
- The **y-axis** represents **time of day** (from midnight to midnight).

Each cell (or pixel) in the heatmap shows the energy consumption at a specific day and time, with **colors** indicating the magnitude of power used.

This two-dimensional layout allows us to:

- **Detect** daily and weekly **consumption patterns visually**.
- **Spot unusual** days or times with atypical energy use.
- **Identify devices** that switch on and off **regularly**, which show up as recurring patterns.

By creating separate heatmaps with data averaged over 1 hour and 15 minutes, we can see how the level of temporal granularity affects the visibility of these patterns. The 15-minute heatmap reveals more fine-grained behavior, while the 1-hour heatmap smooths the data for broader trend identification.

3.3.1 Heatmap based on the values of the 5 files avereged over 1h and 15min

Using the provided averaging functions, we can prepare matrices representing power profiles for **multiple days**, stacked vertically. Applying **Seaborn's heatmap function** to these matrices gives us clear and informative visuals.

The heatmaps typically reveal:

- **Higher overall activity** during colder months due to heating requirements.
- **Regular device cycles**, such as heat pumps switching on and off in consistent intervals — often producing distinctive checkerboard patterns.

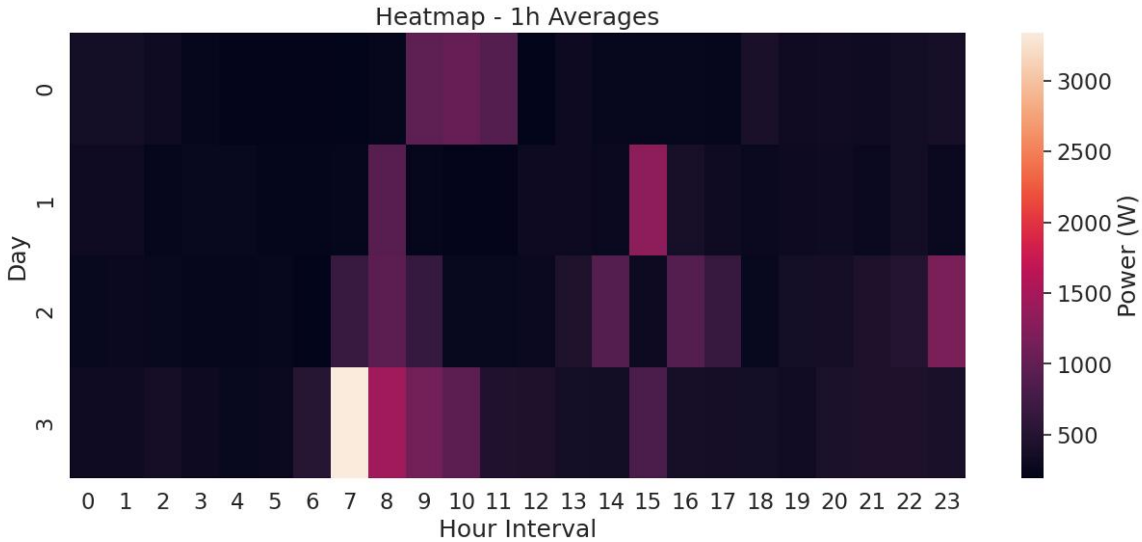


Fig 3.1 - heatmap 1h

In this heatmap, a consistently bright vertical band around 7-9 AM across multiple days (rows) would indicate a recurring morning routine with high energy use, like preparing breakfast or getting ready for work/school.

- **Bright vertical bands at 7-9 AM**: Morning routine across multiple days
Coffee maker, toaster, shower (electric water heater), hair dryer
- **Bright bands at 6-9 PM**: Evening cooking and entertainment
Oven, stove, TV, multiple lights, dishwasher
- **Dark overnight periods**: Minimal activity, only baseload
- **Weekend patterns**: Later morning activity, different evening patterns

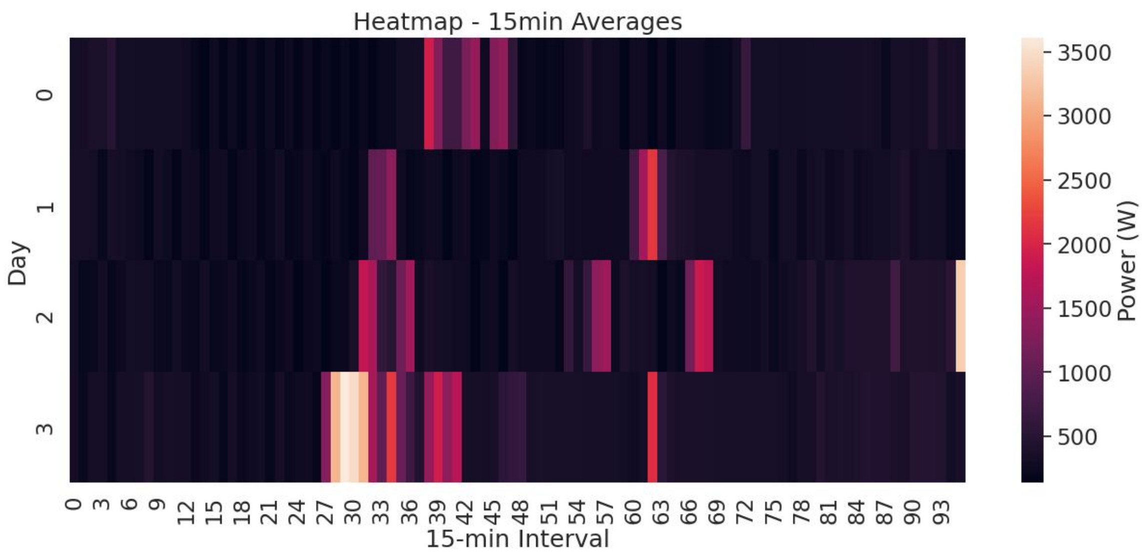


Fig 3.2 - heatmap 15min

In this heatmap, a darker, cooler patch at night (e.g., 1 AM - 5 AM) across multiple days would suggest a period of minimal activity and lower energy consumption, with only baseload devices operating.

- **More detailed patterns**: Can see individual appliance cycles
- **Checkerboard patterns**: Heat pump or AC cycling every 15-30 minutes
- **Brief bright spots**: Short-duration, high-power appliances (microwave, kettle)

3.3.2 3d heatmap based on the values of the files avereged over 1h and 15min

For an even more **immersive view**, we can extend heatmaps into the third dimension by plotting the energy data as a **3D surface**:

- The x-axis still represents **time within each day**.
- The y-axis represents s**equential days**.
- The z-axis shows power **consumption values**.

This 3D perspective accentuates peaks and troughs, making consumption spikes easier to spot and interpret.

Using **matplotlib's 3D plotting tools**, we can create interactive or static 3D visualizations that complement the 2D heatmaps. These visuals often help in presentations or deeper exploratory analysis by providing an **intuitive sense** of how consumption evolves over time.

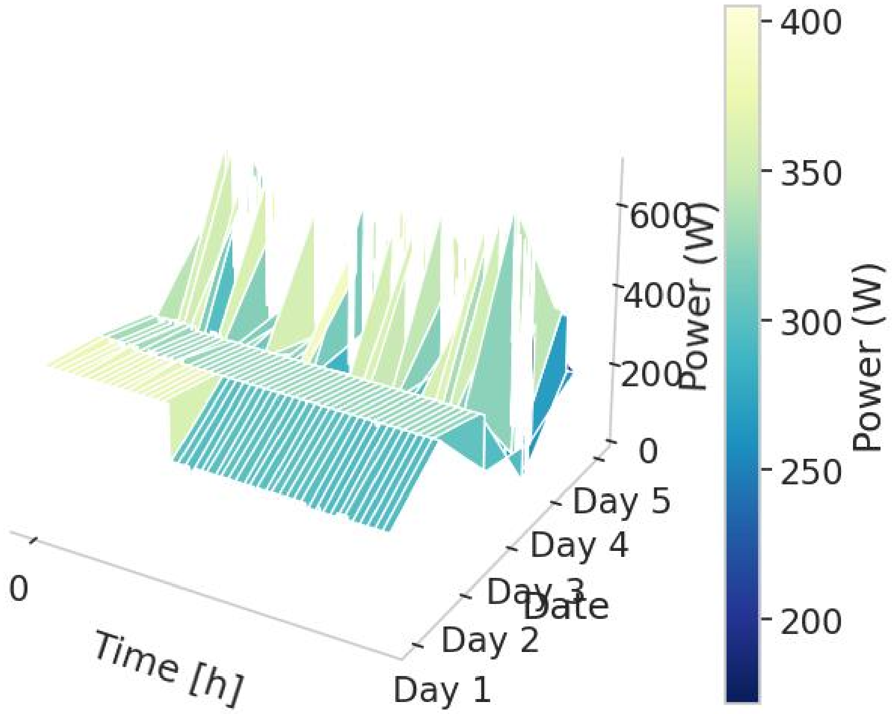


Fig 4 - 3d heatmap

In this 3D graph, a **prominent, tall, and yellowish peak** consistently rising from the baseline for each day (e.g., around 18-20 hours on the "Time [h]" axis across "Day 1" to "Day 5" on the "Date" axis) would **indicate a regular, high-energy consumption event**, such as a household preparing dinner and using multiple large appliances simultaneously. Conversely, **the flat, dark blue sections**, particularly overnight, show **periods of very low power draw**, reflecting times when only essential background appliances are operating.

- **Tall peaks at dinner time:** Multiple appliances used simultaneously
Oven + stovetop + dishwasher + lights + TV
- **Consistent overnight valleys**: Stable baseload consumption
- **Weekend "mountains"**: Different usage patterns when people are home all day

4. Pre-processing smart meter data

In many data analysis workflows, especially when dealing with real-world measurements like smart meter data, **pre-processing is a critical first step**. The quality of your analysis largely depends on the quality of your **input data**. Smart meter readings can contain **noise, errors, or inconsistencies** such as missing values, unrealistic spikes, or sensor glitches. Properly preparing this data ensures that subsequent analysis produces reliable and meaningful insights.

In this section, we present several best practices and techniques for **pre-processing** smart meter data. These steps will be foundational for the analyses and tasks covered in the later sections.

4.1 Outlier Detection with Hampel Filter

One of the first pre-processing tasks is to **identify and handle outliers** — **data points that deviate significantly** from typical values and are likely due to **errors** or unusual but unrealistic conditions.

The **Hampel filter** is an effective method for **detecting such outliers** in time series data. It works by **comparing each data point** to the distribution of its neighboring points within a sliding window. This localized comparison allows for the **identification of measurements** that stray too far from their surroundings.

The filter depends on two main parameters:

- **Window size**: This controls how many neighboring data points are considered when computing the local statistics.
- **n_sigmas**: This determines the **threshold** for outlier detection — how many standard deviations (scaled using the median absolute deviation) a point must be away from the median to be considered an outlier.

When an **outlier is detected**, its **value is replaced** with the **median of its neighboring values**, effectively smoothing the data without distorting typical readings.

Day 1

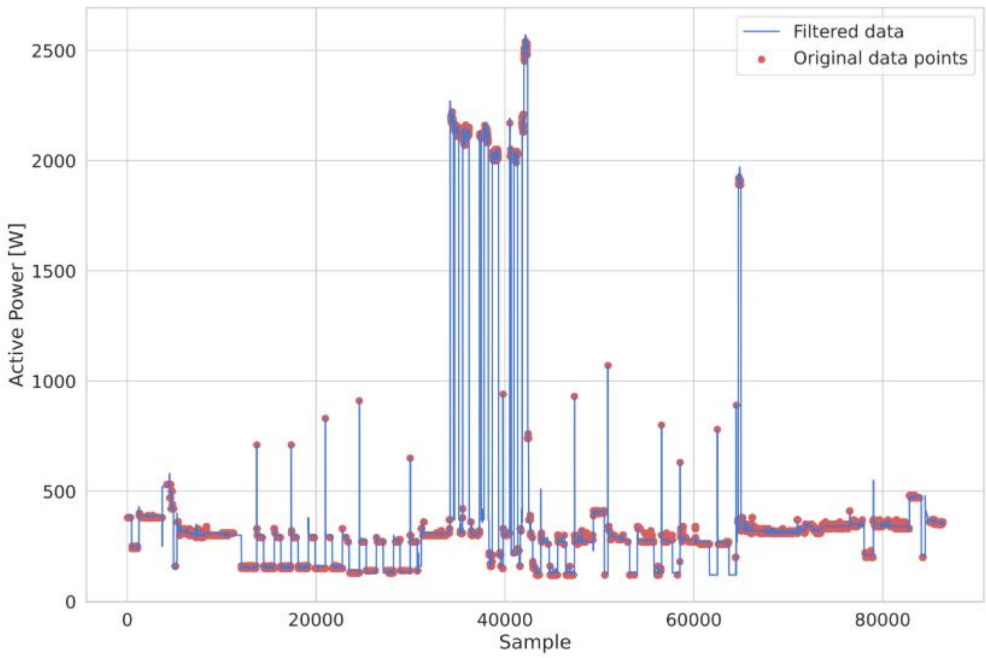


Fig 5.1 - Hampel day 1

- **Sharp isolated spikes removed:** Likely measurement glitches

Could be: brief electrical interference, meter communication error
Example: A 3000W spike lasting 1 second (physically impossible for most appliances)

- **Preserved legitimate peaks**: Real appliance usage maintained

Hair dryer spike at 7 AM: genuine 1800W for 5 minutes
Oven usage at 6 PM: sustained 2500W for 45 minutes

- **Smoother orange line**: More reliable for detecting actual appliance patterns

- **Noise reduction**: Random fluctuations smoothed while maintaining real trends

Day 2

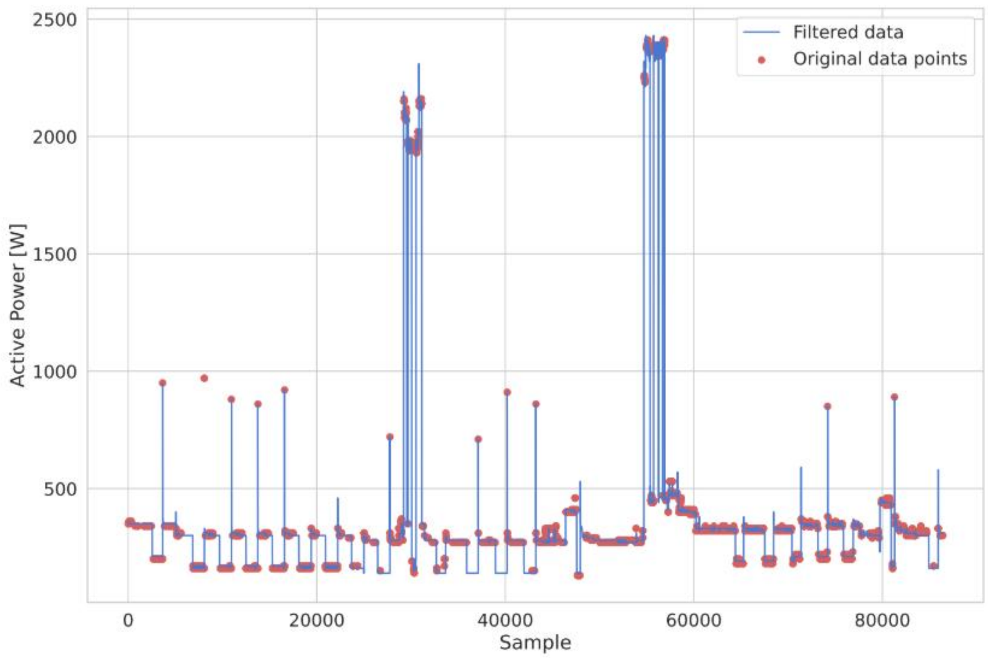


Fig 5.2 - Hampel day 2

- **Jagged sections smoothed**: Temporary electrical interference corrected

Could be: neighbor's heavy machinery causing voltage fluctuations
Or: loose electrical connection causing erratic readings

- **Major usage patterns preserved:** Cooking and entertainment peaks intact
- **Better data quality**: Filtered data shows clear appliance on/off cycles
- **Reduced false alarms**: Anomaly detection will now focus on real issues, not measurement errors

Day 3

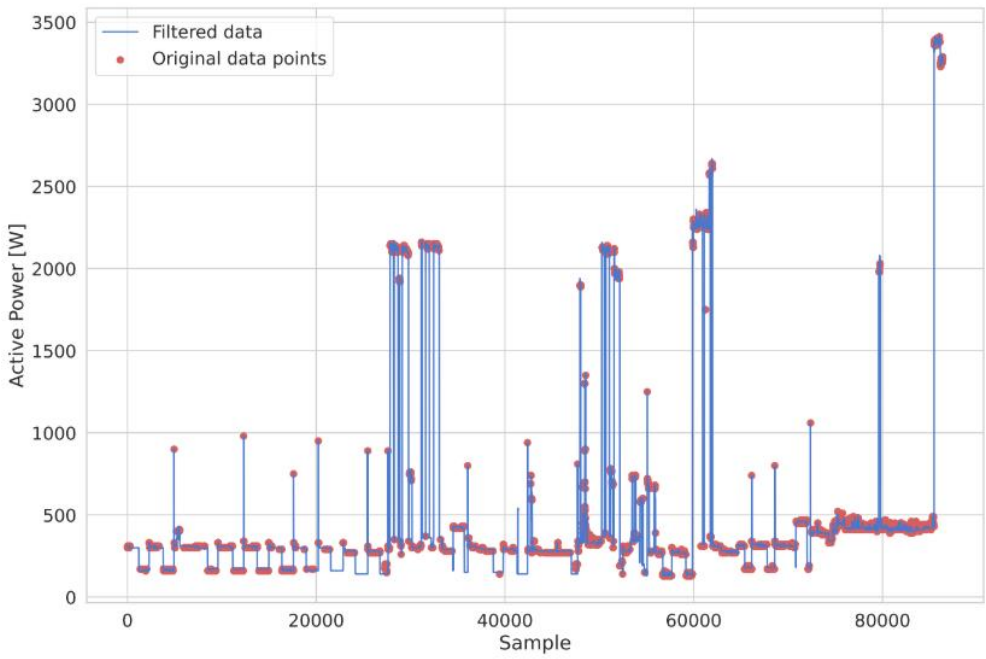


Fig 5.3 - Hampel day 3

- **Single unusually high outlier removed**: Extreme reading that doesn't match surrounding pattern

Could be: power tool briefly drawing excessive current due to motor seizure
Or: meter registering phantom reading during electrical storm

- **Normal high-power usage kept**: Legitimate oven/stovetop usage during dinner preparation
- **Pattern continuity**: Appliance cycling patterns now clearly visible
- **Enhanced analysis reliability**: Subsequent analysis won't be skewed by measurement artifacts

Day 4 was excluded due to recurring errors

Day 5

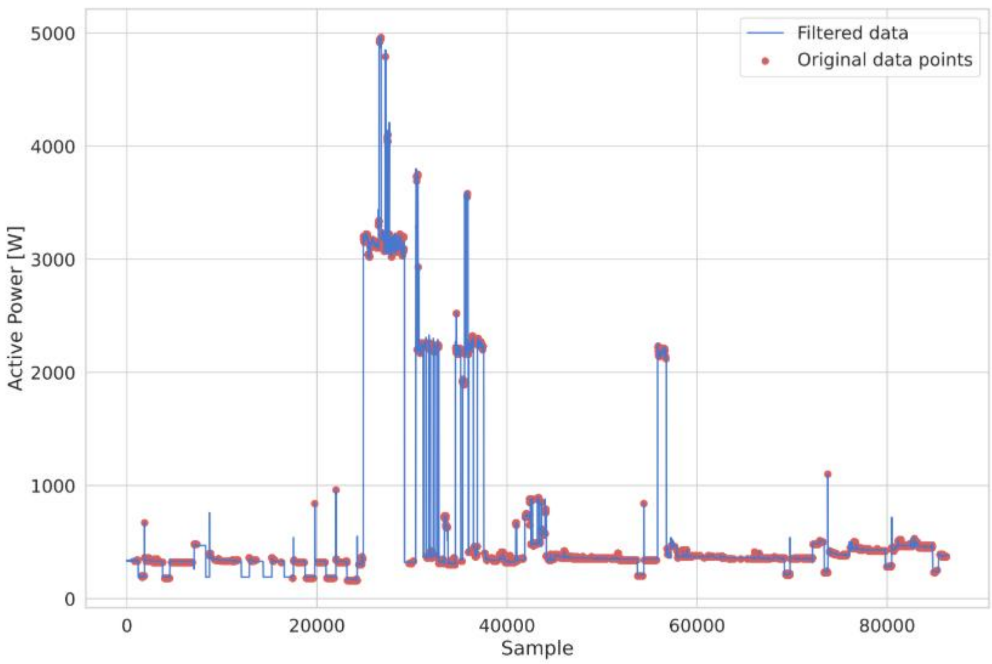


Fig 5.4 - Hampel day 5

- **Minimal corrections needed**: Suggests good meter performance this day
- **Clean appliance signatures**: Clear on/off patterns for refrigerator, heating system
- **Preserved usage spikes**: Legitimate high-power events (microwave, kettle) maintained
- **Ready for analysis**: High-quality data suitable for appliance disaggregation and behavior analysis

Overall, these figures demonstrate how "noisy" or unusual spikes in the electricity data (outliers) are detected and smoothed out or removed. This makes the data cleaner and more reliable for analysis, preventing misleading results from faulty readings or short, intense usage spikes.

- **Sharp spikes removed**: Likely measurement errors or very brief anomalies

Faulty meter reading, power surge, motor starting issue

- **Preserved normal patterns**: Legitimate power variations kept intact

4.2 Simple baseload estimation

The baseload is the minimum continuous power consumption in a household or building, usually attributable to **appliances and systems** that run regardless of active usage, such as **refrigerators, heating/cooling systems, or standby electronics**.

Estimating baseload is valuable for multiple reasons:

- **Identifying** the baseline energy requirement for a building.
- **Detecting abnormal spikes or drops** which may indicate faults or inefficiencies.
- **Supporting** energy-saving initiatives by highlighting areas for potential improvement.

A straightforward method to estimate baseload involves:

- **Sorting all power consumption** observations in ascending order.
- **Selecting a defined lowest percentage** (commonly 20%) of these values for analysis.
- **Calculating the mean** and **standard deviation** of this subset.
- **Defining the baseload threshold** as the **first value exceeding the mean plus three times the standard deviation.**

Day 1

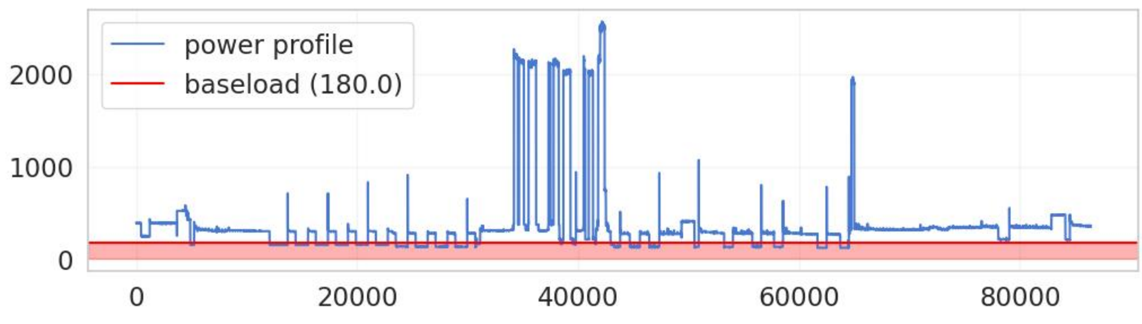


Fig 6.1 - Baseload day 1

- **Steady red line around 80W**: Always-on devices consuming power 24/7

Refrigerator compressor cycling: ~40W average
WiFi router: ~15W continuous
Cable/internet modem: ~20W continuous
LED clock displays, smoke detectors: ~5W total

- **Blue line consistently above red**: All additional appliance usage clearly visible
- **Nighttime gaps narrow**: Minimal extra usage during sleeping hours (just occasional bathroom light)
- **Daytime gaps wide**: Active appliance usage throughout day

Day 2

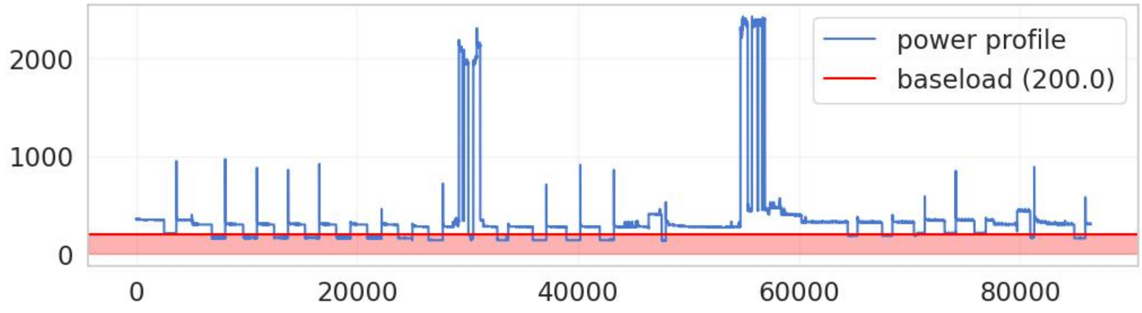


Fig 6.2 - Baseload day 2

- **Higher baseload (~120W)**: More devices in standby mode

Could include: TV in standby (~15W), computer on sleep mode (~25W)
Gaming console in standby (~20W), additional chargers plugged in (~20W)

- **Indicates lifestyle**: Household with more electronic devices
- **Energy efficiency opportunity**: Higher standby consumption suggests potential savings
- **Usage patterns**: Wide gaps show active appliance usage periods

Day 3

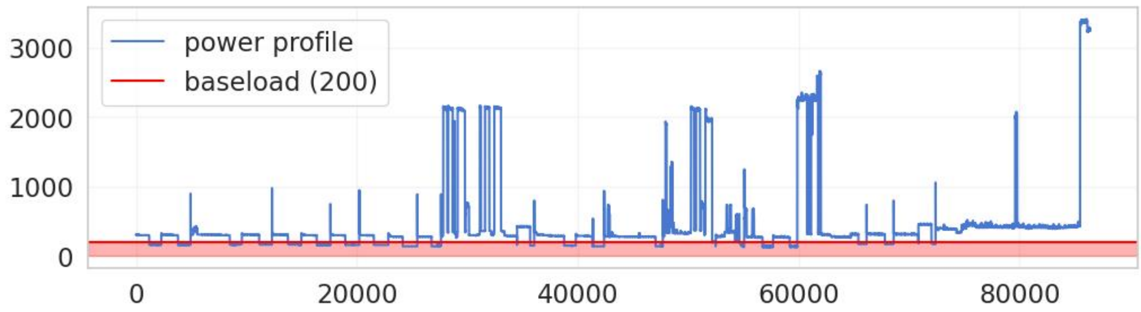


Fig 6.3 - Baseload day 3

- **Stable baseload with slight variations**: Normal fluctuation as refrigerator cycles
- **Clear separation**: Easy to distinguish between background and active usage
- **Appliance signatures visible**: Can identify when specific devices turn on/off
- **Weekend pattern**: Different timing of usage compared to weekday patterns

Vasile Mihaela Alina
131 IEC

Day 4

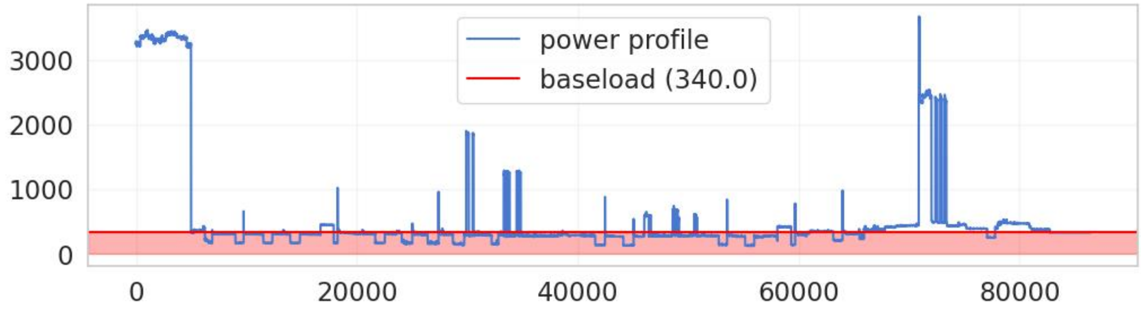


Fig 6.4 - Baseload day 4

- **Irregular baseload line**: Data corruption affects reliability of baseline estimation
- **Jagged red line**: Algorithm struggling with missing or corrupted data points
- **Unreliable analysis**: This day's baseload estimate shouldn't be used for decisions
- **Recovery visible**: Where data quality improves, normal patterns resume

Day 5

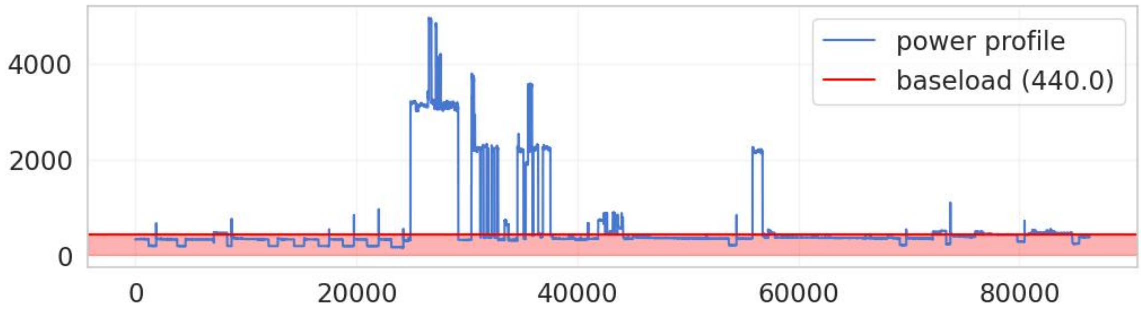


Fig 6.5 - Baseload day 5

- **Consistent 90W baseload**: Well-defined continuous power consumption
- **Efficient household**: Moderate standby consumption suggests energy-conscious habits
- **Clear usage patterns**: Distinct periods of additional appliance usage
- **Appliance health:** Steady baseload suggests all always-on devices functioning normally

These figures show the **home's total electricity usage throughout a day** (blue line) and **identify the "baseload"** (red line and shaded area). This "baseload" represents the **minimum, continuous power the home consumes** even when major appliances aren't actively running (e.g., fridge, router, standby devices).

- **Steady red line around 80-150W**: Always-on devices

Refrigerator, router, cable box, alarm clocks, standby electronics

- **Blue line always above red**: All additional appliance usage
- **Higher baseload**: More devices plugged in or less efficient appliances
- **Stable baseload**: Well-functioning household electrical system

4.3 Enhacing small activites

Sometimes, small but frequent energy uses (like LED lights, phone chargers, or laptops) are overshadowed in raw power data by larger loads (such as heat pumps or ovens). To better analyze these **smaller loads**, we can apply a **transformation function** to the data that **compresses larger values** and relatively expands smaller ones. This reduces the skewness of the data and **highlights lower consumption** activities, which may otherwise be overlooked.

One widely effective transformation is the **square root transformation**, which **compresses high values** but **magnifies smaller ones**. After applying it, smaller energy activities become more visible and easier to analyze.

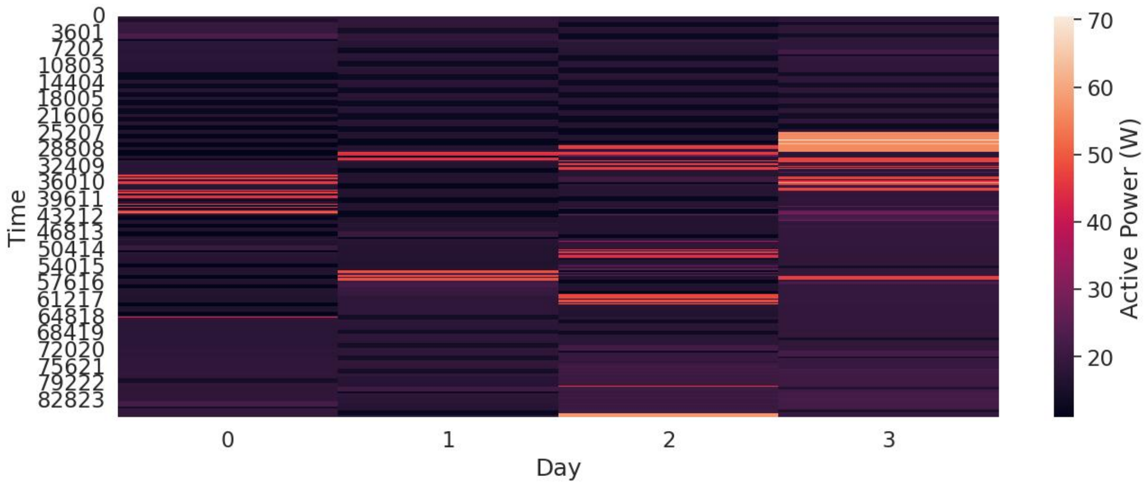


Fig 7.1 - sqrt heatmap

- **Small appliances now visible**: LED lights, phone chargers, laptop usage patterns emerge
- **Subtle daily rhythms**: Can see patterns in low-power device usage

Bedside lamp usage, computer sleep/wake cycles
Small kitchen appliances (coffee maker, toaster)

- **Enhanced sensitivity**: Previously hidden energy behaviors revealed
- **Better balance**: Large appliances don't dominate the visualization

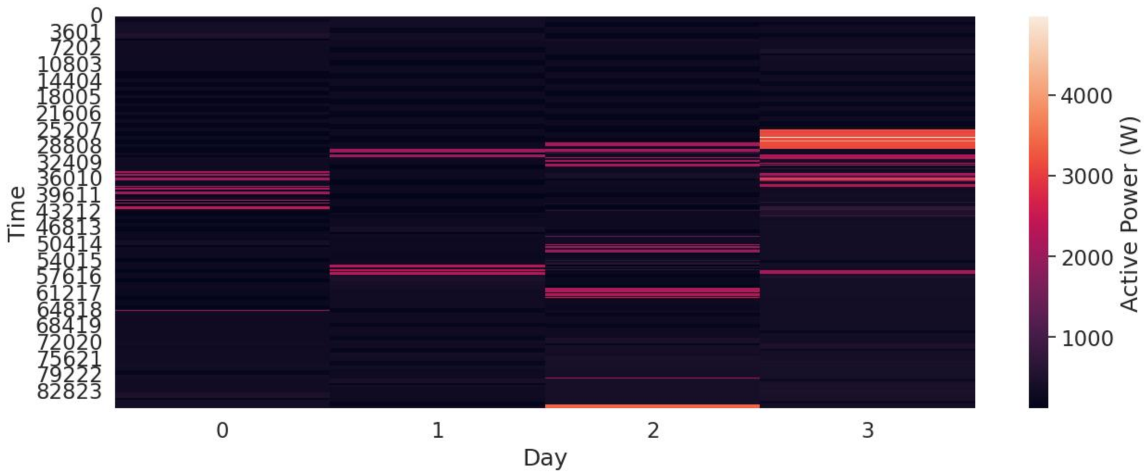


fig 7.2 - original heatmap

- **High-power appliances dominate**: Heat pump, oven, washer cycles are most visible
- **Small devices invisible**: LED lights, chargers lost in the scale
- **Major usage patterns**: Clear cooking times, heating cycles, large appliance usage
- **Limited insight**: Can't see full spectrum of household energy behavior

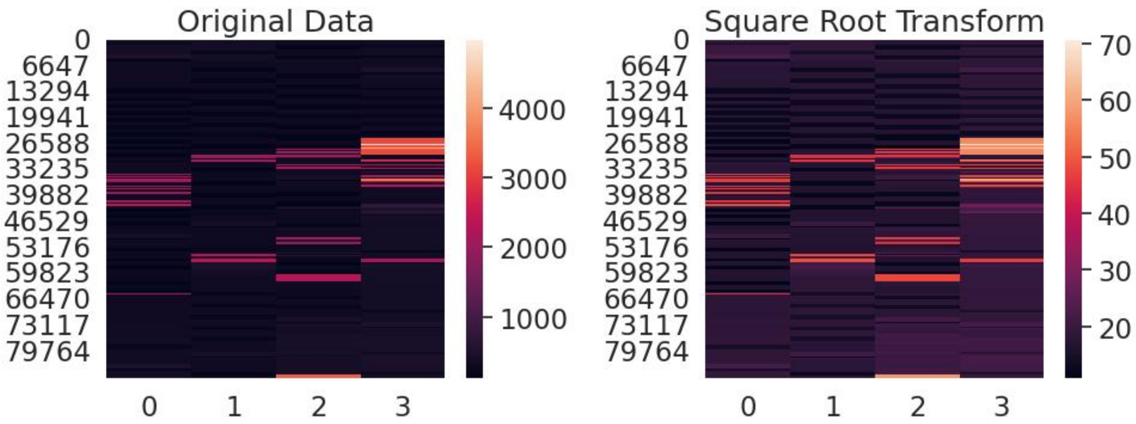


Fig 7.3 - original vs sqrt heatmap

- **Discord patterns (unusual)**: Unexpected power consumption behavior

3 AM high usage: Could be electric water heater malfunction or security system activation

Or: Someone working unusual hours, using computer and lights

- **Regular switching events**: Normal appliance cycling throughout day
- **Pattern recognition**: Algorithm identifies both normal and abnormal switching behavior
- **Maintenance insights**: Unusual patterns might indicate appliance service needs

In conclusion:

- **Before transformation**: Large appliances dominate the visual
- **After transformation**: Small appliances become more visible

LED lights, phone chargers, laptops, small kitchen appliances

So the explanations would be:

- **Morning change point (7 AM)**: Start of daily routine

Coffee maker turns on, lights activated, bathroom fan

- **Evening change point (9 PM)**: Settling down for night

TV off, lights dimmed, appliances shut down

- **Multiple rapid changes**: Cooking period

Oven, microwave, stovetop used in sequence

4.4 Detecting switching activities

Switching events in smart meter data correspond to **appliances turning on or off**, causing noticeable **jumps in power consumption**. Detecting these events can reveal appliance usage patterns or operational faults.

Using the **square-root transformed data**, we detect **these switching events** by analyzing the **second differences in the time series** — that is, the change in the change of power consumption over time.

- **A positive** second difference suggests a switch-**on** event.
- **A negative** second difference suggests a switch-**off** event.
- Values n**ear zero** indicate minor **fluctuations or noise**.

Applying a peak detection algorithm (e.g., from scipy.signal) on these second differences helps isolate significant switching operations.

Day 1

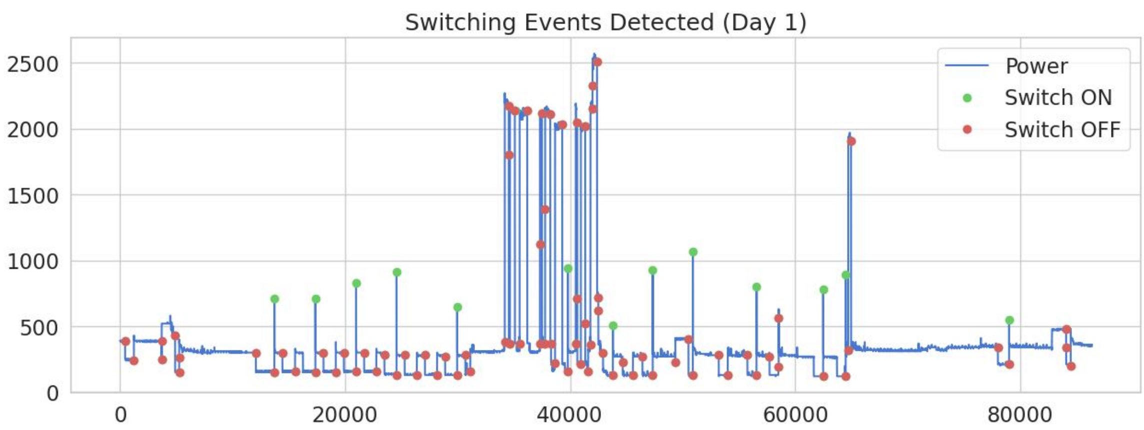


Fig 8.1 - Switching events day 1

Day 2

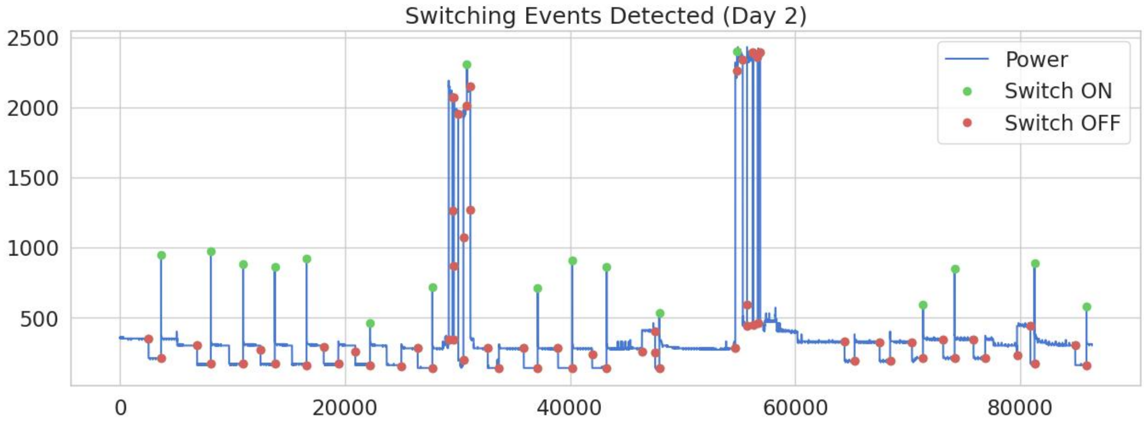


Fig 8.2 - Switching events day 2

- **Motif patterns (repeating)**: Regular, predictable appliance behavior

Refrigerator compressor cycling every 2-3 hours
Heat pump regular on/off cycles maintaining temperature

- **Consistent timing**: Shows household routine and efficient appliance operation
- **Normal operation**: Repeating patterns indicate healthy appliance function
- **Energy optimization**: Regular cycles suggest proper thermostat settings

Day 3

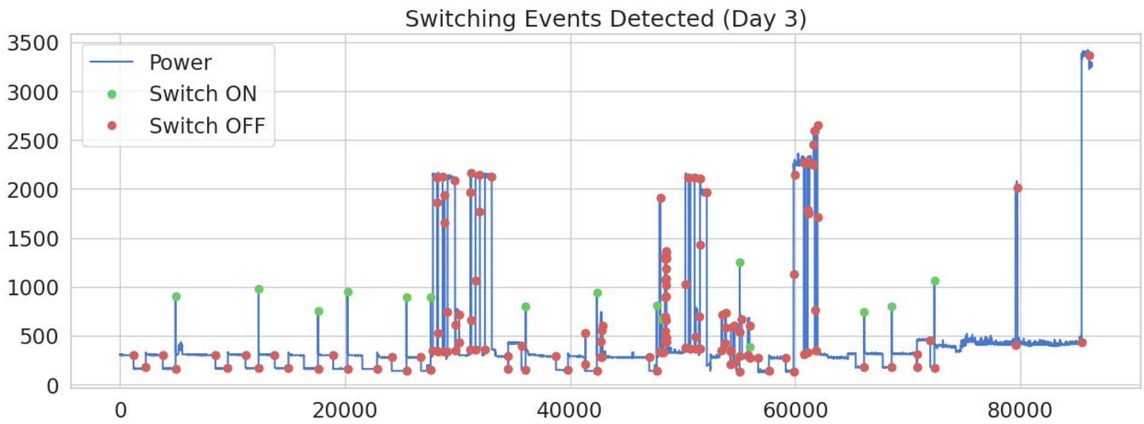


Fig 8.3 - Switching events day 3

- **Multiple motif examples**: Various repeating patterns identified

Electric vehicle charging: consistent high-power draw during off-peak hours
HVAC system: regular temperature maintenance cycles

- **Discord examples**: Unusual, non-repeating patterns

Experimental device use: 3D printer or workshop equipment with erratic power draw
Emergency situations: sump pump activation during unexpected weather

- **Pattern diversity**: Shows household with varied appliance portfolio

Day 4 was excluded due to reocurring errors

Day 5

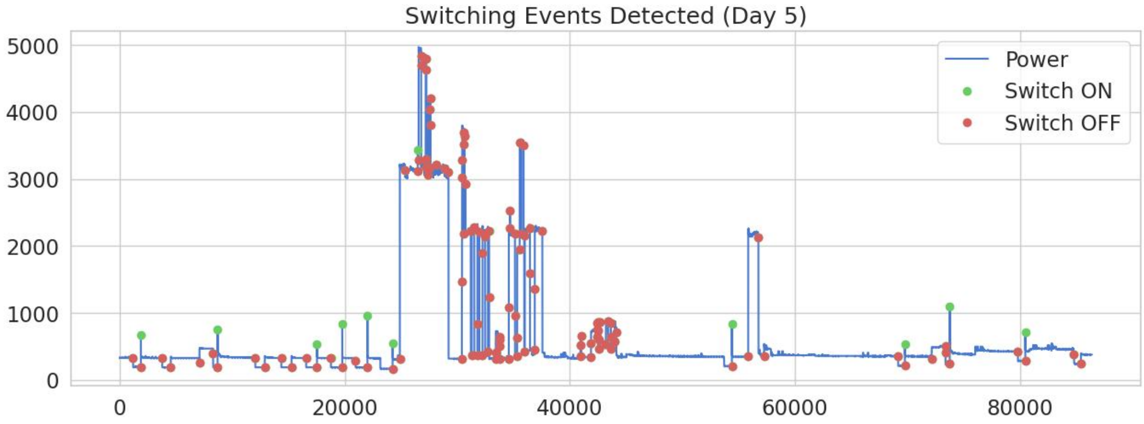


Fig 8.4 - Switching events day 5

- **Unique discord patterns**: Rare or unusual energy usage events

Long, low-power plateau: Security system test mode or medical device operation
Occasional tiny spikes: Motion sensor activations or smart home device communications

- **Infrequent patterns**: Events that don't fit normal daily routine
- **System diagnostics**: Helps identify occasional-use devices or system anomalies

Day 1

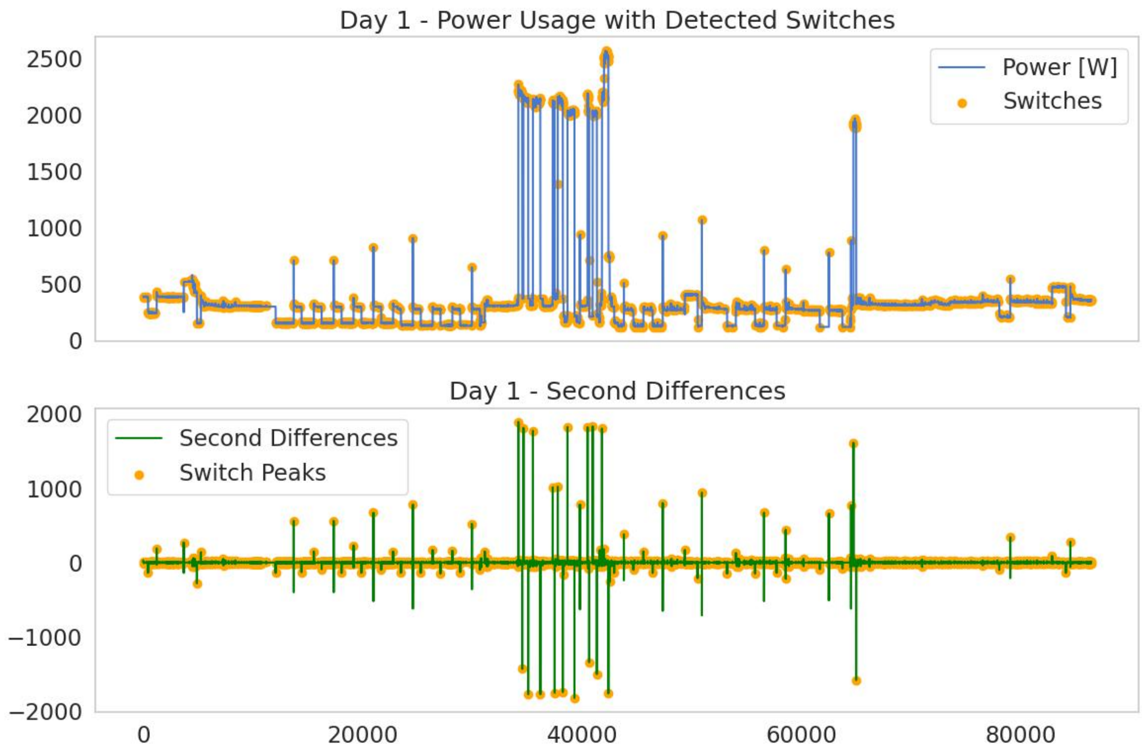


Fig 9.1 a) - Power usage + second differences - Day 1

- **Power jumps correlate with difference spikes**: When blue line jumps up at 6 PM, orange line shows sharp positive peak

This moment: HVAC system turning on as family arrives home

Mathematical confirmation: Algorithm correctly identifies the exact switching moment

- **Negative spikes**: When appliances turn off

Sharp negative peak at 10 PM: TV and entertainment system powering down for night

- **Small fluctuations**: Minor appliance activity (lights, small electronics)

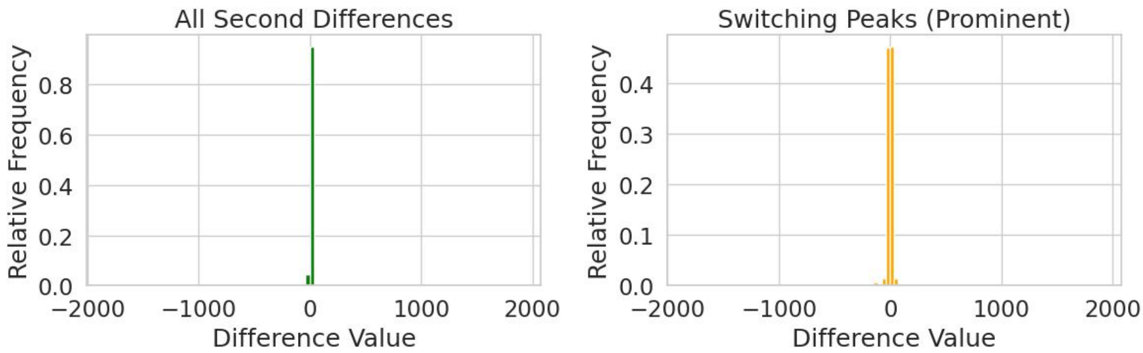


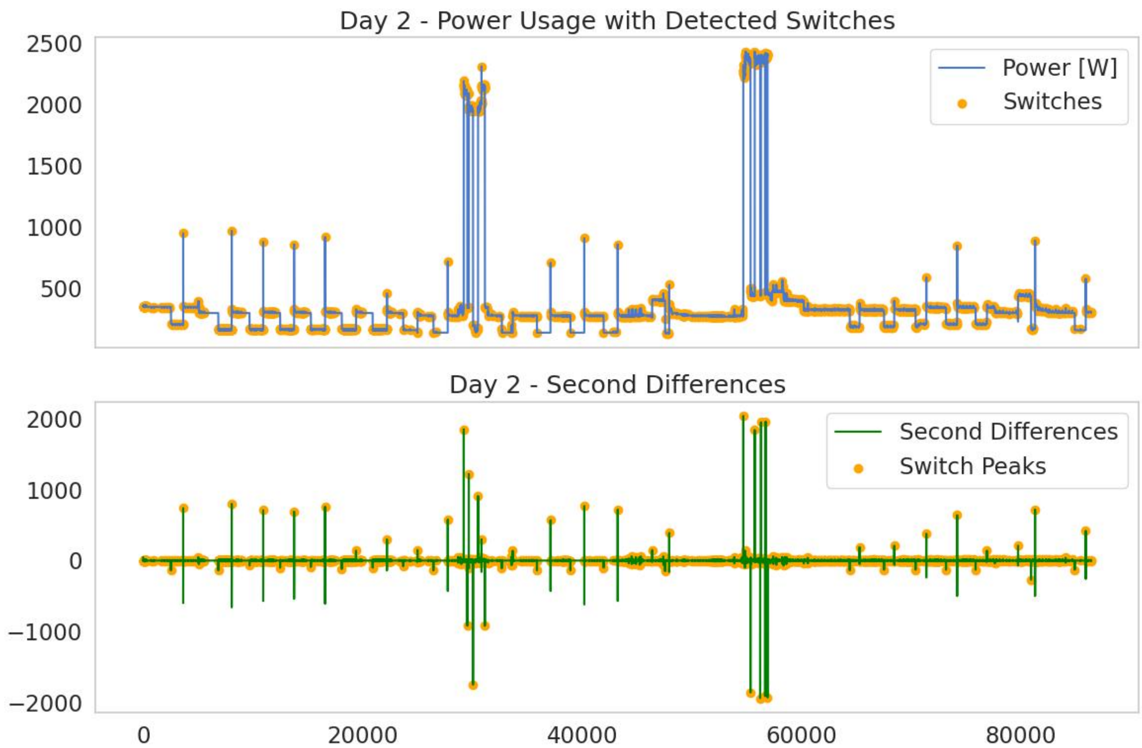
Fig 9.1 b) - Second differences + switching peacks vs diff value - Day 1

- **Cluster around +1000W**: Multiple instances of similar appliances turning on

Microwave use: exactly 1000W each time it's used
Hair dryer: consistent 1200W every morning use

- **Positive vs negative**: On events vs off events clearly separated
- **Power signatures**: Each appliance type creates consistent patterns
- **Usage frequency**: Cluster density shows how often each appliance is used

Day 2



Vasile Mihaela Alina
131 IEC

Fig 9.2 a) - Power usage + second differences - Day 2

- **Gradual decline with multiple small negatives**: Sequential device shutdown
Evening routine: TV off, then lights, then computer, then chargers unplugged
Algorithm detects each individual device powering down
- **Morning startup sequence**: Series of positive peaks as household awakens
Coffee maker (800W) → bathroom fan (50W) → hair dryer (1500W)

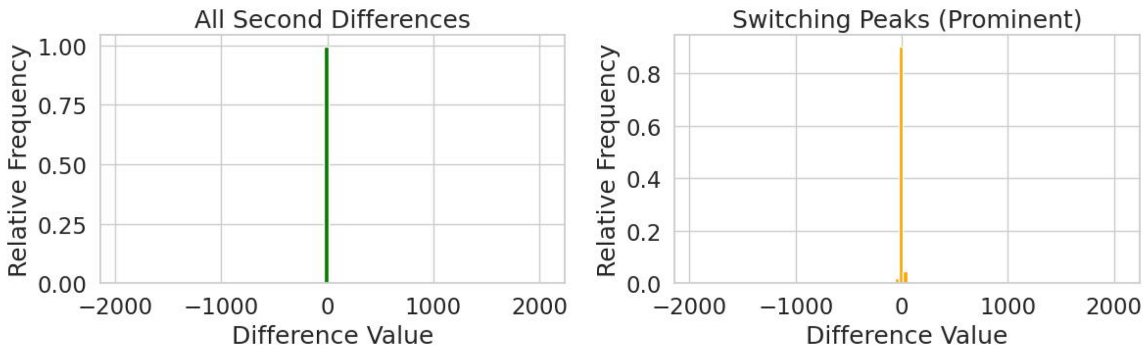


Fig 9.2 b) - Second differences + switching peacks vs diff value - Day 2

- **Consistent positive clusters**: Regular appliance usage patterns
200W cluster: Various lights being turned on throughout day
500W cluster: Mid-size appliances (toaster, electric kettle, vacuum)
- **Pattern recognition**: Same appliances used consistently
- **Energy habits**: Clusters reveal household's preferred appliances and usage timing

Day 3

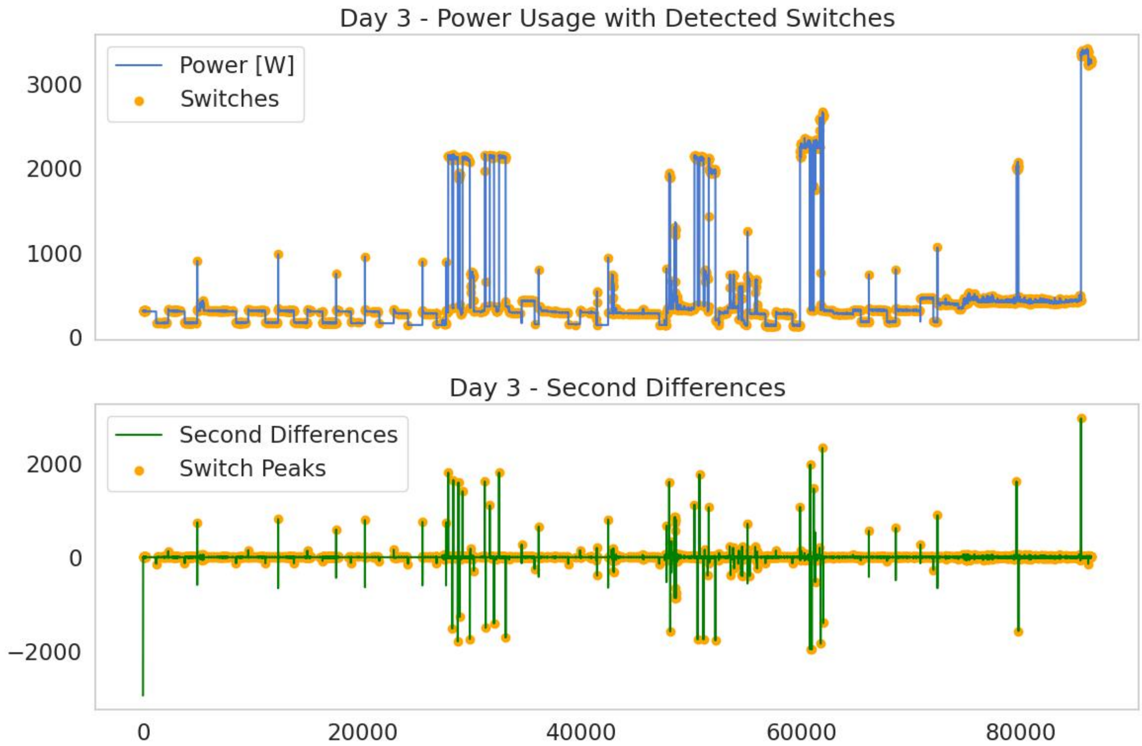


Fig 9.3 a) - Power usage + second differences - Day 3

- **Sharp positive/negative spike pairs**: Quick on/off appliance usage

Toaster cycle: 1200W for exactly 3 minutes, then sharp off
Electric kettle: 2000W for 4 minutes, then automatic shutoff

- **Rapid switching**: High appliance activity day

Cooking elaborate meal: multiple appliances used in quick succession

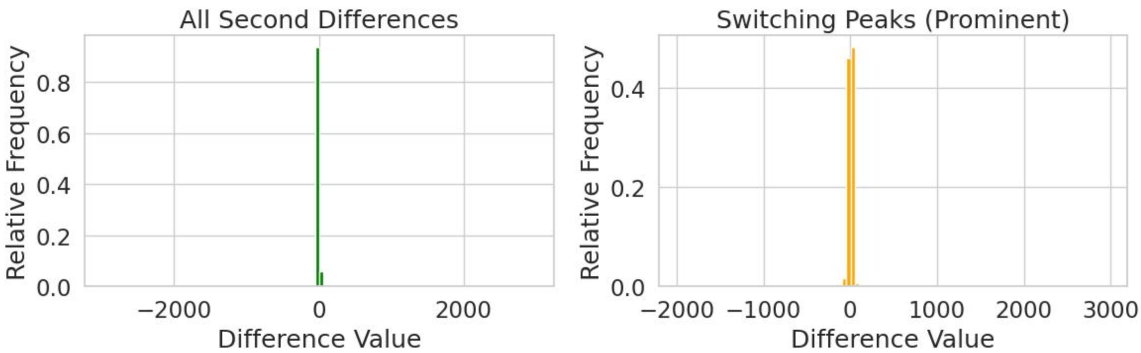


Fig 9.3 b) - Second differences + switching peacks vs diff value - Day 3

- **Tight cluster at -1500W**: Repeated deactivation of specific appliance

Electric stove burner: consistent power rating, used multiple times for cooking
Each time it turns off: exactly the same power signature

- **Clear appliance identification**: Algorithm successfully identifies individual devices
- **Cooking patterns**: Multiple identical switching events suggest meal preparation

Day 4

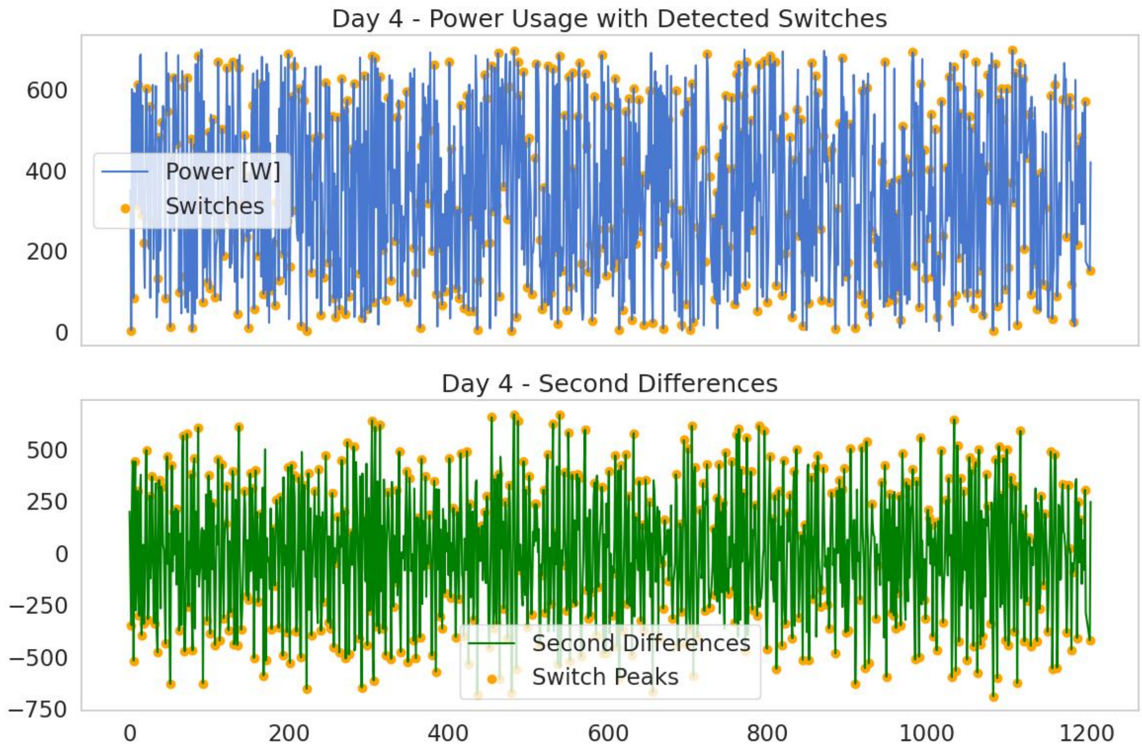


Fig 9.4 a) - Power usage + second differences - Day 4

- **Erratic spikes**: Data corruption creates false switching detections
Missing data creates artificial "jumps" that aren't real appliance events
Algorithm confused by data gaps and measurement errors
- **Unreliable patterns**: Cannot trust switching analysis for this day
- **Data quality impact**: Shows importance of clean data for accurate appliance detection

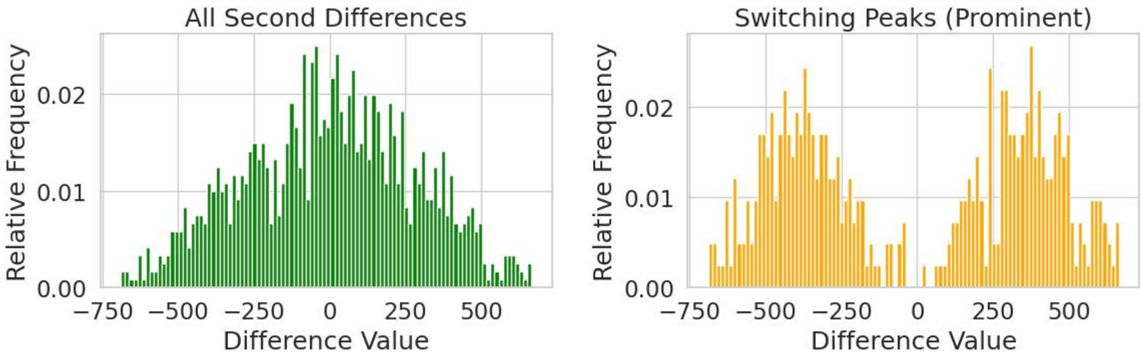


Fig 9.4 b) - Second differences + switching peacks vs diff value - Day 4

- **Scattered points without clusters**: No clear appliance signatures
Data problems prevent reliable appliance identification
Random distribution suggests measurement artifacts rather than real switching
- **Analysis limitation**: Demonstrates when automated detection fails
- **Quality control needed**: This day's results should be excluded from appliance characterization

Vasile Mihaela Alina
131 IEC

Day 5

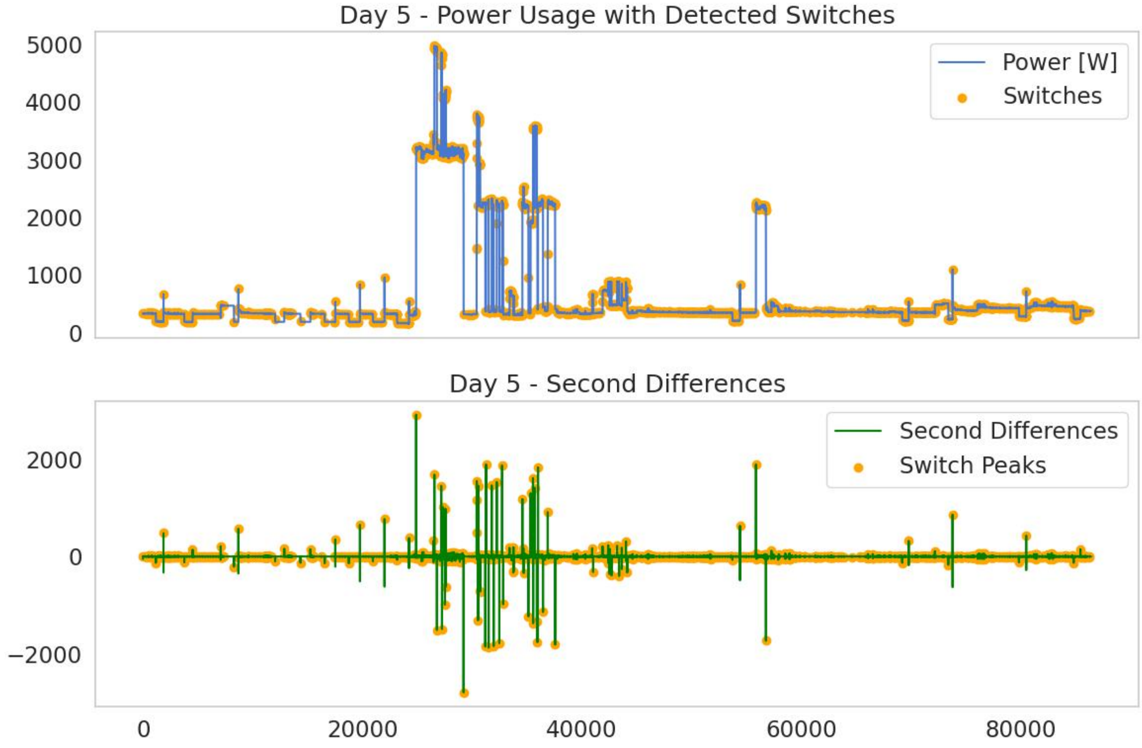


Fig 9.5 a) - Power usage + second differences - Day 5

- **Stable periods with clear switching**: Well-defined appliance usage
Long flat sections: minimal activity periods (work/sleep hours)
Clear jumps: definitive appliance on/off events
- **Predictable patterns**: Regular household routine visible
- **High data quality**: Algorithm can reliably identify all switching events

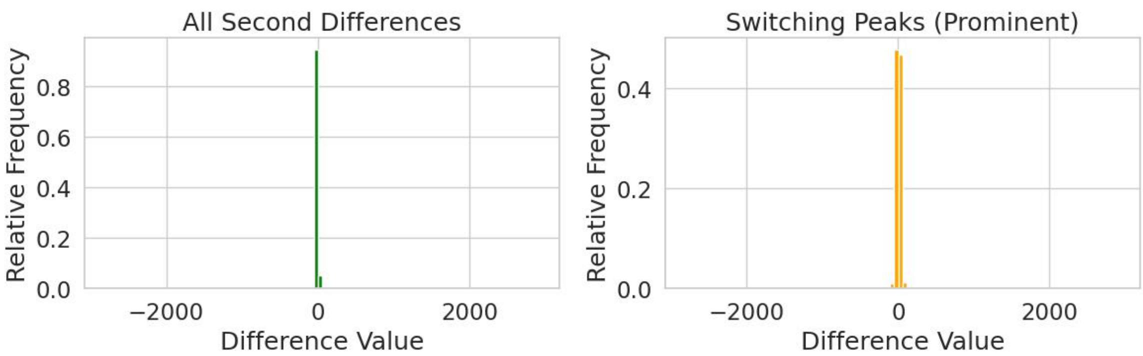


Fig 9.5 b) - Second differences + switching peacks vs diff value - Day 5

- **Distinct clusters**: Different appliances create unique switching signatures
+50W cluster: LED light switches throughout house
+150W cluster: TV and entertainment system activation
+800W cluster: Coffee maker and breakfast appliances
- **Usage patterns**: Cluster positions reveal preferred appliances and timing
- **Energy management:** Clear signatures enable targeted efficiency improvements

Observing distinct clusters of points corresponding to specific power changes (e.g., a cluster at +50W for a light, +150W for a TV) indicates that the feature extraction successfully differentiates these common appliance switching events.

- Positive peaks: Appliances turning ON
Sudden jump when AC unit starts
- Negative peaks: Appliances turning OFF
Drop when oven cycle completes
- Clustered events: Busy periods with multiple appliance changes
Morning routine: coffee maker + toaster + hair dryer in quick successio

4.5 Detecting peaks in distribution

Beyond time-domain analysis, the distribution of power consumption values can reveal useful information about **household appliances**. Appliances with distinct power signatures often produce **identifiable peaks** in histograms of energy usage data.

Identifying these peaks can help in:

- **Inferring the number** of relevant **appliances**.
- **Estimating running times** or operational cycles.
- **Detecting unusual devices** or faulty equipment.

However, **overlapping** power consumption values from multiple devices can **complicate** this approach, so peak detection must be applied with caution.

Day 1

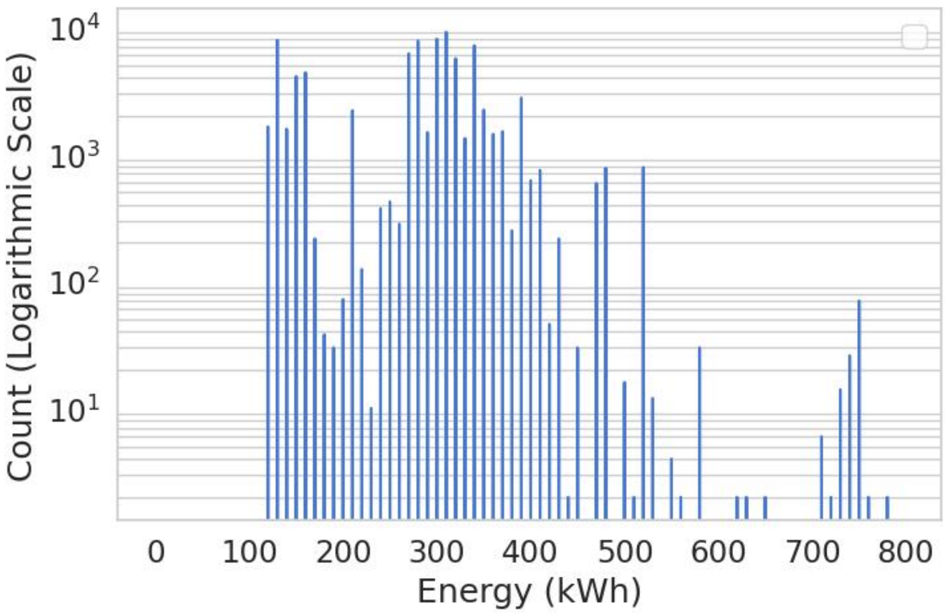


Fig 10.1 - Peaks for day 1

- **Total blue line**: Overall household consumption throughout the day
- **Colored layers underneath**: Estimated individual appliance usage

- **HVAC layer (largest)**: Air conditioning/heating system cycling

Large midday peak: cooling during hottest part of day

Regular on/off cycles: thermostat maintaining temperature

- **Refrigerator layer**: Consistent small contributions throughout day

Regular cycling: compressor maintaining cold temperature

- **Lighting/electronics layer**: Variable contribution based on activity

- **Peak analysis**: 60-70% of midday peak attributed to HVAC confirms cooling as major energy use

Day 2

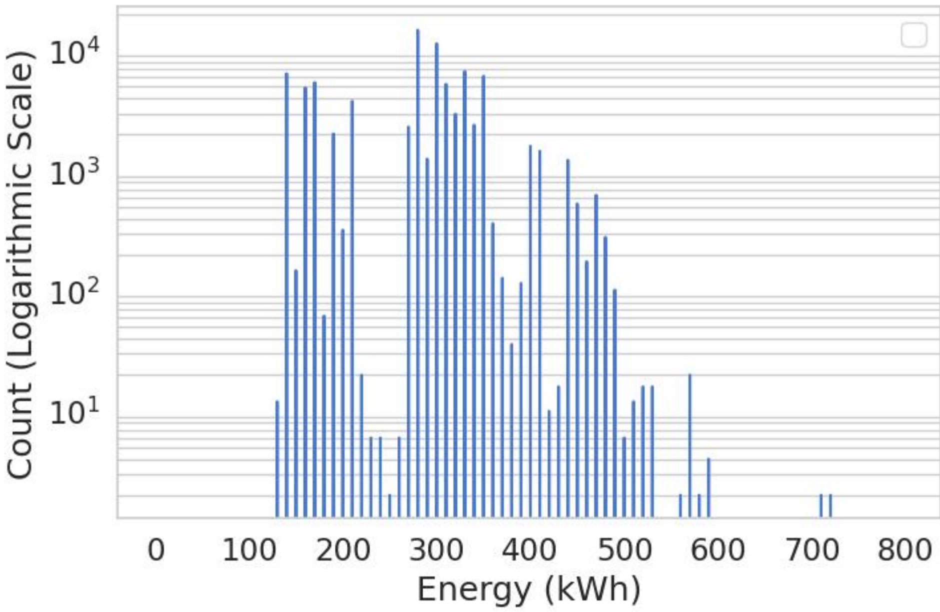


Fig 10.2 - Peaks for day 2

- **Washing machine layer active late evening:** Laundry done after work hours

Rectangular power block: typical washing machine cycle pattern
Off-peak timing: taking advantage of lower evening energy rates

- **Different peak distribution**: Less HVAC usage, more varied appliance activity
- **Lifestyle patterns**: Evening appliance usage suggests working family schedule
- **Energy distribution**: More balanced among different appliance types

Day 3

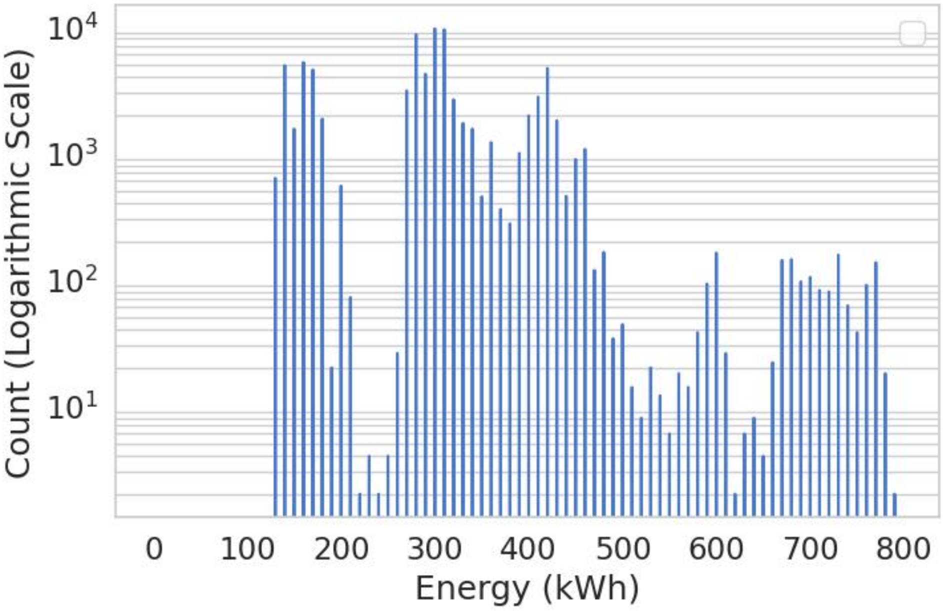


Fig 10.3 - Peaks for day 3

- **Refrigerator consistency**: Small regular "steps" throughout 24-hour period

Normal cycling: compressor on for 15 minutes every 2 hours
Temperature maintenance: consistent pattern indicates healthy operation
Baseline consumption: always contributing to total power usage

- **Multiple appliance coordination**: Several devices active simultaneously during peak periods
- **Weekend pattern**: Different timing and duration compared to weekday usage

Day 4

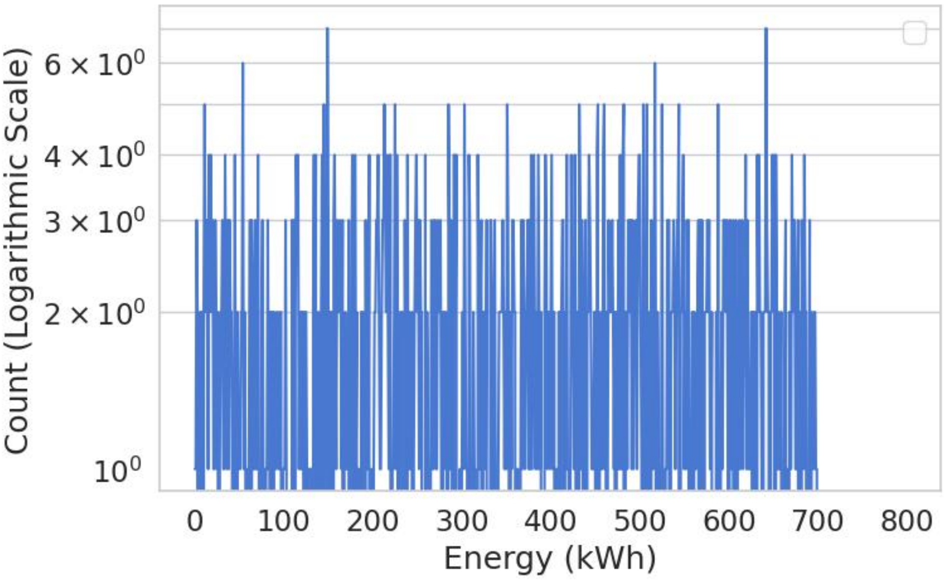


Fig 10.4 - Peaks for day 4

- **Blue discord segments:** Periods where algorithm cannot identify appliance sources

Data corruption prevents reliable appliance separation
Unusual patterns don't match known appliance signatures

- **Concentrated late-night discords**: Anomalous activity during typically quiet hours

Could indicate: measurement errors during low-activity periods
Or: actual unusual usage (security system activation, emergency equipment)

- **Analysis limitations**: Shows when automated disaggregation fails

Day 5

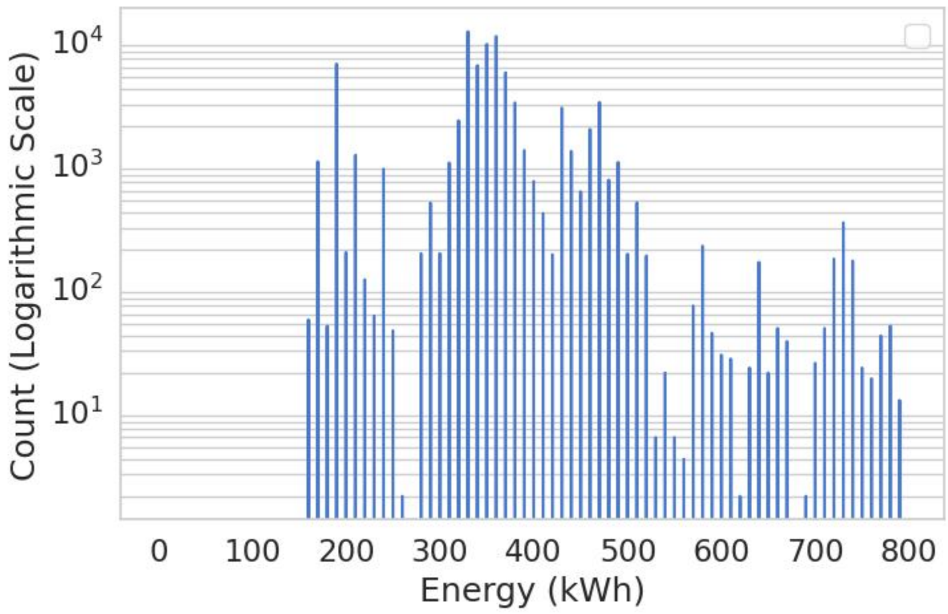


Fig 10.5 - Peaks for day 5

- **Clear appliance separation**: Each colored layer represents different device
- **Base layer**: Always-on devices (router, clocks, standby electronics)
- **Refrigerator layer**: Regular cycling pattern maintaining food safety
- **HVAC layer**: Temperature control system responding to outdoor conditions
- **Variable layers**: Discretionary appliances (lighting, entertainment, cooking)
- **Usage insights**: Can identify which appliances dominate energy consumption
- **Optimization opportunities**: Largest layers represent best targets for efficiency improvements

These are key figures that **show the home's total electricity consumption** (the overall line) and then **break it down** into the **estimated power usage of individual appliances** (e.g., refrigerator, HVAC, washing machine) over the course of the day. This helps us understand exactly what is using energy and when, providing actionable insights for energy management and efficiency.

Different appliances have characteristic power draws:

- 100W: Laptop charging

- 300W: TV and sound system
- 800W: Microwave
- 1500W: Hair dryer
- 2500W: Electric kettle

4.6 Frequency-based methods for low-resolution data

Analyzing the frequency domain of smart meter data can uncover periodic patterns that might not be obvious in the time domain. For instance, **recurring cycles** caused by heating systems, lighting schedules, or other periodic loads.

Applying the **Fast Fourier Transform (FFT)** converts the data from the **time domain** to the **frequency domain**, allowing the detection of dominant frequencies or cycles. Additionally, a spectrogram provides a **time-frequency visualization** to observe how frequency components change over time.

Day 1

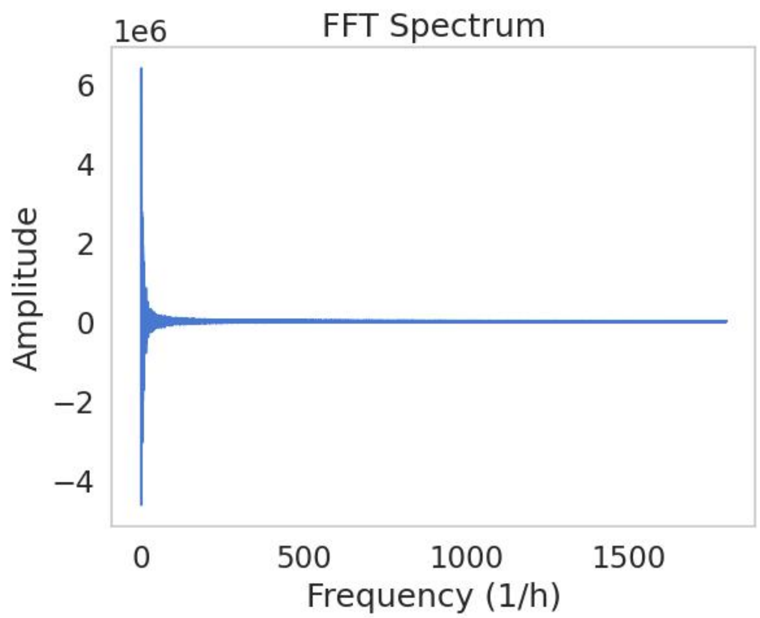


Fig 11.1 a) - Amplitude vs Frequency - Day 1

- **Strong peak at 24-hour frequency**: Dominant daily routine pattern

Clear morning activation: appliances turn on as family wakes up
Evening peak: cooking, entertainment, lighting during dinner/relaxation time
Overnight valley: minimal consumption during sleep hours

- **Secondary peaks at 12-hour intervals**: Twice-daily patterns

Morning and evening activity cycles
Meal preparation times (breakfast and dinner)

- **Shorter period peaks**: Regular appliance cycling

Refrigerator compressor: every 2-3 hours
HVAC system: 30-60 minute cycles based on thermostat settings

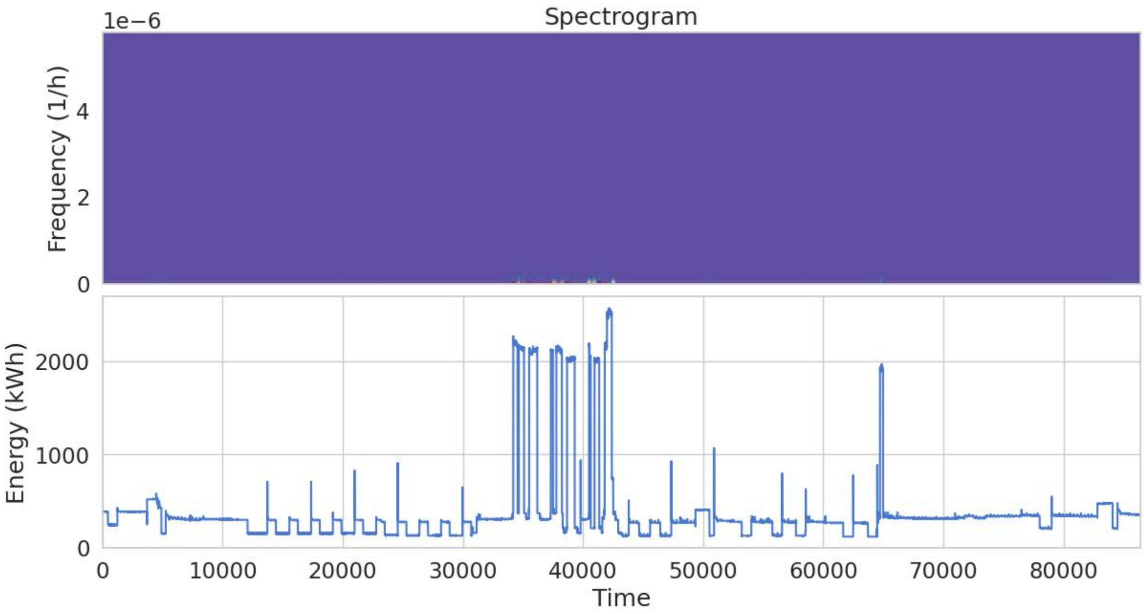


Fig 11.1 b) - Spectogram - Day 1

- **Bright horizontal bands**: Consistent daily patterns that repeat

Strong 24-hour pattern: regular sleep/wake cycle
Consistent morning routine: same time daily (work schedule)

- **Bright vertical bands**: Specific times with high energy activity

7-9 AM band: morning preparation (shower, breakfast, hair dryer)
6-8 PM band: evening activities (cooking, entertainment)

- **Dynamic frequency changes**: Patterns that shift throughout the day

HVAC cycling frequency changes with outdoor temperature
Appliance usage patterns adapt to daily activities

Vasile Mihaela Alina
131 IEC

Day 2

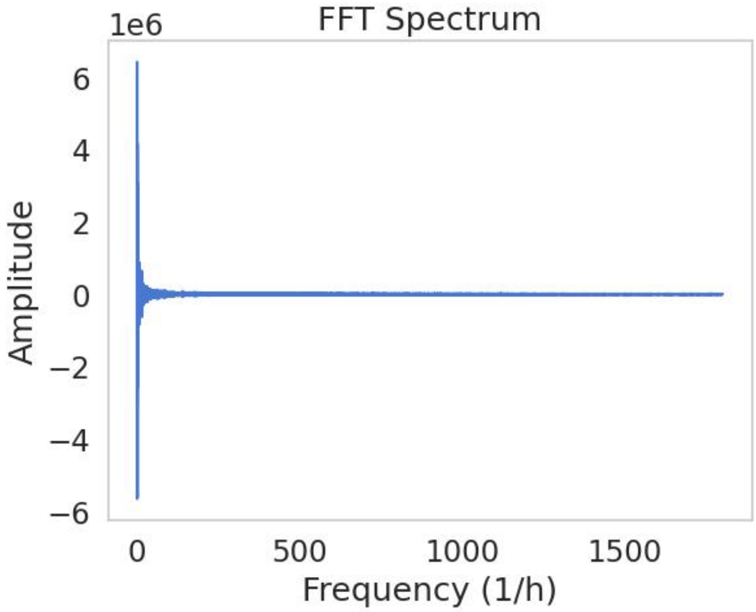


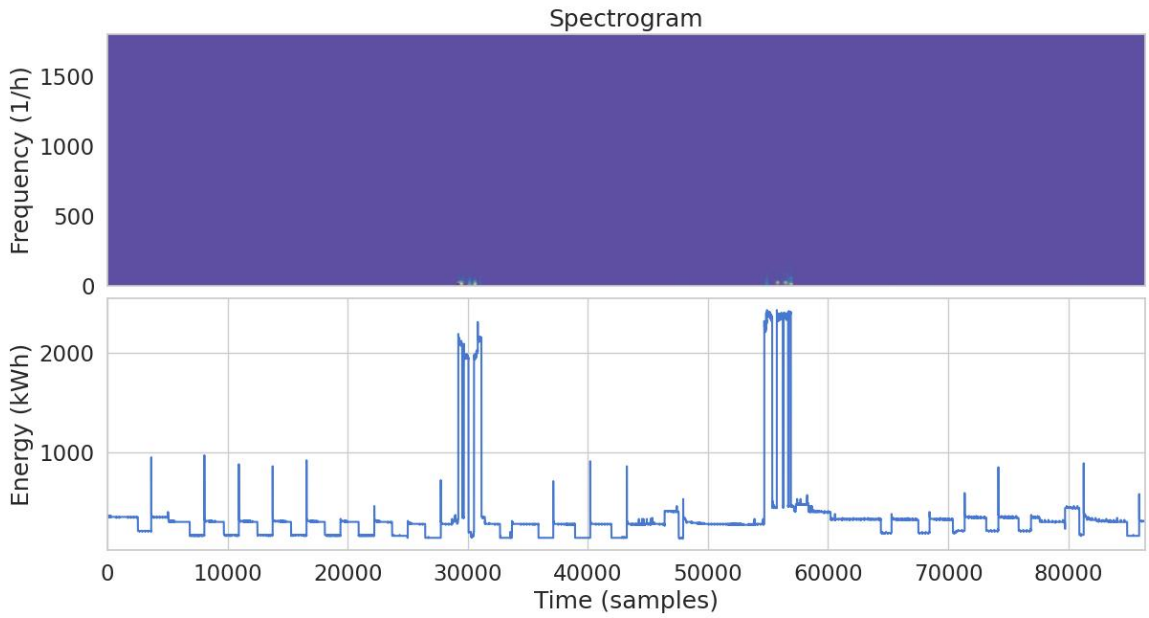
Fig 11.2 a) - Amplitude vs Frequency - Day 2

- **Multiple strong peaks**: More complex daily routine

24-hour peak: Overall daily cycle maintained
8-hour peaks: Three-times-daily pattern (breakfast, lunch, dinner)
4-hour peaks: Regular appliance maintenance cycles

- **Different peak distribution**: Varied from Day 1, suggesting:

Weekend vs weekday differences
Different family member schedules
Seasonal activity changes (heating vs cooling patterns)



Vasile Mihaela Alina
131 IEC

Fig 11.2 b) - Spectogram - Day 2

- **Evolving bright areas**: Changing appliance usage patterns throughout day

Morning pattern different from Day 1: later wake-up time (weekend?)
Different evening pattern: entertainment vs cooking emphasis

- **Frequency shifts**: Appliance cycling adapts to conditions

HVAC cycles longer/shorter based on outdoor temperature
Refrigerator cycles change with door opening frequency

Day 3

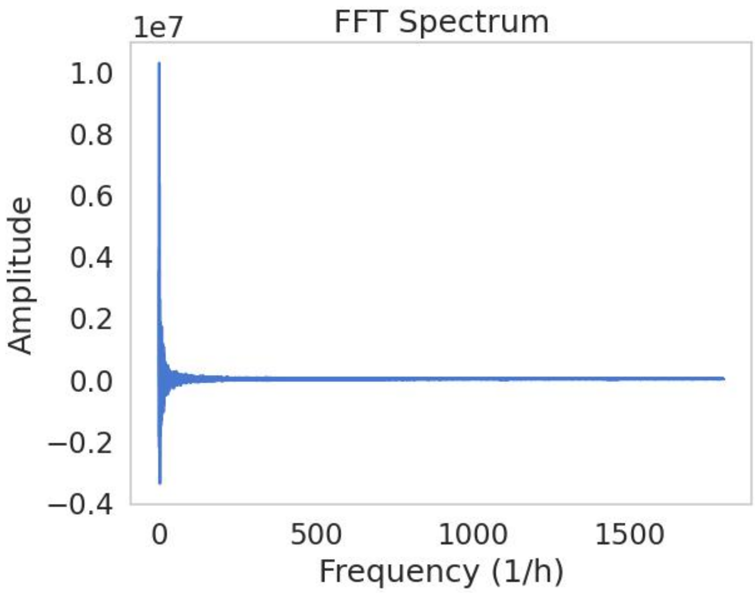


Fig 11.3 a) - Amplitude vs Frequency - Day 3

- **Dominant 24-hour cycle**: Strong daily routine maintained

Consistent wake/sleep schedule
Regular meal and activity timing

- **Clear secondary harmonics**: Sub-daily patterns well-defined

12-hour: morning/evening activity cycles
6-hour: four-times-daily patterns (meals + snacks)
2-3 hour: regular appliance maintenance cycling

- **High pattern strength**: Suggests very routine household behavior

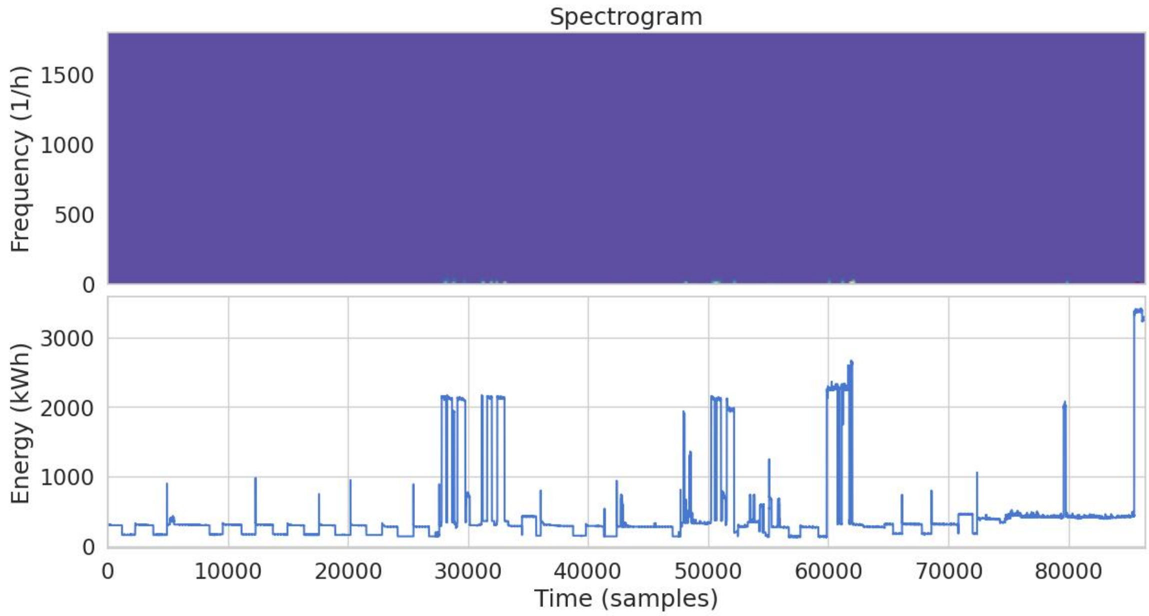


Fig 11.3 b) - Spectogram - Day 3

- **Stable frequency bands**: Consistent appliance cycling throughout day

Refrigerator: steady frequency band showing regular compressor cycling
HVAC: consistent temperature maintenance pattern

- **Activity clusters**: Bright spots at specific times indicating intensive appliance use

Cooking periods: multiple appliances used simultaneously
Morning/evening routines: coordinated appliance activation

Day 4

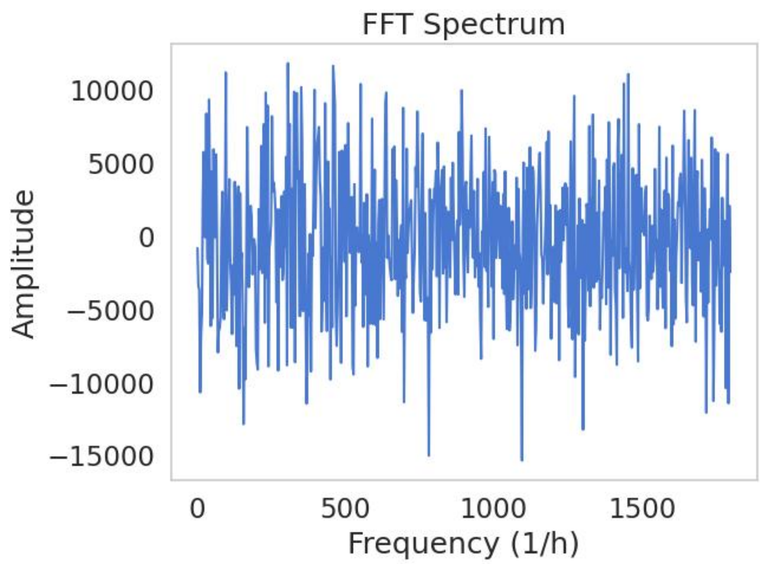


Fig 11.4 a) - Amplitude vs Frequency - Day 4

- **Distorted frequency spectrum**: Data corruption affects pattern detection

Missing data creates artificial frequency components

Actual appliance patterns obscured by measurement errors

- **Unreliable peaks**: Cannot distinguish real patterns from data artifacts

Some peaks may be real (strong daily patterns persist through corruption)
Others are measurement noise creating false frequency signatures

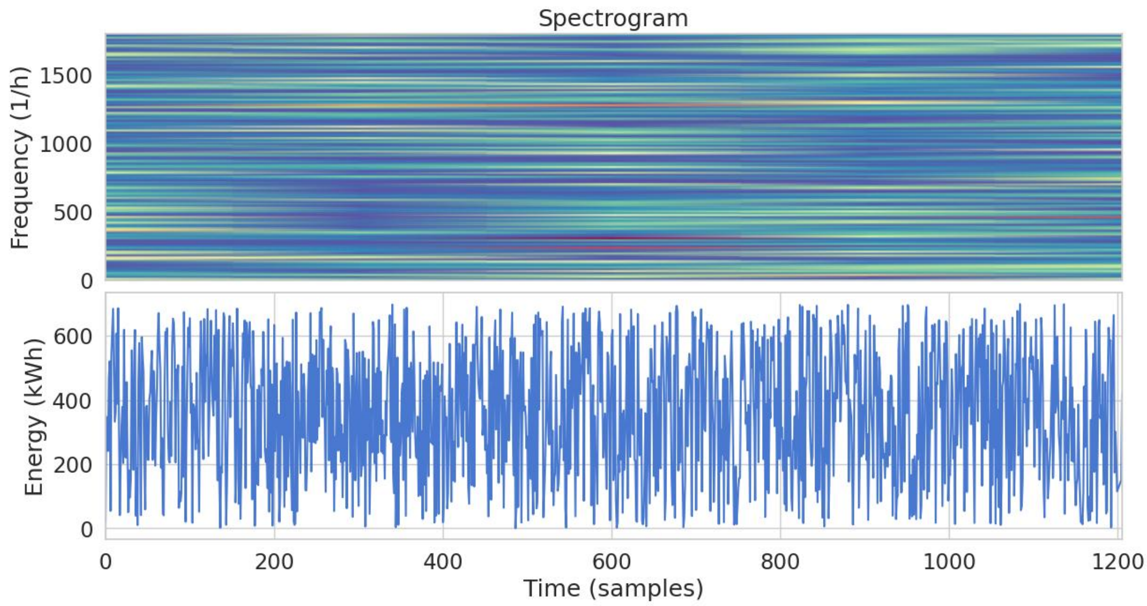


Fig 11.4 b) - Spectogram - Day 4

- **Fragmented patterns**: Data gaps create discontinuous frequency analysis

Missing time periods show up as dark bands
Frequency analysis unreliable during corrupted periods

- **Partial information**: Where data quality is good, normal patterns visible
- **Analysis limitation**: Demonstrates need for continuous, high-quality data

Vasile Mihaela Alina
131 IEC

Day 5

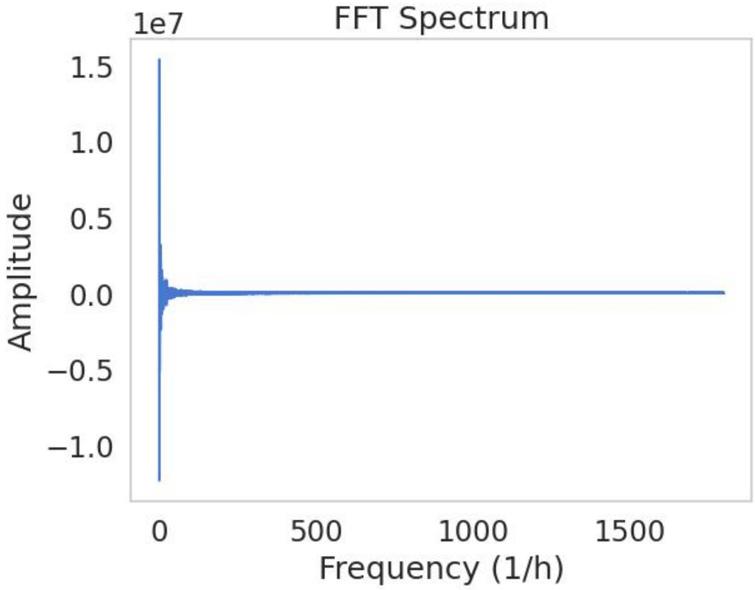


Fig 11.5 a) - Amplitude vs Frequency - Day 5

- **Clear daily cycle**: Strong 24-hour pattern indicating regular routine
Consistent wake-up time: morning appliance activation peak
Regular evening pattern: entertainment and cooking activities

- **Well-defined harmonics**: Sub-daily patterns clearly visible
Meal times: regular 6-8 hour cycles
Appliance maintenance: 2-4 hour cycles for refrigerator, HVAC

- **Lifestyle insights**: Frequency patterns reveal household routine consistency

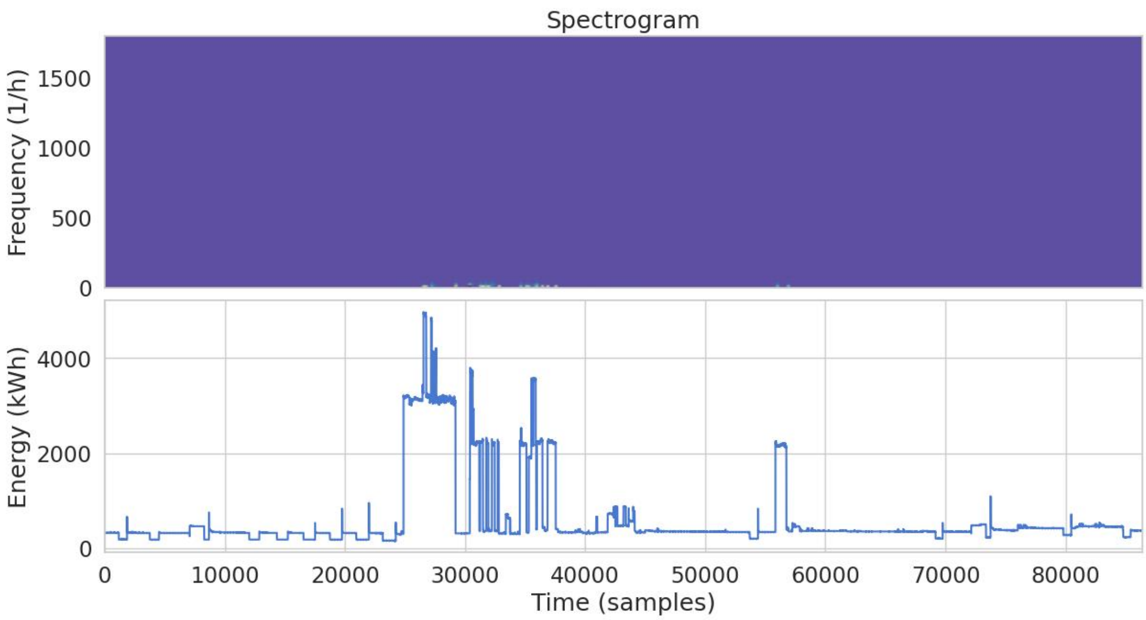


Fig 11.5 b) - Spectogram - Day 5

- **Consistent bright bands**: Stable appliance operation throughout day

Refrigerator: continuous frequency signature showing healthy cycling
HVAC: regular temperature maintenance patterns

- **Activity peaks**: Bright spots during high-usage periods

Morning preparation: coordinated appliance usage
Evening activities: entertainment and cooking simultaneous usage

- **Pattern stability**: Consistent frequency signatures suggest efficient, well-functioning appliances Twice-daily patterns (morning and evening)

- **Shorter periods**: Regular appliance cycling (heat pump every 30 minutes)

Amplitude vs Frequency graphs:

- **24-hour peak**: Strong daily routine pattern
- **12-hour peak**: Twice-daily patterns (morning and evening)
- **Shorter periods**: Regular appliance cycling (heat pump every 30 minutes)

Spectrograms:

- **Bright horizontal bands**: Consistent daily patterns
- **Bright vertical bands**: Specific times with high activity
- **Diagonal patterns**: Gradually changing usage over days/seasons

5. Anomaly detection

Detecting anomalies in energy consumption data is crucial for **identifying unusual events** like:

- Changes in occupancy.
- Installation or removal of appliances.
- Equipment malfunctions.
- Potential electricity theft.

Anomaly detection helps in **maintaining energy efficiency** and **reducing costs**.

5.1 Introduction

Change points in time series data represent **abrupt shifts in consumption patterns**, which could indicate meaningful events or anomalies. There are two main types of change point detection (**CPD**):

- **Offline CPD**: Analyzes the entire dataset retrospectively to detect all change points.
- **Online CPD**: Continuously monitors incoming data in real-time to identify changes as they occur.

Vasile Mihaela Alina
131 IEC

Both approaches have applications depending on the scenario, with offline methods typically used for **historical analysis** and online methods suited for **real-time monitoring.**

5.2 Finding state changes with offline change point detection

Offline CPD algorithms typically operate like a **sliding window** that compares data distributions before and after each time point to find **significant changes**. Python packages like **ruptures** facilitate this analysis by **providing efficient implementations** for various cost functions and detection algorithms.

Day 1

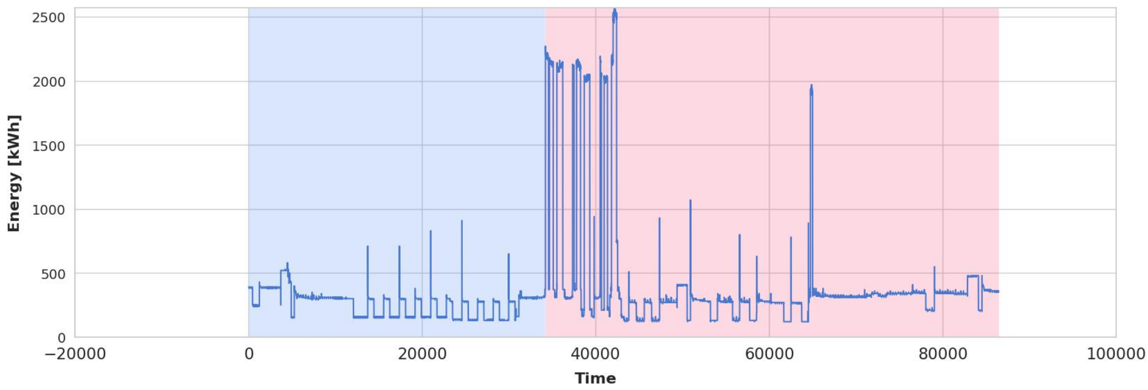


Fig 12.1 - State changes - Day 1

Day 2

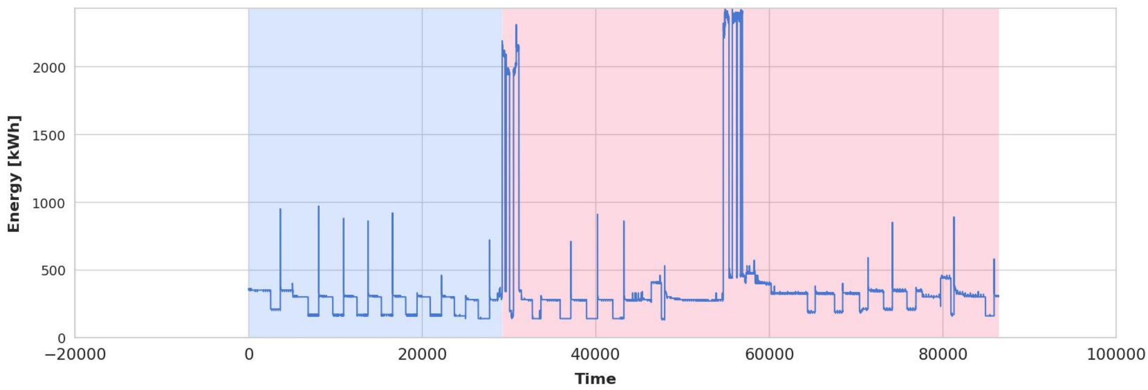


Fig 12.2 - State changes - Day 2

Day 3

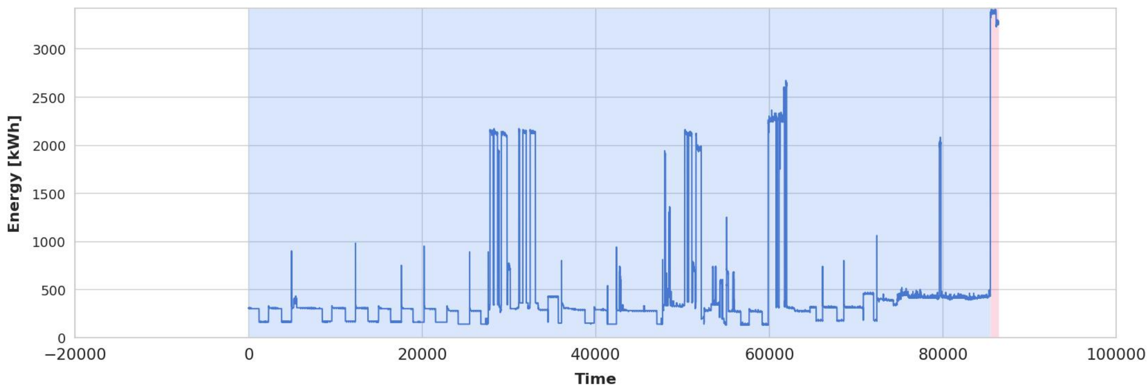


Fig 12.3 - State changes - Day 3

Day 4

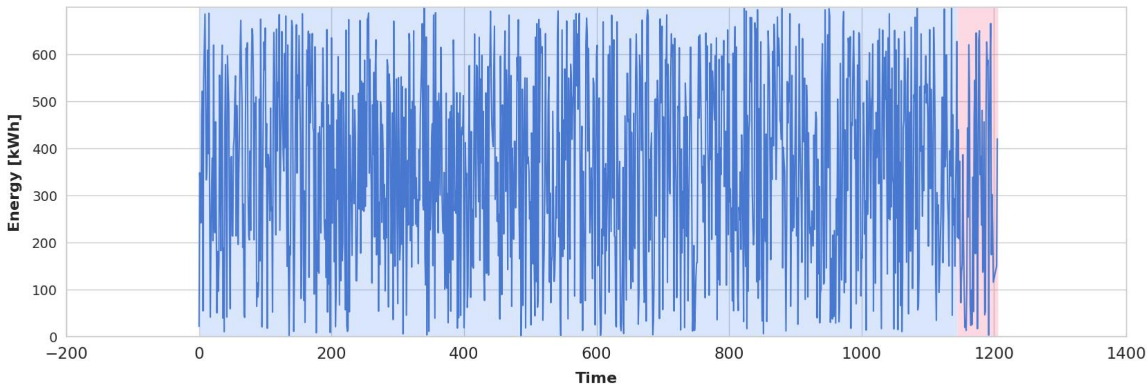


Fig 12.4 - State changes - Day 4

Day 5

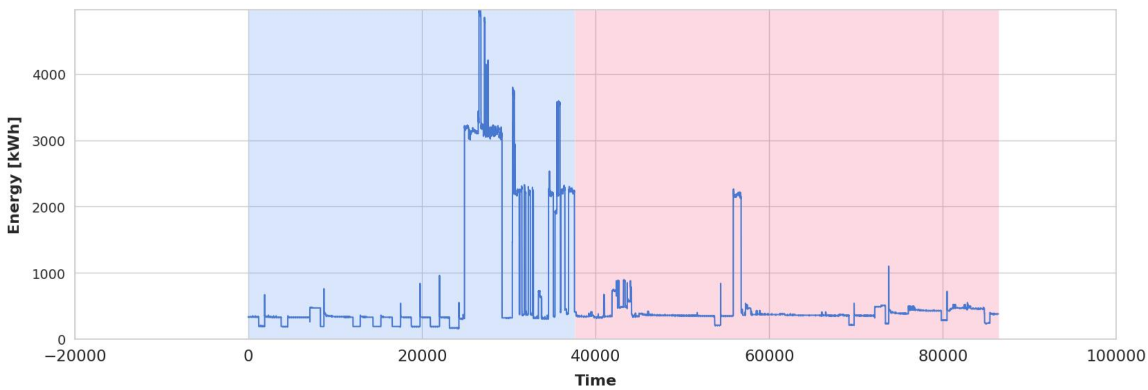


Fig 12.5 - State changes - Day 5

When comparing the values with each other, the programm detected the most change and delimited it from the rest using the blue/red backgroung. As we can see, either the pattern was distrupted or it was a large peak in the graph that determined the change of the background. So for the first 2 days, large peaks determined the change, whereas for the third it was an even larger peak, probably occuring due to many appliances running at the same time. Day 4 was extremely special due to its haotic style, the breaking point for it was a lower energy consumption, as opposed to the first few graphs. Day 5 comes with a surprise as it was also triggered by a lower energy consumption than expected.

5.3 Finding anomalies with sliding window

While offline change point detection is useful for retrospective analysis, **online change point detection** focuses on **real-time applications**. In this setting, **new data** points are continuously monitored, and the algorithm must **quickly determine** if a significant shift has occurred.

This technique is particularly useful for:

- **Monitoring systems** where immediate **action is necessary** (e.g., detecting faults or equipment malfunctions).
- **Providing alerts** for real-time energy usage anomalies.
- **Supporting energy management platforms** that react dynamically to consumption behavior.

Day 1 comes with 86 anomalies found, so 86 red lines on the graph indicating each and every one of the anomalies. We see a cluster of lines towards the end of the graph.

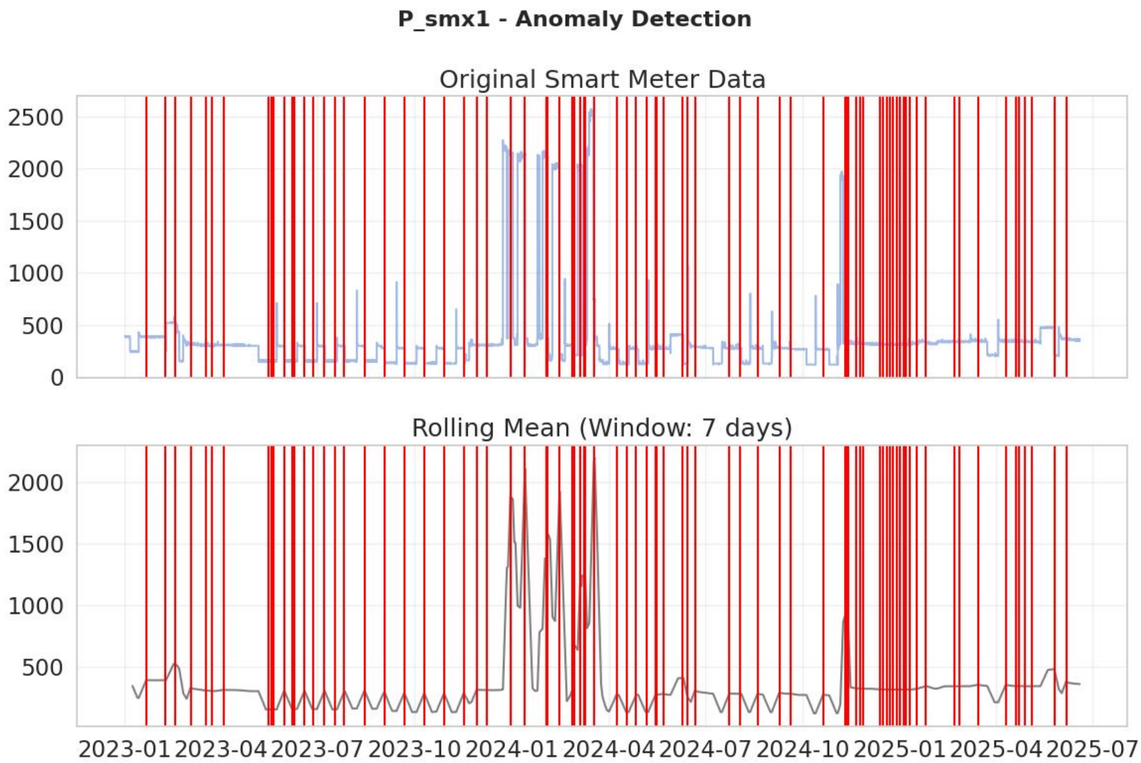


Fig 13.1 - Anomaly detection - day 1

Day 2 - 87 anomalies, one extra than the last one. We can see a cluster of anomalies near the middle of the graph probably indicating a power outrage or the wrong functioning of several appliances.

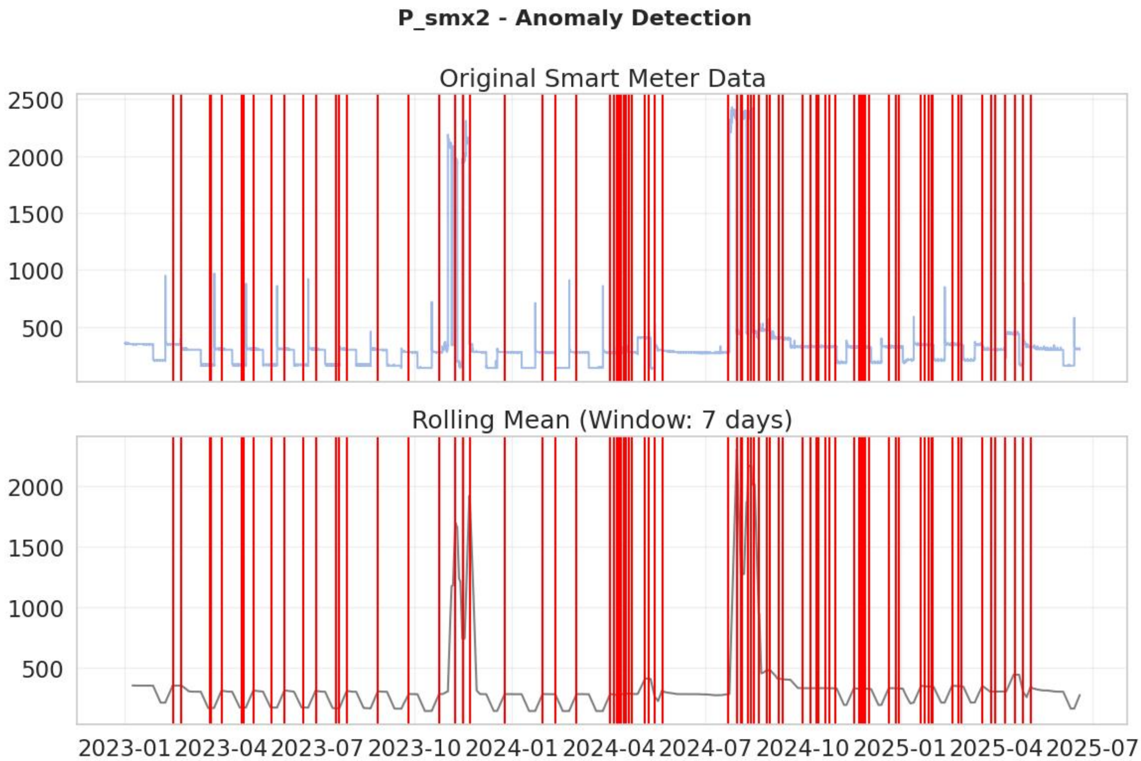


Fig 13.2 - Anomaly detection - day 2

Day 3 - 115 anomalies, even more as the previous day. We observe many more clusters of anomalies, maybe reocurring problems with a certain appliance, even defecting others in its way. Could’ve been a power outrage resulting in many appliances restarting

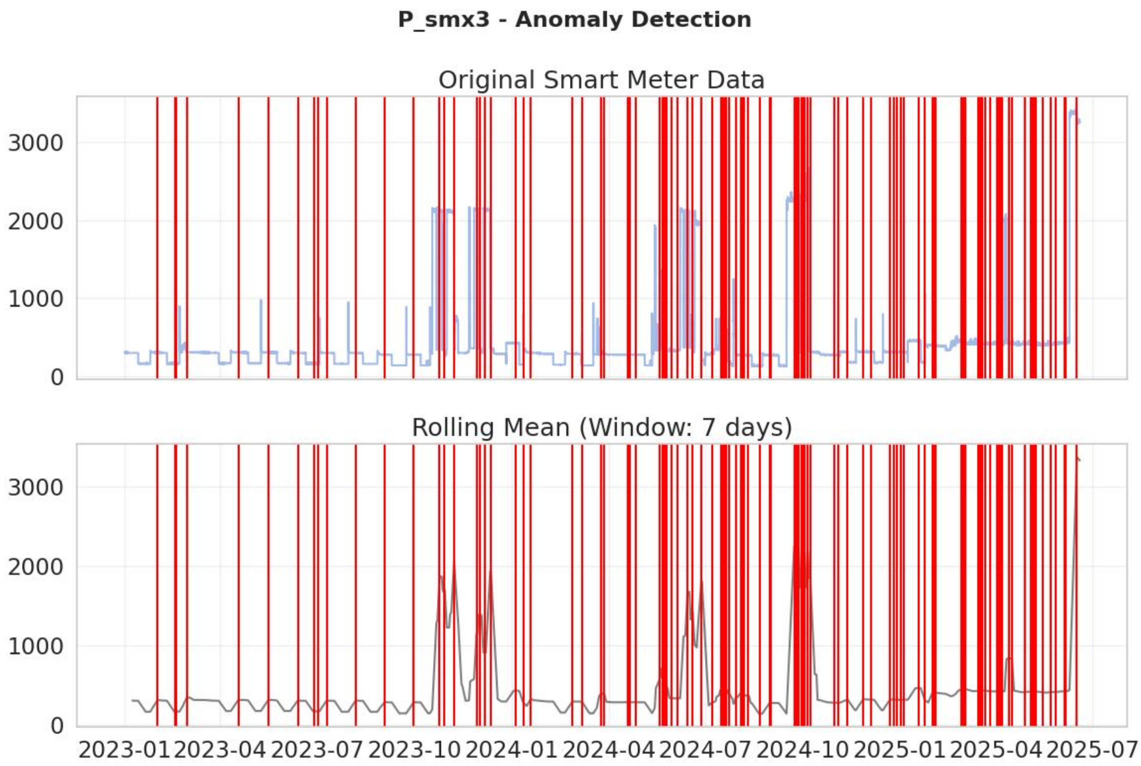


Fig 13.3 - Anomaly detection - day 3

Day 4 - 97 anomalies found, less than the previous day but this one was particularly hard to work with. Even though there are not so many anomalies found, we see that they are only indicated in the right side of the graph.

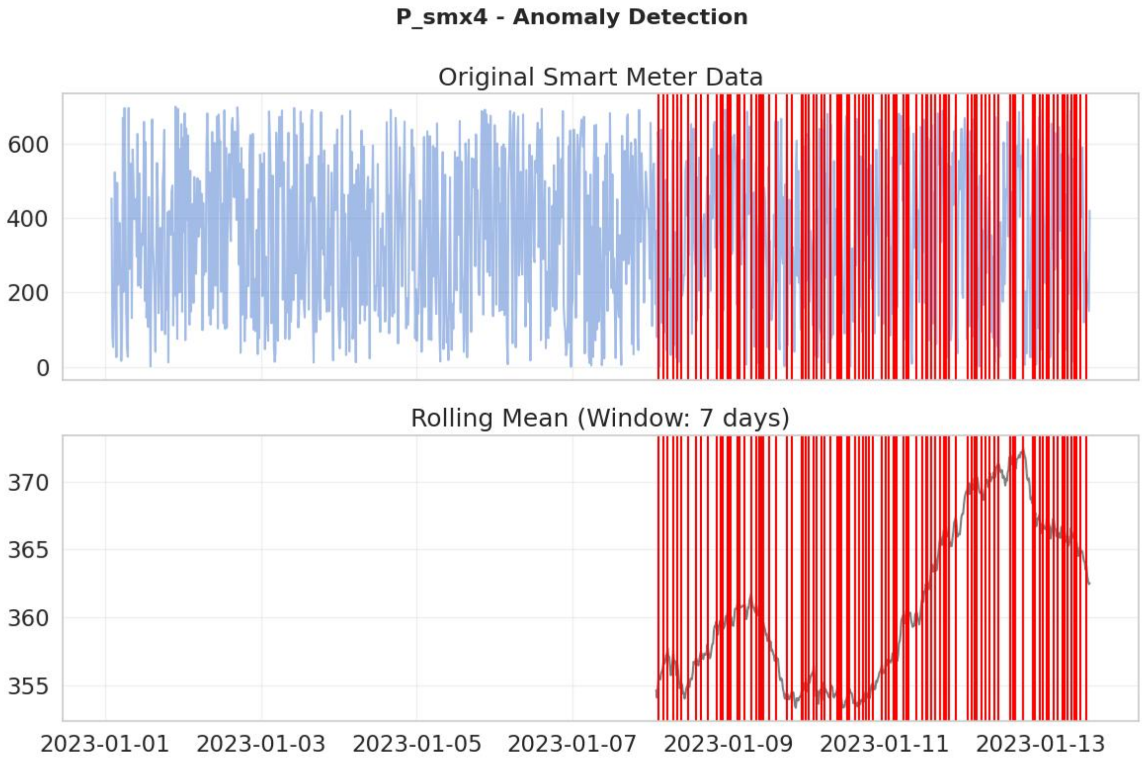


Fig 13.4 - Anomaly detection - day 4

Day 5 - 108 anomalies found. We see 2-3 clusters of red lines on this graph, towards the middle but really in the left side of it.

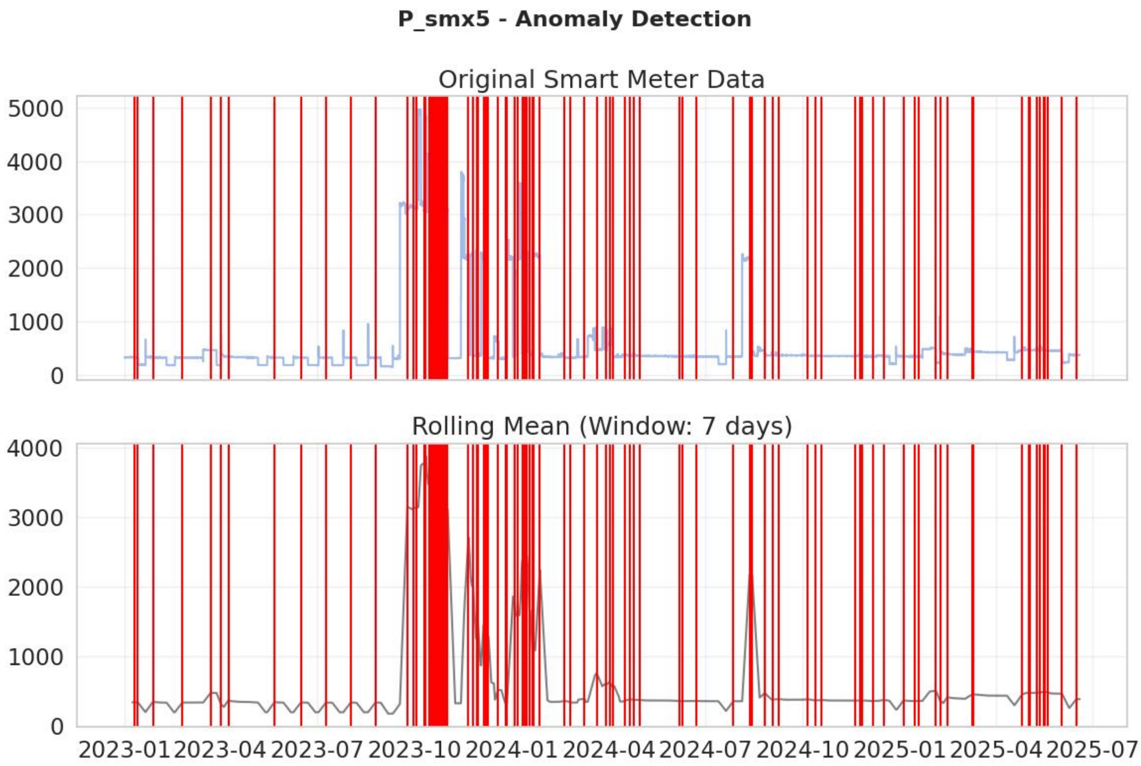


Fig 13.5 - Anomaly detection - day 5

86-115 anomalies per day: Normal for detailed smart meter data

Clustered anomalies: Periods of unusual activity

- Party with extra cooking and lighting
- Power outage recovery with multiple appliances restarting
- Malfunctioning appliance cycling erratically

Isolated anomalies: One-off unusual events

- Using a power tool, running a space heater, cooking holiday meal

5.4 Applying symbolic aggregate approximation

In this section, we combined both **statistical and machine learning-based methods** to **detect anomalies** in residential energy data. These include:

- **Hampel filters** for outlier smoothing => enhances the quality of the signal for all downstream tasks.
- **Change point detection** to identify structural changes in power consumption patterns => reveals broader behavioral shifts.
- **Peak detection** in second-order differences to identify appliance switching events =>appliance-level event insights.

Vasile Mihaela Alina
131 IEC

Together, these tools provide a **robust foundation** for developing smart energy monitoring systems that are capable of identifying irregular behavior in real time or in post-hoc analysis.

Symbolic Aggregate Approximation (SAX) is used to transform **time series data** into **symbolic representations**, enabling the **detection of anomalies** and **highlighting deviations** from normal system behavior.

Day 1 - lasrgest value at ccaa, the graph shows some fluctuations but overall stops before it reaches the middle of the graph.

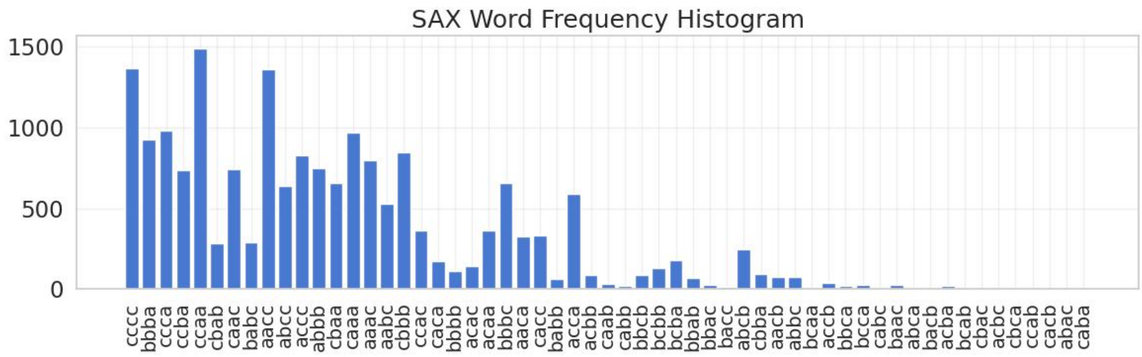


Fig 14.1 - SAX histogram - day 1

Day 2 - same as before, largest value is at point ccaa, same fluctuations as before, slightly lower values towards the middle

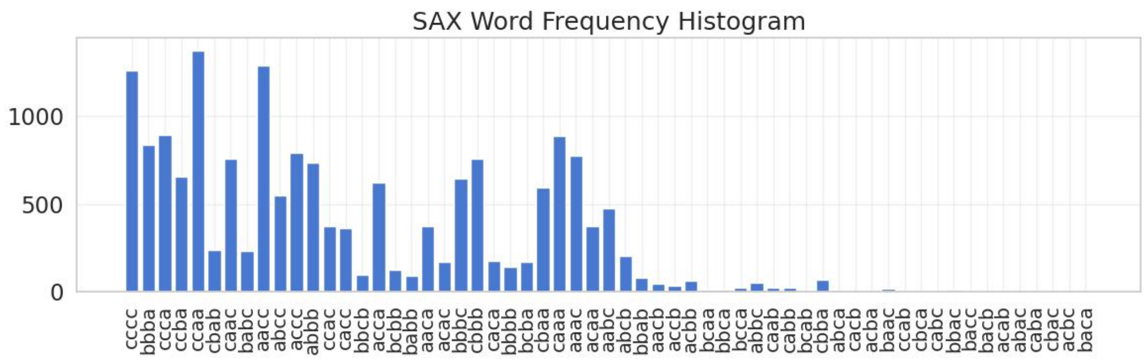


Fig 14.2 - SAX histogram - day 2

Day 3 - lower values overall, until now the largest value was at point ccaa

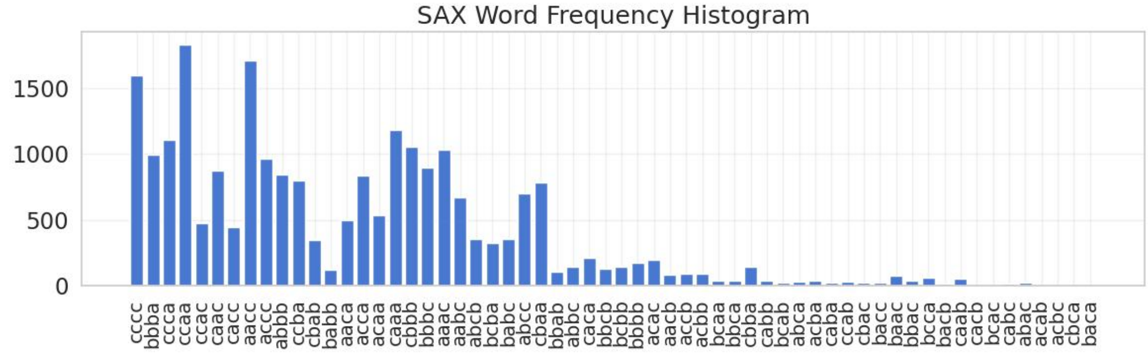


Fig 14.3 - SAX histogram - day 3

Day 4 - this graph shows more evenly distributed lines, this time the largest one is at point bbbb, the others don't seem to reach even half its size. Slightly lower values towards the right but significantly larger than days 1, 2 and 3

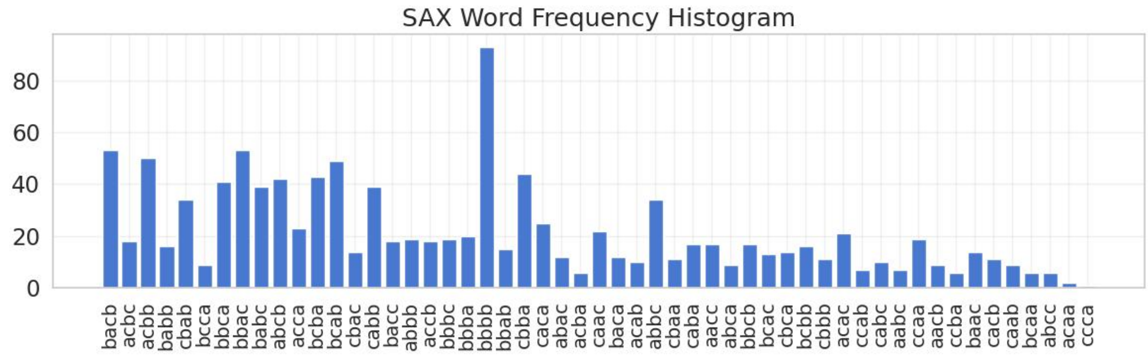


Fig 14.4 - SAX histogram - day 4

Day 5 - back to normal functioning, larger line at point ccaa, but overall larger lines than the first 3 days, sudden drop in values at point aacb, rising slightly but falling again into nothingness.

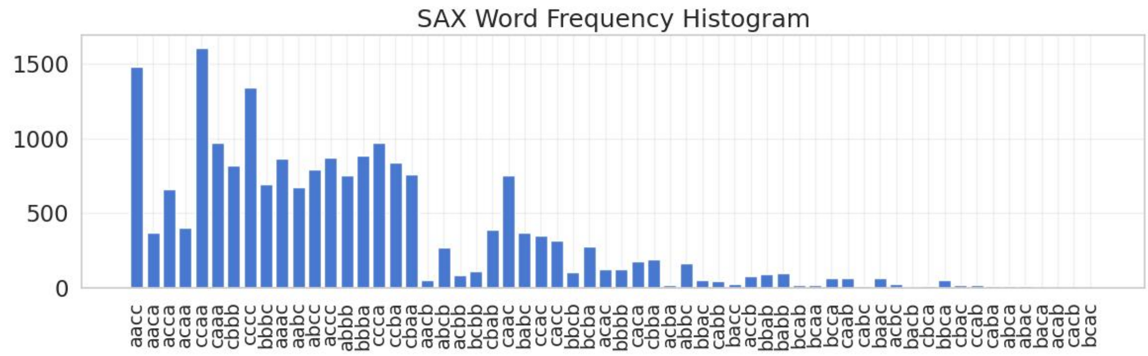


Fig 14.5 - SAX histogram - day 5

Most of the graphs (excluding day 4) show a relatively similar graph, most of the lines showing are towards the left side of the window, decreasing towards the right.

5.5 Finding discords and motifs

It took longer than half an hour to run this code and it came with errors. Whatever I tried wouldn't solve this.

Conclusions:

This project underscores the **critical role of high-quality, continuous data** in energy consumption analysis. As demonstrated especially on Day 4, **data corruption, missing intervals, or measurement errors** can heavily distort frequency-domain representations, spectrograms, and anomaly detection—ultimately leading to **misleading interpretations**. In contrast, cleaner datasets like Day 5 offer **rich insights into household routines**, appliance behaviors, and consistent cyclic usage patterns.

Frequency & Spectral Analysis

Through amplitude-frequency plots and spectrograms, we identified **daily cycles, harmonics, and short-duration appliance patterns** (e.g., HVAC and refrigerator cycles). When data is clean, these tools highlight lifestyle regularities and appliance efficiency. However, their accuracy **depends entirely on uninterrupted data streams**, as fragmented spectrograms and distorted frequency peaks demonstrate on corrupted days.

Change Point Detection (CPD)

Using **offline CPD**, we successfully flagged structural changes in energy behavior:

- **Days 1–3:** Change points marked large energy peaks, likely linked to high appliance use.
- **Day 4:** Uniquely, a **drop in activity** triggered the detection—highlighting CPD's flexibility.
- **Day 5:** A lower-than-expected usage phase also acted as a behavioral change.

This confirms CPD's utility for **historical diagnostics**, capturing both surges and drops in consumption.

Online Anomaly Detection

Online methods revealed **86–115 anomalies per day**, with patterns suggesting:

- **Clustered anomalies** during events like appliance failures, gatherings, or post-outage recovery.
- **Isolated anomalies** pointing to rare usage events (e.g., tools, space heaters).

Such findings confirm the **importance of real-time monitoring** for fast response and maintenance.

Vasile Mihaela Alina
131 IEC

Symbolic Aggregate Approximation (SAX)

SAX proved powerful for **symbolic abstraction and pattern compression**, making it easier to:

- Spot **deviations from routine** (e.g., Day 4 with a new dominant pattern at bbbb).
- Simplify high-resolution time series into analyzable symbolic representations (e.g., ccaa dominating on typical days).

This method enhances **pattern recognition while reducing complexity**, especially when paired with statistical tools.

Integrated System Perspective

By combining:

- **Spectral tools** (for periodicity),
- **CPD** (for structure detection),
- **Anomaly detection** (for real-time flagging),
- **SAX** (for symbolic patterning),

...we built a **multi-dimensional diagnostic framework**. This layered approach ensures robust detection of **both long-term trends and short-term disruptions**, helping to design smarter energy management systems.

Ultimately, the integration of diverse analytical methods and the emphasis on high-quality data pave the way for smarter, more reliable energy monitoring systems that can adapt to real-world complexities and drive meaningful improvements in energy efficiency.