NOVIA
UNIVERSITY OF APPLIED SCIENCES

# ANALYSIS OF ELECTRICITY CONSUMPTION PROFILES

Johan Westö

SERIES R: REPORTS 13/2023

**Abstract**

Renewable energy sources are transforming the electricity market and forcing electricity companies to reevaluate their policies for charging their customer. However, potential policy changes should be based on a thorough understanding of the customers' current behaviors. Here, we used a technique called UMAP to group customers with similar consumption profiles. The analysis revealed five smaller customer groups with distinctive consumption profiles and a large group where the customers' consumption profiles changed gradually, primarily from yearly to daily variations. The found groups were relevant for understanding the customers' profile costs, that is, whether a customer uses electricity when it is cheap or expensive.

**Sammanfattning**

Introduktionen av förnybara energikällor förändrar elmarknaden och tvingar elkraftsbolag att se över sina policyer för debitering av kunderna. Potentiella förändringar bör dock baseras på en noggrann förståelse av kundernas nuvarande beteenden. Här använde vi en teknik som kallas UMAP för att gruppera kunder med liknande konsumtionsprofiler. Analysen påvisade fem mindre kundgrupper med distinkta konsumtionsprofiler och en stor grupp där kundernas konsumtionsprofiler gradvis förändrades, främst från årliga till dagliga variationer. De identifierade grupperna var relevanta för att förstå kundernas profilomkostnader, det vill säga huruvida en kund använder elektricitet när den är billig eller dyr.

NOVIA
UNIVERSITY OF APPLIED SCIENCES

MINISTRY OF
EDUCATION AND CULTURE
FINLAND

This work was performed within the Flexible Energy Systems Integration and Optimization (FESIO) project. FESIO is a joint project between Novia University of Applied Sciences and Vaasa University of Applied Sciences. The project was financed by the Ministry of Education and Culture in Finland and aims to facilitate the green transition by helping companies utilize new AI-based tools.

Publisher: Novia University of Applied Sciences, Wolffskavägen 35 B, 65200 Vasa, Finland

Novia Publications and Productions, series R: Reports 13/2023
ISSN: 1799-4179,
ISBN: 978-952-7526-28-6 (online)
Layout: Johan Westö

# Contents

# Introduction

The electricity market is changing due to the increasing amount of energy from renewable energy sources like wind and solar (see Appendices A and B). Traditionally, consumers used electricity when needed, and plant operators adjusted the energy produced accordingly. However, the increased reliance on wind and solar power reduces our possibility to control the rate of power produced while also increasing fluctuations in available power and price. Consequently, there is an increasing interest in trying to modify consumer behavior through new policies that would shift the time of consumption to times when more power is available.

Policy changes should nonetheless be based on a thorough understanding of the consumers' actual consumption profiles. Such profiles have been studied since the 80s in Finland as part of the Finnish load research project (Mutanen, 2018; Seppälä, 1996). Electricity companies today assign a profile template, also known as an index series, to each customer to approximate the actual consumption profile (Trimble, 2020). In practice, however, the electricity companies might have had to make educated guesses when assigning a template to consumers. The reasons are that the templates may need to be updated as they are 30 years old, and consumers might have upgraded their heating systems or made other load changes afterward. The end result being that a reliable comprehensive overview of the customers' actual consumption profiles is often lacking.

Companies nonetheless have access to historical records of hourly consumption values for their customers. These records usually date back to the late 2000s when automatic meter reading systems started to become the norm. Researchers and electricity companies in Finland have since investigated the possibilities of updating the standard consumption templates (index series) based on actual measurement data as well as reassigning consumers to better-matching templates (Mutanen et al., 2011, 2017; Räsänen et al., 2010; Selenius, 2020). Multiple ways of clustering (grouping) consumers based on yearly consumption profiles have also been proposed, and their impact on load estimation accuracy has been assessed (Mutanen et al., 2017). However, clustering alone is insufficient for a comprehensive understanding of the customers and their profiles. The assigned cluster category fails to carry information on how various customers and their profiles relate to each other. We, therefore, instead utilized recent advances in visualizing high-dimensional data by basing or analysis on the Uniform Manifold Approximation and Projection (UMAP) technique (McInnes et al., 2020). UMAP lets us overcome the drawbacks of classical clustering approaches by providing a comprehensive overview of the customer base while also grouping customers with similar consumption profiles. The resulting analysis lets us 1) identify customer groups with similar consumption patterns, 2) compare groups against other known customer information, and 3) visualize which groups tend to use electricity when it is cheap (plentiful supply) and expensive (sparse supply), respectively.

# Methods

We obtained a batch of consumption data to analyze from a Finnish electricity company. The batch contained hourly measurements within the time interval from 1.1.2016 to 20.9.2022 as well as metadata for each customer. The metadata included: the number of phases connected (1 or 3), the main fuse size, a group ID for reporting purposes, a load model ID (consumer or producer), a pseudo-anonymized transformer district ID, the assigned index series, and the tariff in use.

## Data quality and selection

The raw measurement data was delivered as text files. In addition to the measured values for a specific customer, these files also included a timestamp (UTC) and a measurement code indicating whether the value was measured, estimated, uncertain, or missing. 96.6 % of all values were measured, 1.8 % were missing (zero), and the remaining 1.6 % were estimated. In total, the data included measurements from 3998 distinct customers. However, 15 customers suspiciously had their measurements spread out over multiple text files, and at least 11 of these showed signs of being merges of two distinct customers by having two measurement values for each hour. The second measurement often had a measurement code corresponding to missing, whereas the first value looked normal for the whole interval. For this reason, we ignored the second recording from these 11 customers, but otherwise, all values were included as the remaining measurements with missing or estimated status codes were so few.

### UMAP analysis

The UMAP analysis was done in Python using the official UMAP implementation (umap-learn v. 0.5.3). We largely used default parameter values, and the exact values for provided parameters were as follows: `n_neighbors=15`, `n_components=2`, `min_dist=0`, `n_epochs=500`, and `init="spectral"`.

## Results

We analyzed data from a period of four years, ranging from January 2018 to December 2021. The data set included hourly electricity measurements from 2687 consumers with a yearly consumption above 1000 kWh during all four years. The data for each consumer thus corresponds to a time series consisting of 35064 hourly measurements. The time series were normalized prior to the UMAP analyses by dividing each time series by the total consumption. This was done to force the UMAP analyses to compare consumption profiles (when electricity is used) rather than the quantity used.

### Visualizing the data based on consumer profiles

The individual time series are challenging to compare one by one. We, therefore, organized the data by performing a dimensionality reduction mapping called UMAP. The general idea is to convert the 35064 measurements from each customer into only two values, but in such a way that customers with similar consumption profiles end up getting two numbers that are similar to each other. This lets us represent each customer as a point in a scatter plot, where points (customers) with similar consumption profiles end up near each other. Figure 1a shows the resulting scatter plot consisting of one big cluster and five smaller ones. The smaller clusters can be identified based on their profiles and additional customer data as corresponding to: street lights (two clusters with different time schedules), late-evening consumers, industries and service providers with a fixed 5-day work week, and finally, prosumers with sporadic day-time production peaking during the summer (identified based on their load model ID). The average yearly profiles for all consumers in these smaller clusters are shown in Figure 1b to f together with the average consumption during one example week during September 2020.

The big cluster similarly contains a lot of internal structure with customers at various places exhibiting very different consumption profiles. The clearest differences can be seen along the curved black path running roughly vertically within the big cluster in Figure 2. At the top, we find customers whose consumption is dominated by yearly variation (presumably by weather as the peaks vary from year to year), and as we move downwards, the profiles change towards being dominated by variation on a daily scale. Farthest down, we find customers with daily consumption peaks in the morning and in the evening throughout the year. These customers were identified as dairy farmers based on matching profiles in the index series from the Finnish load research project (Seppälä, 1996). Other specific customer groups likewise gathered at other edges of the big cluster. The top right corner, e.g., contained customers with yearly peaks during November and December, thought to be fur farmers as the timing coincides with the skinning period (Enlund, 2013), and the bottom right corner contained customers that primarily used electricity during the late summer, thought to be crop farmers as the timing coincided with the harvesting season and when crops are dried.

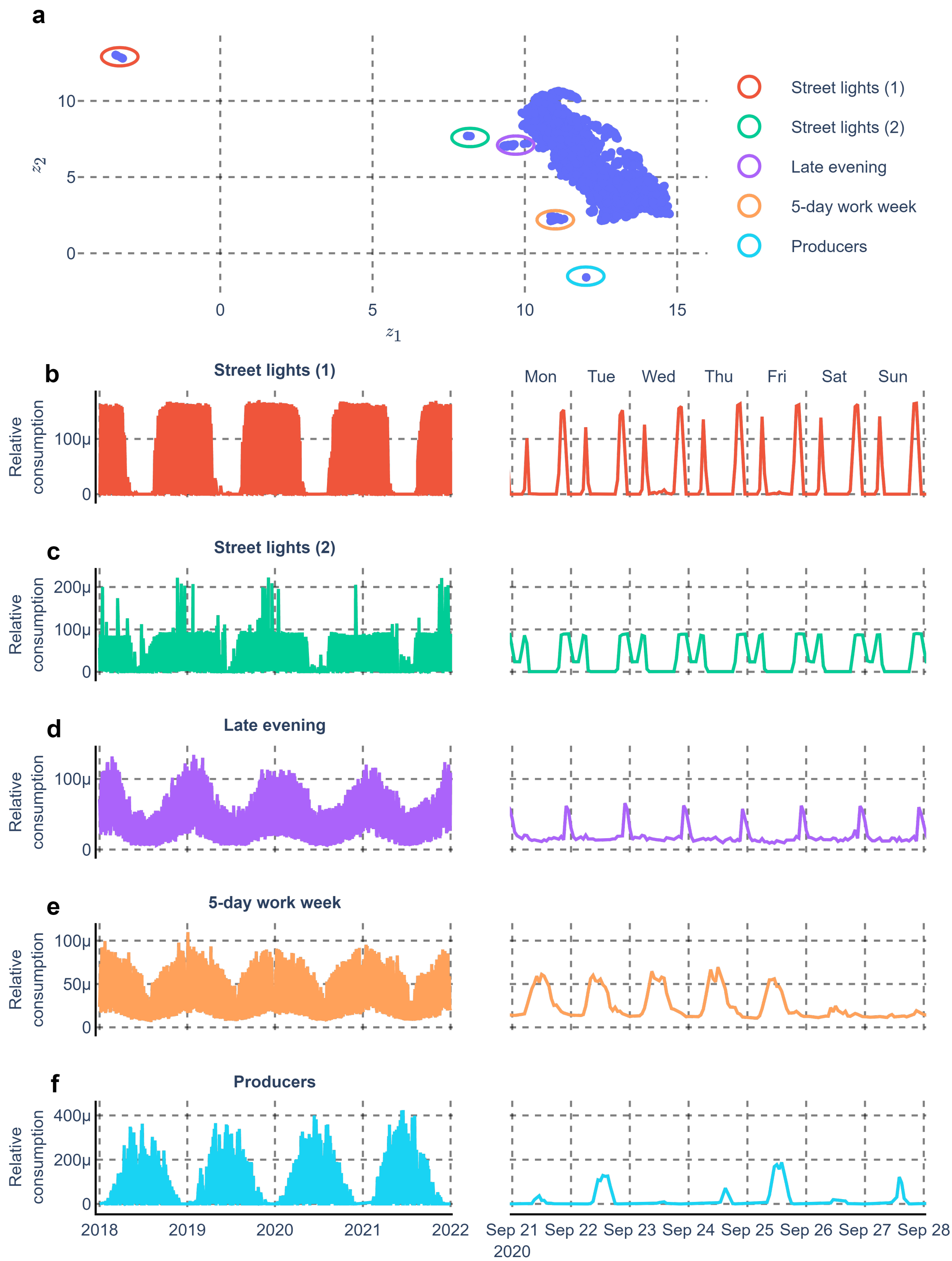



**Figure 1:** UMAP scatter plot with distinct clusters highlighted. **a** The UMAP analysis resulted in five distinct smaller clusters and one large. **b** to **f** The mean yearly profile for all customers within the corresponding (same color) circle in **a** as well as the mean profile during an example week in September 2020.

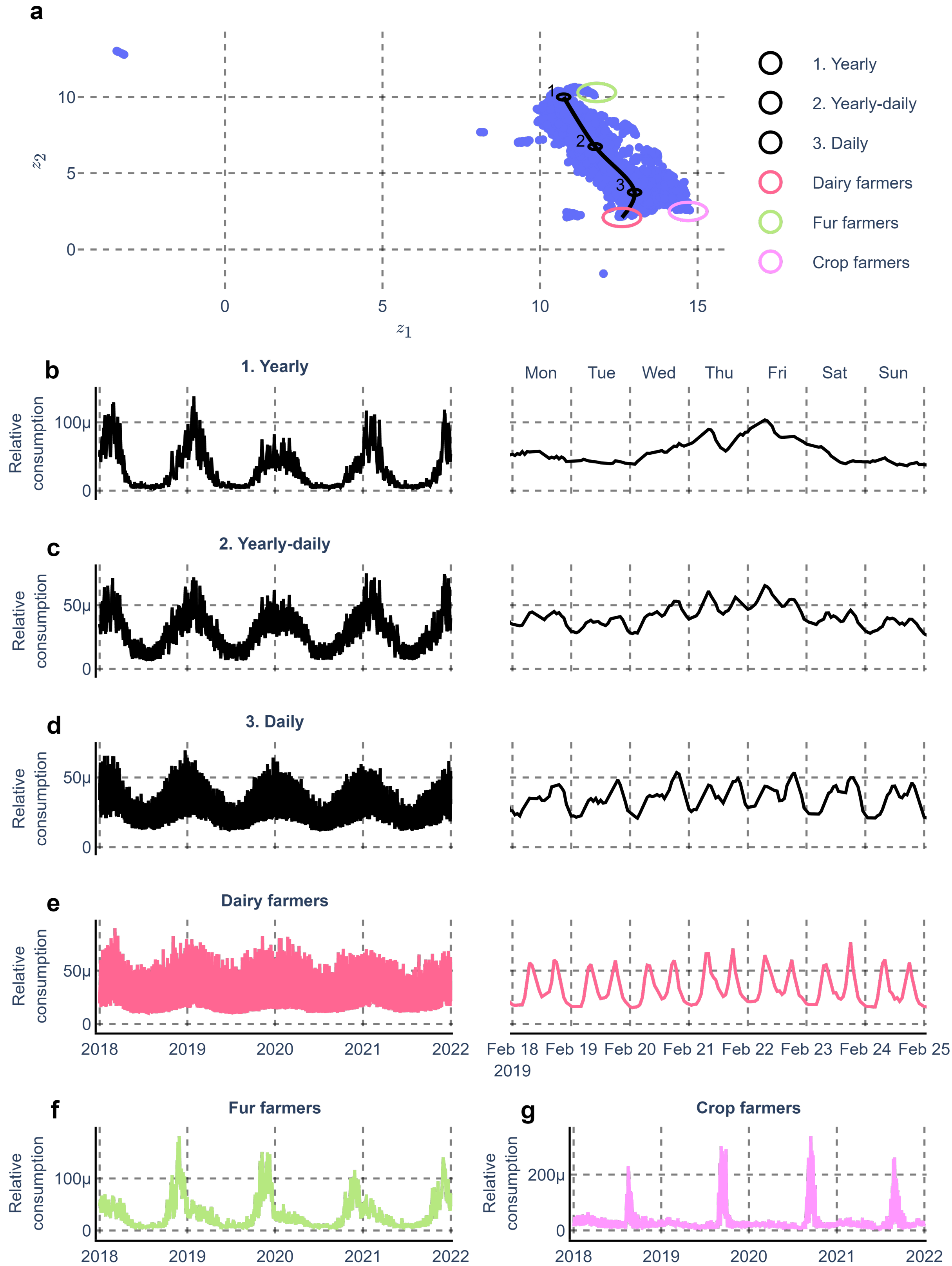



**Figure 2:** UMAP scatter plot with variation within the main cluster highlighted. **a** The large main cluster had an internal structure with customers exhibiting a shift from variation on a yearly basis to variation on a daily basis along the black line. **b** to **e** The mean yearly profile for all customers within the corresponding (same color or number) circles in **a** as well as the mean profile during an example week in February 2019. **f** and **g** The mean yearly profiles for all customers within the corresponding (same color) circles in **a**.

## Usage profile and consumer metadata

Every customer is associated with an index series that provides a template of the expected yearly consumption profile. These templates, originally developed for load estimation (Seppälä, 1996), should thus approximate the actual consumption profile. One would therefore expect that customers close to each other in the UMAP scatter plot (i.e., customers with similar consumption profiles) would have been assigned to the same index series, and that the index series would be clearly clustered in the UMAP visualization. However, this was not the case. Figure 3a shows how the four most commonly assigned index series (templates) are scattered all over the plot, indicating that the assigned index series is unrelated to the actual consumption profile. Instead, it appears that the assigned index series is mainly related to the yearly consumption. Figure 3b shows histograms over the yearly consumption for the same four most commonly used index series, and these show a much clearer pattern with each index series centered around separate yearly consumption values.

Tariffs represent another variable that one would expect to be related to the consumption profile, and indeed, the late evening user group in Figure 1a and d all have the same tariff (code 1090), see Figure 4a. However, there appear to be roughly equally many customers with the same tariff that are spread out more or less randomly, possibly indicating that these have a sub-optimal tariff. The other three commonly used tariffs (codes: 1010, 1020, and 1090) exhibit no clear patterns in the UMAP scatter plot and appear to be again mainly related to the yearly consumption, as indicated by the histograms in Figure 4b.

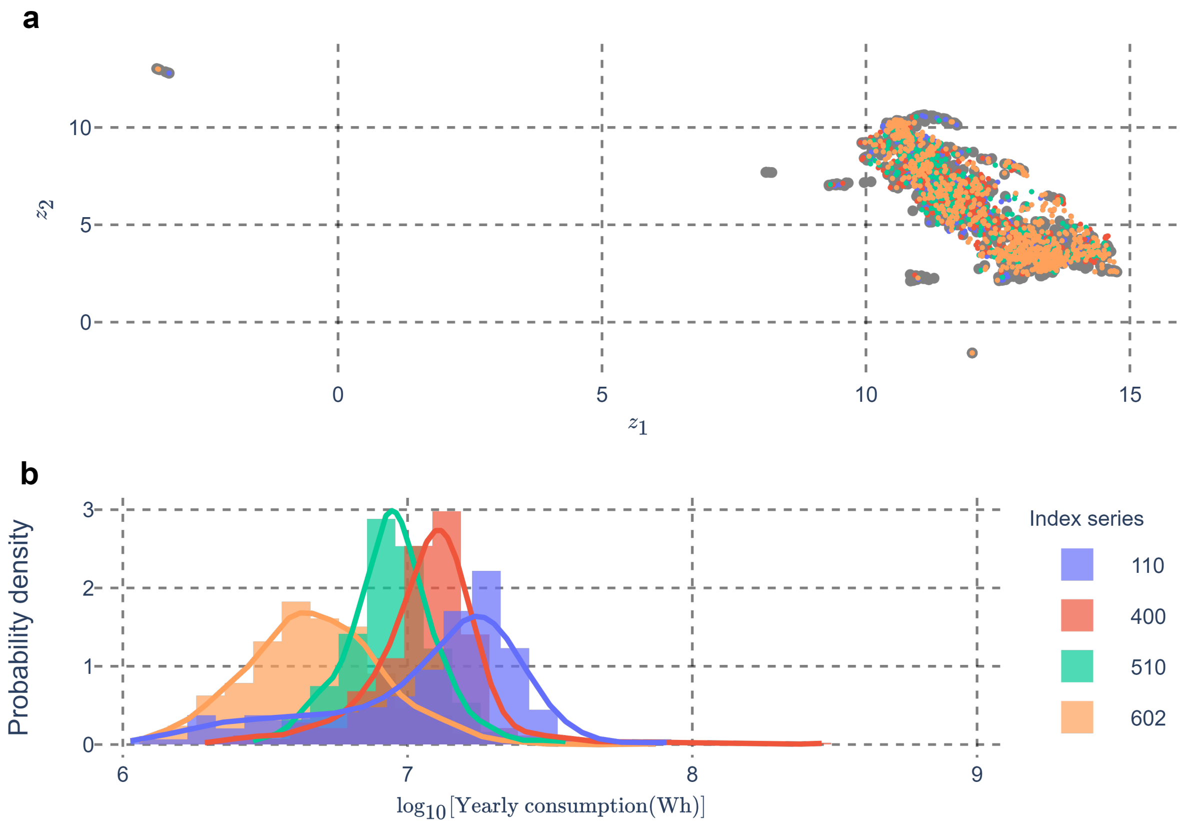



**Figure 3:** Comparison between the UMAP visualization and the assigned index series for each customer. **a** The UMAP scatter plot with customers having one of the four most commonly used index series highlighted. **b** Distributions illustrating the average yearly consumption for each of the index series highlighted in **a**. Index series descriptions: **110** = one family house, direct electric heat, water boiler < 300l, **400** = one family house, heat pump, **510** = one family house, dual heat, flat tariff, and **602** = one family house, no electric heat, electric sauna. The scatter points in gray denote consumers assigned to some other index series.

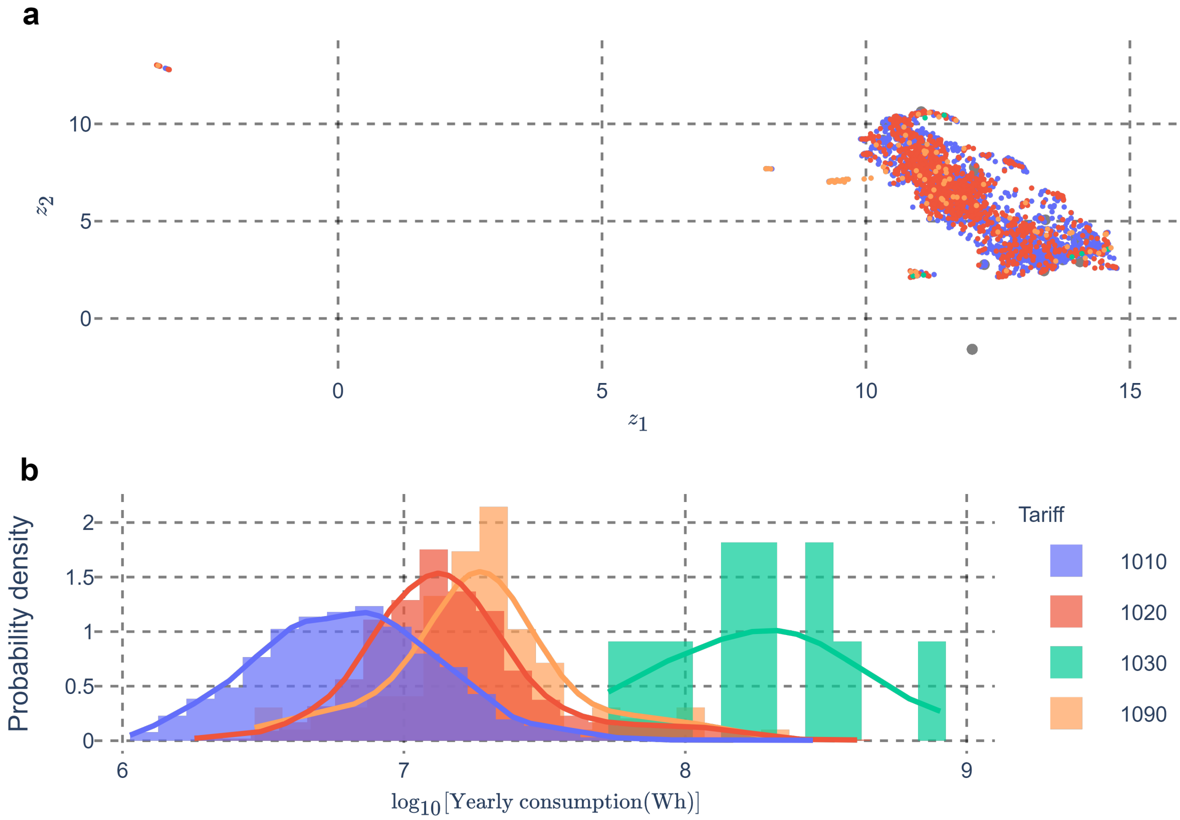


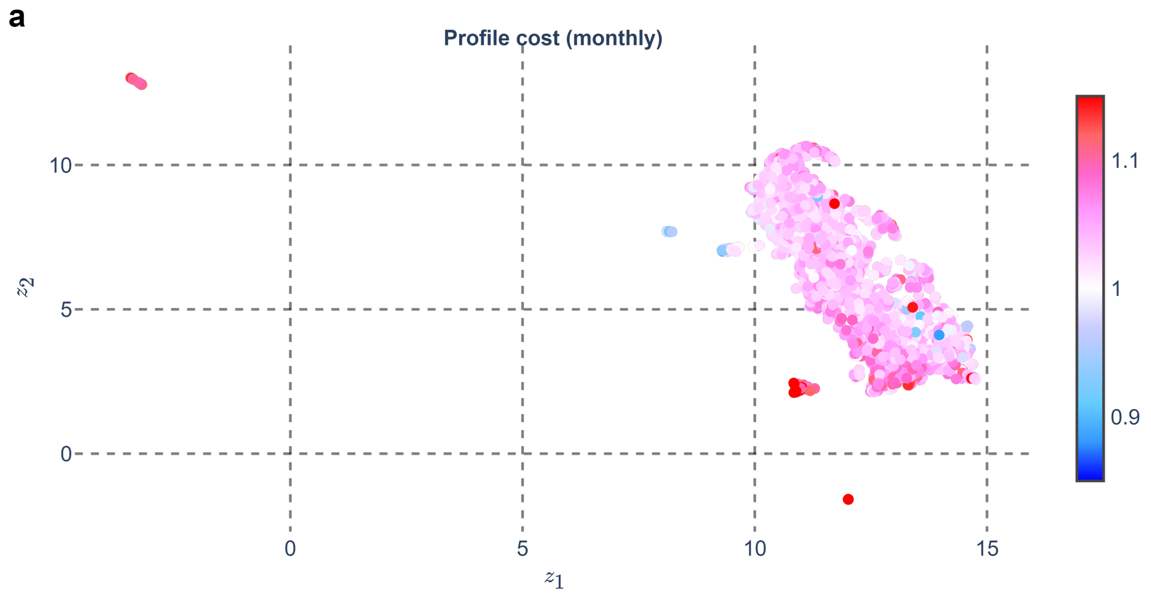

**Figure 4:** Comparison between the UMAP visualization and the tariff in use for each customer. **a** The UMAP scatter plot with customers having one of the four most commonly used tariffs highlighted. **b** Distributions illustrating the average yearly consumption for each tariff group highlighted in **a**.

## Profile costs

Electricity can be bought and sold ahead of time, meaning prior to the point in time when it is used. In such cases, electricity is often bought at a fixed monthly or yearly price, with the quantity bought divided uniformly over the time period (i.e., the price is the same for every hour in the period). It is, therefore, of interest to know how the consumption of various customer groups compares with respect to a flat consumption profile. This comparison was made using a *profile cost* defined as the ratio of the costs (spot price) for the real consumption profile versus a flat profile. A value larger than one ($> 1$) indicates that the customer used electricity when it was more expensive than the mean price for the period, and a value below one ($< 1$) indicates that the customer used electricity when it was cheaper than the mean price.

The profile cost was computed separately for a monthly and a yearly basis. Figure 5a first shows the monthly profile cost for all customers. Comparisons to Figure 1 and Figure 2 highlight that the customer groups "street lights (1)", "Producers", and "5-day work week" are the ones with highest profile costs. In addition, one can see that the profile cost increases along the yearly-daily black line in Figure 2. This indicates that it is the customers who have the majority of their consumption during the daytime that get a high monthly profile cost.

The customers with a high profile cost on a monthly basis also tended to have a high profile cost on a yearly basis. In addition, we found that the top and bottom right parts of the big UMAP cluster had high profile costs on a yearly basis, see Figure 5b. As highlighted in Figure 2, these are consumers that have a strong seasonal pattern in their consumption profiles. In short, it is thus daytime consumers and consumers with a seasonal pattern that have high profile costs on a yearly basis.



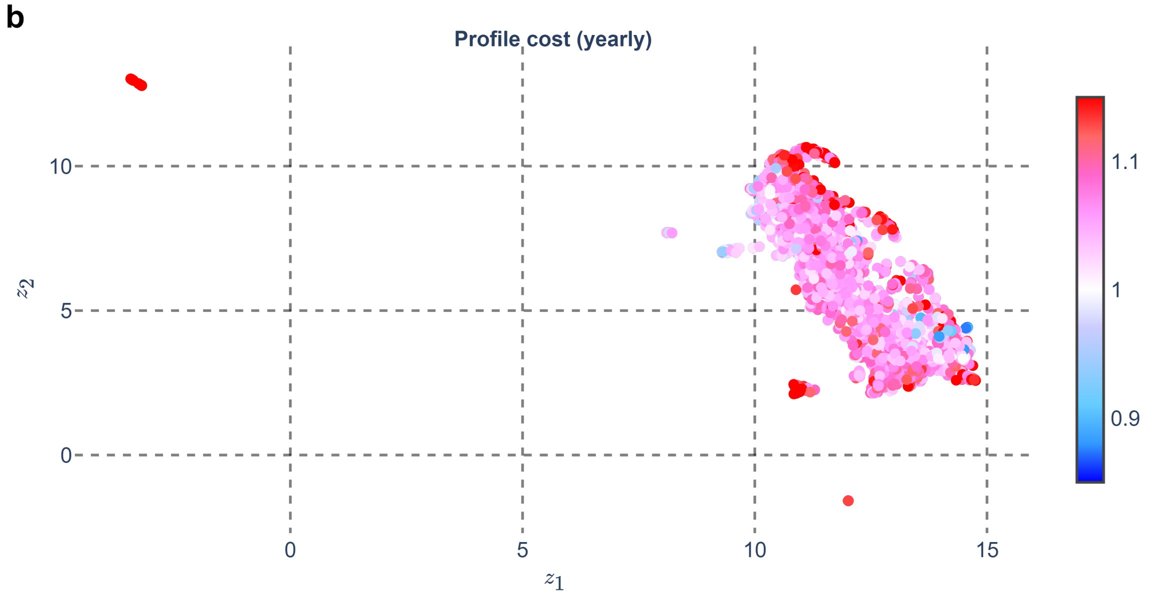




**Figure 5:** UMAP scatter plots with the customers' profile costs color coded. **a** Profile costs computed on a monthly basis, and **b** profile costs computed on a yearly basis.

## Discussion

Analyses of historical measurement data can help provide a comprehensive understanding of an electricity company's customers and how and when they use electricity. The importance of such information is growing daily as the electricity market changes due to the ever-increasing fraction of renewable energy sources such as wind and solar. Here, we utilized a technique called UMAP to provide a comprehensive visualization of a company's customers that automatically grouped customers with similar consumption profiles.

The analysis highlighted five smaller groups with distinctive consumption profiles and a large group where the customers' consumption profiles changed gradually, primarily from yearly to daily variations. The existing customer information did not correlate with their consumption profiles (except for customers with a night tariff). Instead, tariffs and assigned template profiles (index series) seemed to be mainly related to the customers' yearly consumption. The customers' profile costs nonetheless correlated well with their position in the UMAP visualization, with neighboring customers having a similar profile cost. The customers with the highest profile costs were those who had their consumption fixed to the daytime (certain street lights, customers with regular office hours, and some dairy farmers) or to specific times during the year (heating season and fur and crop farmers). Daytime consumers obtained high profile costs due to these customers using electricity during the hours of the week when it tends to be most expensive, see Appendix A. Customers with a high heating load and the fur and the crop farmers are less straightforward to explain. Historically, the price of electricity has tended to be more expensive during the late summer when compared to the average price for the same year (see Figure A-1b and Figure A-1c), which would explain why crop farmers ended up with a high yearly profile cost. The same argument can be made for fur farmers and consumers with a high heating load during the cold season, but in this case, one ought to be more careful as both 2020 and 2021 are extreme years with large price spikes in November and December.

The historically high price of electricity that we have observed lately has motivated people and companies to either use less electricity or to shift usage to other hours of the day. Therefore, it would be interesting if one could do an analysis similar to the UMAP-based analyses done here, where it would be possible to follow how customers adjust their behavior over time. Future analyses will likely reveal customer groups with profiles matched to the amount of renewable energy currently in the grid.

## References


Enlund, L. (2013). *Skin* [BA thesis]. Yrkeshögskolan Novia. https://www.theseus.fi/bitstream/handle/10024/60937/Enlund_Lina.pdf?sequence=1&isAllowed=y

Kulla, I. (2023, August 23). Personal communication.

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform manifold approximation and projection for dimension reduction.* arXiv: 1802.03426 `[stat.ML]`.

Mutanen, A. (2018). *Improving electricity distribution system state estimation with amr-based load profiles* [Doctoral dissertation, Tampere University of Technology]. https://trepo.tuni.fi/bitstream/handle/10024/114074/mutanen_1529.pdf?sequence=1

Mutanen, A., Järventausta, P., & Repo, S. (2017). Smart meter data-based load profiles and their effect on distribution system state estimation accuracy. *International Review of Electrical Engineering (IREE)*, *12*, 460. https://doi.org/10.15866/iree.v12i6.13419

Mutanen, A., Repo, S., & Järventausta, P. (2011). Customer classification and load profiling based on amr measurements.

Räsänen, T., Voukantsis, D., Niska, H., Karatzas, K., & Kolehmainen, M. (2010). Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, *87*, 3538–3545. https://doi.org/10.1016/j.apenergy.2010.05.015

Selenius, H. (2020). *Sähkönkulutuksen tuntimittauksien hyödyntäminen tyyppikäyrien muodostamiseksi* [msthesis]. Aalto University. https://aaltodoc.aalto.fi/bitstream/handle/123456789/44326/master_Selenius_Henri_2020.pdfsequence=1&isAllowed=y

Seppälä, A. (1996). *Load research and load estimation in electricity distribution* [Doctoral dissertation, Helsinki University of Technology]. https://www.vttresearch.com/sites/default/files/pdf/publications/1996/P289.pdf

Trimble. (2020). *Trimble NIS power system analysis theory guide for finland* (tech. rep.).

## Appendix A: Spot price data in Finland

The spot price for electricity exhibits a clear trend of increasing volatility, see Figure A-1a. It's worth noting that this trend was already present in 2021, before the Russian invasion of Ukraine. Thus highlighting that the trend is connected to other current events like the ongoing transitions towards renewable energy sources, even though the war in Ukraine obviously also impacted the electricity market.

Despite the fluctuations, there are still certain price patterns that appear to be persistent. Figure A-1b and Figure A-1c display the average price for each month, relative to the yearly average. Months with values above 1 have a higher price than the yearly average, and months with a value below 1 have a lower price. Two trends are noteworthy: 1) Spring is when electricity is cheaper, and towards the end of summer, it becomes more expensive. 2) 2021 and 2022 are exceptional years, with electricity prices being anomalously high towards the end of the year. A similar analysis for weekdays shows that the price is strongly linked to the working week (see Figure A-1d): the price tends to be higher during daytime on working days. Specifically, the price starts increasing at around 6 am, peaks at around 8-9 am, decreases slightly during the day, peaks again at 6 pm, and then decreases to its lowest point in the middle of the night.

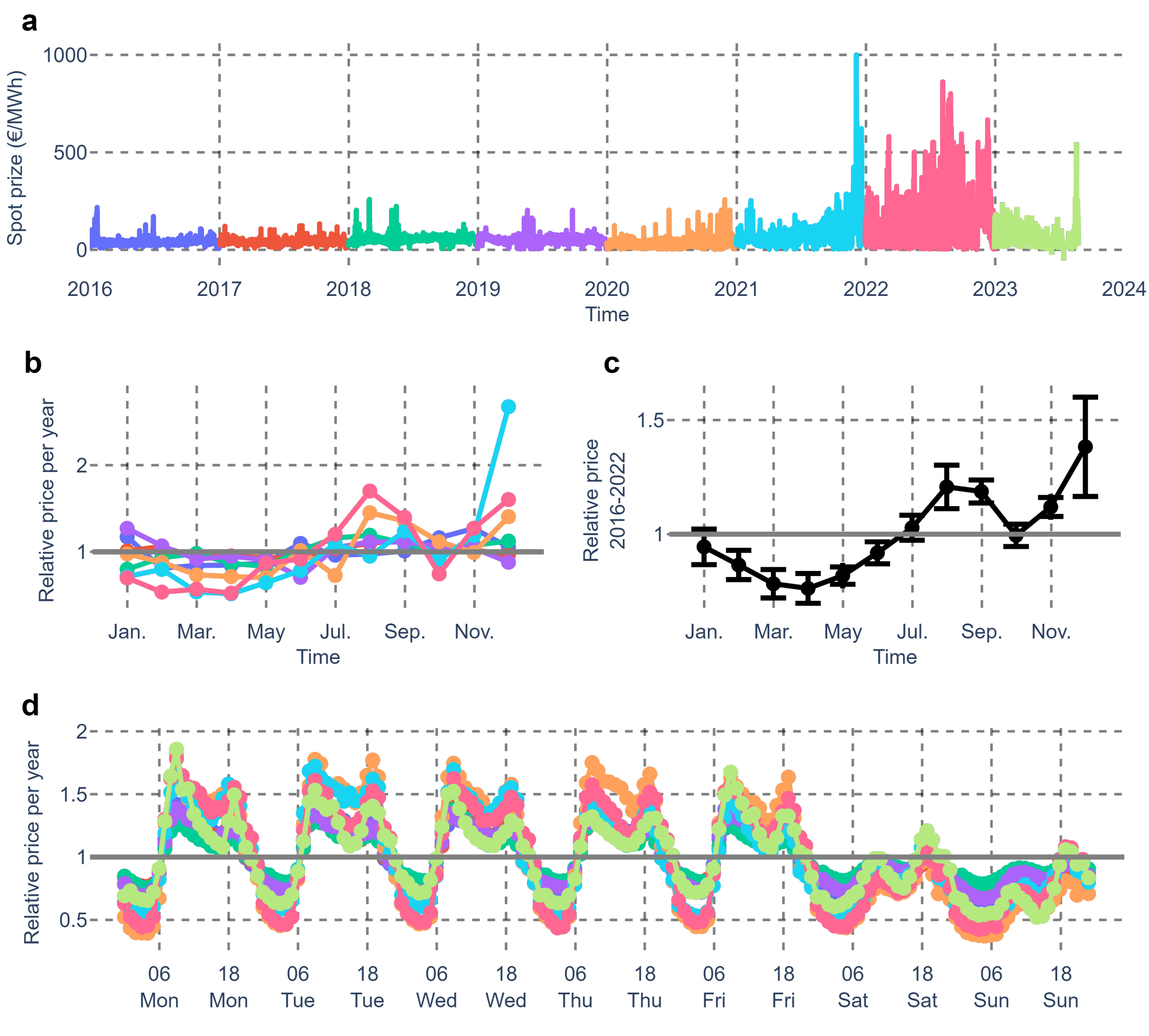



**Figure A-1:** The day-ahead spot price in Finland. **a** Hourly spot prices from January 2016 to August 2023. **b** and **c** Monthly relative prices with respect to the yearly average price. Relative prices for individual years in **b** (2016 to 2022) and averages and SEMs over all years in **c**. **d** Hourly relative prices with respect to the yearly average for every day of the week. Data obtained from ENTSO-E's transparency platform.

## Appendix B: Electricity consumption and production in Finland

The transition towards renewable energy sources is evident in Fingrid's production and consumption data. Wind power has been steadily increasing and can constitute almost 50 % of Finland's electricity production during certain hours, see Figure B-1a. On a monthly basis, this figure drops to ≈ 25 % due to fluctuations in the wind. Production is, nonetheless, concentrated to the colder months when demand is higher (see the yearly bumps from 2020 onwards for wind power in Figure B-1b). The increasing amount of wind power also has a cannibalizing effect on the price: the price of electricity tends to be lower when it is windy. This effect can be seen as a negative correlation between the price and the amount of wind power, and as shown in Figure B-1c, this correlation has been getting stronger over time as more and more wind power has been installed.

The connection between price, consumption, and production has also changed, see Figure B-1c. Historically, the price of electricity was very correlated with both production and consumption, except for one or two spring months when melt-water filled the water basins for hydropower in Scandinavia and forced hydropower production irrespective of demand and price (Kulla, 2023). Today, total production no longer appears correlated with the price, and even the correlation between price and consumption has decreased lately. The reasons are probably many and complex, but the transition towards green energy sources, the commissioning of Olkiluoto 3, and the war in Ukraine have probably all contributed.

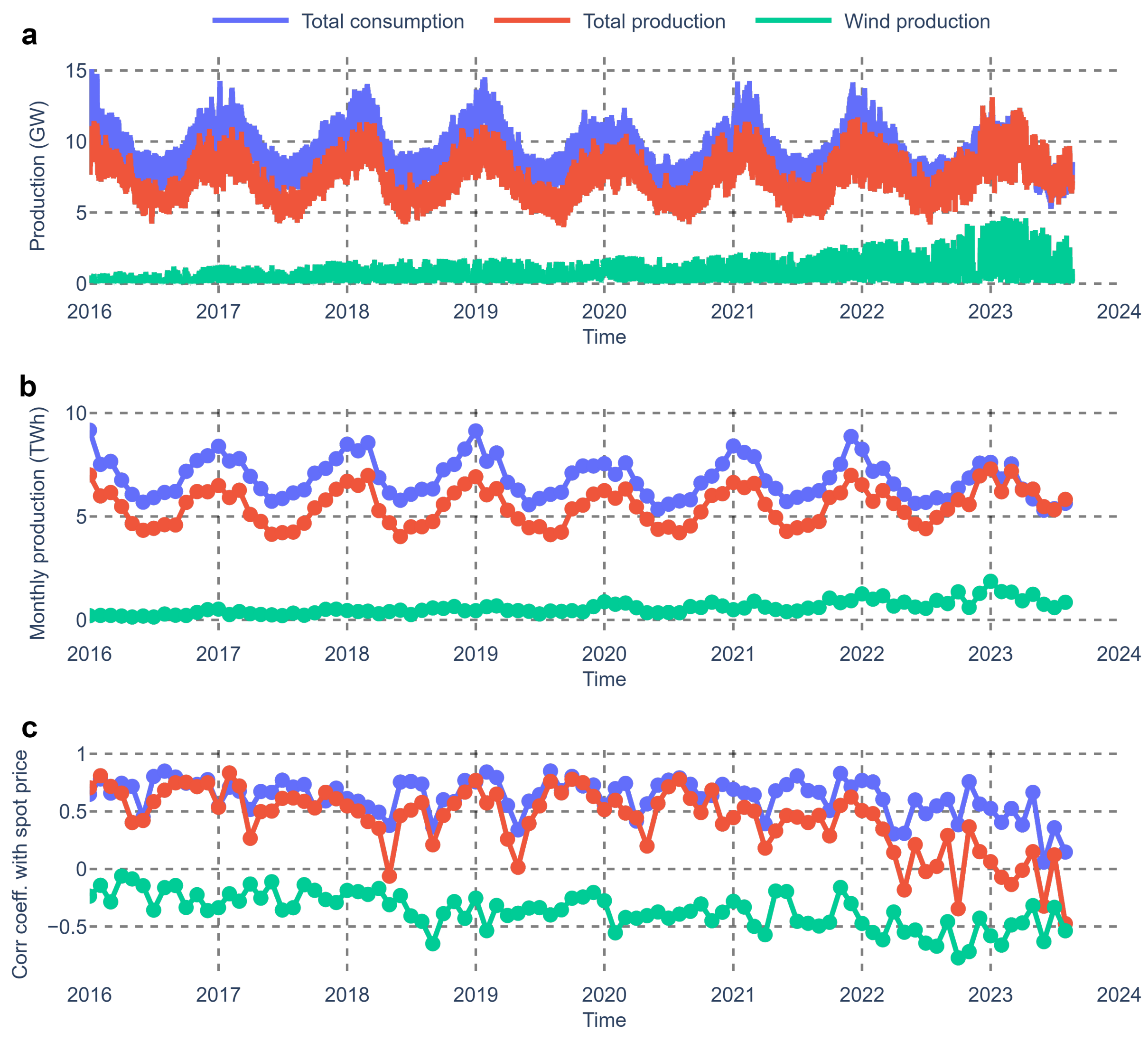



**Figure B-1:** Production and consumption data for Finland from January 2016 to August 2023. **a** Hourly consumption and production (wind and total). **b** Monthly consumption and production (wind and total). **c** Monthly correlations between price and consumption, and between price and production (wind and total).

NOVIA
UNIVERSITY OF APPLIED SCIENCES

## Intelligent Systems Institute

The intelligent systems institute @ Novia gathers collective knowledge from multiple experts with the intent to build and strengthen collaboration with companies. We want to get experts to tackle your relevant research challenges, and to provide hands-on assistance on real problems. We think that successful collaborations rely on researchers understanding the needs and challenges you face. Hence, we are more than happy to visit you onsite and to start exploring the opportunities you have.

## Novia UAS

Novia's area of activity is located along the Finnish coast and the University of Applied Sciences operates in Vaasa, Turku, Raasepori, and Pietarsaari. With 5000 students, and a workforce consisting of about 350 people, Novia is the largest Swedish-speaking UAS in Finland. Novia UAS has a close working relationship with the business community and an active alumni network.

**NOVIA UNIVERSITY OF APPLIED SCIENCES**
Tel +358 (0)6 328 5000
www.novia.fi

**ADMISSIONS OFFICE**
PO BOX 6
FI-65201 Vaasa, Finland
Tel +358 (0)6 328 5055
admissions@novia.fi

Read our latest publication at www.novia.fi/forskning/novias-publikationsserie
Visit the intelligent systems institute at www.novia.fi/intelligentsystems/home