Energy 288 (2024) 129805

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Correlation as a method to assess electricity users' contributions to grid peak loads: A case study

Carl Flygare *, Alexander Wallberg, Erik Jonasson, Valeria Castellucci, Rafael Waters

Uppsala University, Ångströmslaboratoriet, Lägerhyddsvägen 1, Uppsala, 752 37, Sweden

ARTICLE INFO

Keywords:
Electricity consumption
Flexibility
Power grid peaks
Smart grids
Time series correlation analysis

ABSTRACT

Flexibility has increasingly gained attention within the field of electrification and energy transition where a common objective is to reduce the electricity consumption peaks. However, flexibility can increase the risk of grid congestion depending on where and when and it is used, thus an overall system perspective needs to be considered to ensure an effective energy transition. This paper presents a framework to assess electricity users' contributions to grid load peaks by splitting electricity consumption data into subsets based on time and temperature. The data in each subset is separately correlated with the grid load using three correlation measures to assess how the user's consumption changes at the same time as typical grid peak loads occur. The framework is implemented on four different types of business activities at Uppsala municipality in Sweden, which is a large public entity, to explore their behaviors and assess their grid peak load contributions. The results of this study conclude that all four activities generally contribute to the grid peak loads, but that differences exist. These differences are not visible without splitting the data, and not doing so can lead to unrepresentative conclusions. The presented framework can identify activities that contribute the most to unfavorable grid peaks, providing a tool for decision-makers to enable an accelerated energy transition.

1. Introduction

Over the past decades – and especially during recent years – the concept "energy transition" is becoming increasingly common.[1] For industrialized countries it may include replacing fossil fuels, while for developing countries to skip fossil fuels entirely. Both of these energy system developments point at the same outcome — strengthen the energy infrastructure, electrify processes, and produce more electricity, preferably from fossil-free sources close to the user. This may also increase the energy security of the specific location, which is another key aspect in the energy transition process [1–4]. Being one of the most significant challenges of the 21st century, the Paris Climate Agreement especially emphasizes decarbonization of the global energy systems with goal 7 "Affordable and clean energy". This challenge is also present in other global issues, a recent example being the COVID-19 pandemic accentuating local self-sufficiency of resources [5]; or the ongoing war in Ukraine which, apart from its effect on people's lives, has emphasized geopolitics leading to pressure on phasing out Russian gas from the EU energy system. Such events constitute effects that affect the entire world, and the International Energy Agency has called for an acceleration of renewable energy capacity and energy efficiency [6].

In Sweden, electricity consumption is projected to increase substantially in the upcoming decade. The Swedish Energy Agency (SEA) estimated in their latest long-term scenario an increase between 25%–65% [7], and several influential national organizations show similar or higher numbers [8–11]. The Swedish transmission system operator (TSO) Svenska Kraftnät (SvK) has, in turn, named their latest development plan "Towards a doubled use of electricity" [12]. As supply, Sweden has a goal of 100% renewable electricity production by 2040 [7–11,13,14]. In this ongoing energy transition process *flexibility* has become an important concept that addresses the ability of an electricity user to adapt their consumption. This is an important part of concepts such as "smart grids" where controlling as many aspects of the energy systems as possible is desirable. Traditionally, the consumption side has been seen as deterministic and stiff while the centralized production side was the control instrument, but with more distributed energy resources, such as renewable electricity production, energy storage possibilities, or smart meters, this view is changing as decentralized control opportunities emerge. As the electricity consumption increases, and include new types of loads such as EV-charging, the grid capacity is approaching its limit. In several subsequent cases, flexibility may be encouraged which commonly results in users trying

* Corresponding author.
E-mail address: carl.flygare@angstrom.uu.se (C. Flygare).

[1] Google Trends shows an index change from 25 in 2019 to 100 in late 2022 while Google Ngram Viewer shows an ongoing exponential growth since 2000.

https://doi.org/10.1016/j.energy.2023.129805
Received 24 April 2023; Received in revised form 30 September 2023; Accepted 25 November 2023
Available online 30 November 2023

Nomenclature

$\hat{\mathbf{Y}}$	Vector with linear regression values
μ	Mean value
ρ_k	Kendall correlation
ρ_p	Pearson correlation
ρ_s	Spearman correlation
σ	Standard deviation
σ_R	Standard deviation of ranked variable
τ	The % of data points that show a positive correlation using ρ_k
τ_B	Kendall correlation using including ties
$\mathbf{X}$	Vector with time series data points
^oC	Degree Celsius
CC	Correlation constant
d	Difference in ranks (between two variables)
N	Number of data points
N_c	Number of concordant pairs
N_d	Number of discordant pairs
n_x	Number of tied pairs only in x
n_y	Number of tied pairs only in y
R^2	Coefficient of determination
T_m	Mean temperature
T_{xx}	The temperature of the xx:th hour of the day using wintertime
x_i	Separate data points at index i for specific variable

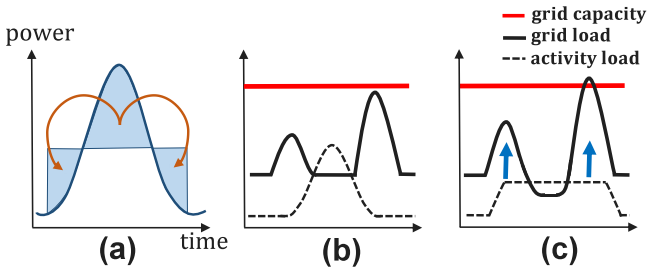


Fig. 1. Simplified illustration of peak shaving, (a), which might increase the grid's peak loads as shown from (b), before peak shaving, to (c), after.

to lower, or "shave", their consumption peaks. The incentives for doing so have mainly been economic: consume less when the price is high, or keep the maximum power usage below a certain tariff, but the reasons may differ [15]. But by doing so, the grid perspective is not considered. If flexibility is implemented as peak shaving without any system perspective it may instead increase the grid's peak load leading to transmission capacity issues as shown in Fig. 1, and such congestion may halter electrification and in turn an efficient energy transition.

From the grid's operational viewpoint, it is essential *where*, *when*, and *how (much)* electricity is consumed as a balance always is needed between consumption and production — while not exceeding the grid's transmission capacity. The focus on users' contribution to grid peak loads is a topic that has received less attention, but is imperative to know. This paper presents a framework that is intended to be used as a tool for energy strategists and urban planners, e.g. at public organizations or grid companies that handle many different electricity users, to assess the user's contributions to grid peak loads and use the result in long-term energy planning goals. Uppsala municipality (UM), where a case study is implemented using the presented framework, is a large public entity in Sweden and has around 3000 separate electricity subscriptions. How these use electricity and contribute to peaks is not known on a larger scale, thus where time and resources should be allocated for the energy transition processes needs decision-based support. Increased knowledge in this topic can be used to evaluate the feasibility of e.g. energy storage, local electricity production, and flexibility, or if grid extensions might be the only solution. Extending the grid is however costly in terms of capital and time, thus alternatives that may increase the grid usage effectiveness should be sought firsthand. One such way is to study the relationship between electricity users and the grid load peaks, and a viable means is by using time series correlation.

1.1. Correlation studies on electricity consumption

Previous studies on energy and electricity have analyzed the correlation between electricity consumption and several other variables. Some examples are correlation with ambient temperature [16], electricity prices [17], heat demand [18], wind speed [19], local production [20], and other consumers to improve predictions [21]. It also surfaces in gray literature, e.g. between grid capacity and electricity price [22] or grid load and system price [23]. On a broader energy topic level studies have analyzed wind power output correlation between different European countries [24], or the complementarity of renewable energy sources to balance local demand and supply [3,25].

A recent study investigated how the temporal aspects affect the correlation between solar irradiation and wind speeds, and showed that it differs over different time horizons [26]. This motivates that similar effects might be present in other weather-dependent parts of the energy system, such as electricity consumption, which in contrast also is affected by human behavior. Correlation has also been spatio-temporally addressed by modeling load flows in systems with increasing amounts of distributed renewable production and EV-charging consumption. Traditionally these analyses have a deterministic approach that neglects uncertainties, but over recent years probabilistic approaches have become more common [27]. Such studies however tend to use only Pearson correlation and correlating all data without any partitions, see for instance [28]. The non-linearity of certain exogenous variables' correlation may be mentioned as in [29,30], but no other correlation measures are suggested and data is correlated jointly without analyzing details in the load variations. Another study did correlate grid peak loads with exogenous weather variables during separate months, but it only used daily resolution data [31].

As shown by previous studies time and temperature can affect the correlation results, but these relations need to be further studied. Higher temporal resolution (from daily to hourly e.g.) may cause the correlation to decline [32,33]. This can be compared with correlating yearly data as one set or as separate monthly, weekly, daily, hourly, etc. sets. Data showing seasonality trends have been shown to cause different (lower) correlations if not handled properly [18]. As heat demand increases with temperature, lower temperatures generally increase the use of electricity. Taking such effects into account is thus suggested, and one way is to split time series data based on meaningful features into subsets in order to not mix behaviors that take place at different times and in different contexts. This can also separate different non-linearities in the data if split in a thoughtful way. For example, electricity consumption during lower outdoor temperatures could be split into one group as it typically decreases with increasing temperature due to less heat demand, while electricity consumption during hotter temperatures could be split into another as it typically increases with increasing temperature due to cooling needs.

Most of the previous studies on this topic aim to forecast a future load time series or predict the peak load, where correlation is a smaller part of the overall method and is not analyzed in detail. Potentially different trends within time series data at different times need to be further studied when it comes to electricity consumption. Previous studies also tend to not split data and only use Pearson, which is a linear correlation measure. Depending on the studies' aim and goals

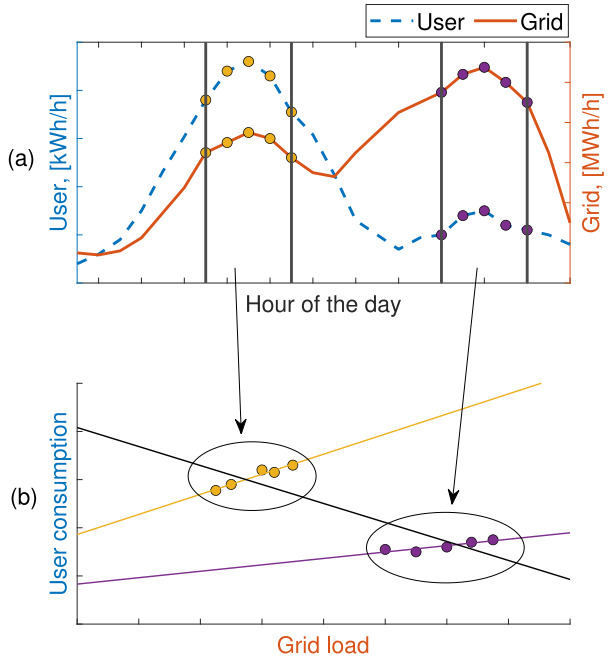


Fig. 2. A user could have a positive correlation with grid peaks both morning and evening during a day for example, but correlating these together would result in the opposite. In **(a)** two-time series are shown with a corresponding scatter plot in **(b)** for the data points in the intervals around the grid peaks. In **(b)** a linear regression is shown both for each group seperately, indicating a positive correlation, and for both groups together, indicating a negative correlation .

this is principally not wrong, but if not looking specifically for linear relationships and inspecting the data to see this is commonly the case – both for short-term and long-term – the result may lead to wrong conclusion as illustrated by Fig. 2. In Fig. 2 **(a)** two daily time series with hourly values are shown and in Fig. 2 **(b)** the data from the vicinity of the first and second peaks are extracted together with linear regression lines. If the peak's data points are correlated separately the result is strongly positive, close to one, but if correlated together it becomes the opposite. This is also known as Simpson's paradox. Correlating all data in Fig. 2 **(a)** instead gives a result close to 0, even though these time series behave very similarly. Splitting data into subsets can distinguish different and more intricate patterns, and to the knowledge of the authors no previous study has used time series correlation to assess an electricity user's contribution to peak loads.

Much of the current research focuses on machine learning methods that may be more or less complex. The simplicity and interpretability of a method are decisive to whether or not it will be used — especially if to be communicated with and used by non-experts within e.g. a public entity. To only visually inspect data and draw conclusions is possible, but it can easily become subjective and has low granularity. This study's rationale stems from this context, and it will present a framework to split data into subsets based on time and temperature. Three different time series correlation measures are then described that are used on each subset separately. The framework is then implemented on four electricity users representing different types of activities from UM to assess their contributions to grid load peaks, based on network area data from the Swedish TSO. The advantages of selecting these variables are: (1) electricity users can easily access their consumption data through their distribution system operator (DSO), and (2) the network area data do not require any model of the grid's topology. Thus, correlation can be conducted without any prior knowledge of the grid's actual structure and parameters — which is difficult to acquire due to confidentiality. The DSO in the studied city has past years acknowledged capacity issues, and as both the city and EV-charging are expected to grow substantially the local grid capacity is an important topic. As UM operates a large portfolio of activities and has responsibility for the city's energy transition, increased knowledge of how their activities contribute to the local grid peaks is valuable information in strategic decision-making.

1.2. Aim of the study

The preceeding section detailed that previous correlation studies risk creating a simplified image, which sometimes can be unrepresentative of the actual relationship between electricity consumption and other variables. The aim of this paper is to improve such analyses by presenting a framework and implementing it to assess different electricity users' contributions to local grid peaks. The contribution of this paper can be summarized as follows:

1. We suggest a framework to split and analyze data separately based on meaningful features to find potentially different behaviors and avoid mixing them, which can create an inaccurate image of their relationship.
2. We analyze real data to show that meaningful differences exist, and that they are missed if the data is not split.
3. We assess the contributions of four different types of activities to local grid peaks using the presented framework and discuss it in the local context, showing that the result can be an important input to energy transition processes where smart grid solutions are considered.

A limitation of the study is that the correlation results only depict the specific relationship based on where and when the data is from. The case study data are from 2021, and even with meaningful splits of the data the results cannot be generalized to other years or places. The framework can however be used for a longitudinal study given more data.

The rest of the paper proceeds as follows: Section 2 outlines the correlation measures. Section 3 describes the data pre-processing and exemplifies the subset creation with data from the case study. Section 4 presents correlation results. To conclude, Section 5 discusses the result together with the local context, and Section 6 summarizes the conclusions.

2. Correlation measures and considerations

The correlation coefficient (CC) is a general and mature measure that is relatively simple and has a clear correspondence between input and output. It does not require training on specific data, nor have hyperparameters. The three correlations implemented are Pearson, being the most commonly used, in addition to Spearman which extends the coefficient of determination R^2 over non-linear data, and Kendall which is another test for non-linear data. Using three different measures can give a better overall picture of the data trend analysis, however when using correlation measures there are some important concepts to keep in mind that this section will discuss. In the following, time series data is represented as a variable vector $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ where x_i is the recorded data at time index i, and N is the total number of data points. A common way, tightly connected to correlation, to study two quantitative variables' relationships is linear regression.

2.1. Linear regression

Linear regression tries to capture the relationship between $\mathbf{X}$ and $\mathbf{Y}$ with a straight line $\mathbf{Y} = \mathbf{X}\beta + \epsilon$. The result can e.g. be used to predict a new data point at time i is denoted as $\hat{y}_i$. A common parameter to assess the fit of a linear regression is $R^2 \in [0, 1]$:

$$R^2 = \frac{\text{Regression Sum of Squares}}{\text{Total Sum of Squares}} = \frac{\text{RSS}}{\text{TSS}} \quad (1)$$

where RSS sums the residuals between the mean of **Y** and the regression line and TSS sums the residuals between the data point and the mean of **Y**, over each data point. If the line fits all data perfectly RSS = TSS and $R^2 = 1$, else less. The value of R^2 explains the proportion of the total variation in **Y** accounted for using **X** as the independent variable. Its value does however not solely decide the assessment as a data set with a smaller range can give a lower R^2 than an elongated data set — even though the data dispersion and the linear regression are the same in both cases. A lower R^2 value also does not guarantee a weaker relationship, and vice versa, but is an indication [34]. Pearson correlation, described in 2.3.1, is closely related to R^2 as squaring the Pearson correlation coefficient equals R^2, which can be shown using Eq. (3) and (6).

$$R^2 = \frac{\text{RSS}}{\text{TSS}} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \mu_{\hat{y}})^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_y)^2} = \frac{\text{var}(\hat{\mathbf{Y}})}{\text{var}(\mathbf{Y})} = \left(\frac{\text{cov}(\mathbf{Y},\hat{\mathbf{Y}})}{\sigma_{\mathbf{Y}}\sigma_{\hat{\mathbf{Y}}}}\right)^2 \tag{2}$$

2.2. Variance and covariance

Variance is a measure of the dispersion around the mean of a variable and is denoted by σ^2, where its square root, σ, is the commonly used standard deviation. Variance is calculated as:

$$\text{var}(\mathbf{X}) = \sigma^2 = \frac{\sum_{i=1}^{N}\left(x_i - \mu_{\mathbf{X}}\right)^2}{N} \tag{3}$$

where $\mu_{\mathbf{X}}$ is the mean of **X**:

$$\mu_{\mathbf{X}} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{4}$$

Covariance instead quantifies the average dependency of two variables to change in relation to each other, i.e. their joint variability. If **X** and **Y** are two of equal length, and for time series sampled at the same time and in order, their covariance is:

$$\text{cov}(\mathbf{X},\mathbf{Y}) = \frac{\sum_{i=1}^{N}(x_i - \mu_{\mathbf{X}})(y_i - \mu_{\mathbf{Y}})}{N} \tag{5}$$

If larger values (positive or negative) in **X** are associated with larger values in **Y** with the same sign, then $\text{cov}(\mathbf{X},\mathbf{Y})$ is positive. If they have different signs, the result is negative.

2.3. Correlation measures implemented

Correlation also quantifies the relationship between two variables with a single value, but compared to covariance it is based on two other aspects – *direction* and scale-independent *strength*. While covariance is scale dependent (it differs between e.g. weights depending on if kilograms or pounds are used) correlation is always between [−1,1], making it easier to interpret. In the following subsections, three correlation measures – Pearson, Spearman, and Kendall – are described. Of these, Pearson is the most common and usually the default in software such as MATLAB, Python, and R. The reason for using two additional measures is that they can handle certain non-linear trends, making the analysis more profound. If all measures show a similar correlation the result can be treated with more assurance, otherwise the data should be inspected more in detail.

Note that the expressions in the previous subsection are for when using population data, i.e. using all existing data, and not a sample of it. As the implemented case study contains every hourly value that occurred in 2021, it thus constitutes population data, and the resulting correlations are then denoted with ρ. If a sample would be used instead, N is changed to $N-1$ in Eq. (3) and Eq. (4), and the notation changes where the mean value is denoted with $\bar{x}$, the standard deviation with s_x, and the resulting correlations with r.

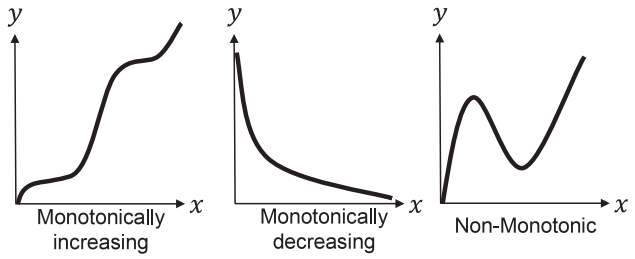


Fig. 3. Monotonic and non-monotonic data.

2.3.1. Pearson correlation

When correlation is mentioned it generally refers to Pearson correlation [35], which can be formulated as:

$$\rho_p = \frac{\sum_{i=1}^{N}(x_i - \mu_{\mathbf{X}})(y_i - \mu_{\mathbf{Y}})}{N \cdot \sqrt{\text{var}(\mathbf{X})}\sqrt{\text{var}(\mathbf{Y})}} = \frac{\text{cov}(\mathbf{X},\mathbf{Y})}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}} \tag{6}$$

Pearson correlation measures a linear trend. However, if the trend deviates too much from being linear a monotonic measure may be a better representation [36]. A monotonic relationship is where the variables continuously increase or decrease, but not necessarily linearly, as shown in Fig. 3. Two correlation measures able to handle a monotonic trend are Spearman and Kendall.

2.3.2. Spearman correlation

Spearman is similar to Pearson but uses the data ranks instead of its actual value [37]. The rank, $R(\mathbf{X})$, is calculated by sorting the data according to its magnitude and assigning the corresponding rank to the original value.

$$\rho_s = \frac{\text{cov}(R(\mathbf{X}), R(\mathbf{Y}))}{\sigma_{R(\mathbf{X})}\sigma_{R(\mathbf{Y})}} \tag{7}$$

A simplified version some programs use gives the same result if there are no tied ranks, which otherwise are discarded:

$$\rho_s = 1 - \frac{6 \cdot \sum_{i=1}^{N} d^2}{N(N^2-1)} \tag{8}$$

where d is the difference between the ranks of x_i and y_i. The interpretation of Spearman resembles Pearson, but with the difference that the variables are re-scaled in a non-linear way using the rank encoding after a regression is fitted to the transformed space. This makes Spearman able to handle monotonic trends, and data that contain outliers.

2.3.3. Kendall correlation

If the variables contain many tied ranks Kendall might be the best choice to reduce the number of tied artifacts. Kendall also ranks data but use it to define concordant and discordant pairs for each data point (x_i, y_i) [38]. This is exemplified in Fig. 4.

There are a total of $\binom{N}{2} = \frac{N(N-1)}{2}$ pairs to calculate, and each pair (x_i, y_i) is compared to all other pairs (x_j, y_j) as:

If $\frac{x_i - x_j}{y_i - y_j} > 0$ pair is concordant

If $\frac{x_i - x_j}{y_i - y_j} < 0$ pair is discordant

If $\frac{x_i - x_j}{y_i - y_j} = 0$ pair is considered a tie

where $i \neq j$. Using N_c for the total number of concordant pairs, and N_d for discordant pairs, the original expression by Maurice Kendall is:

$$\tau_A = \frac{N_c - N_d}{\frac{N(N-1)}{2}} \tag{9}$$

As can be seen, Eq. (9) does not include ties. They can be included in different ways, and one is called the τ_B coefficient which was used in

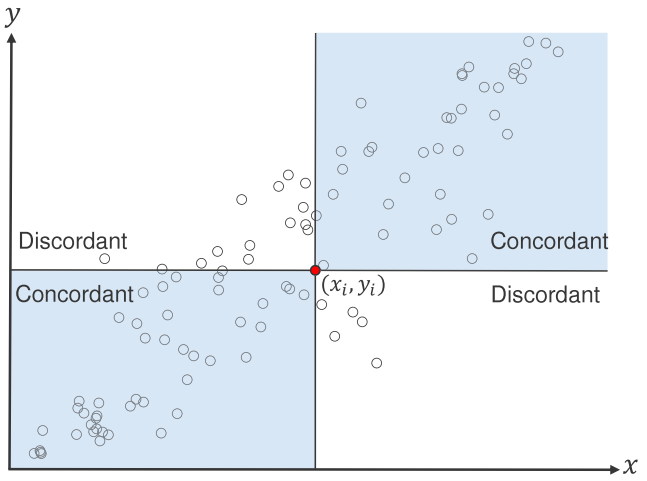


Fig. 4. Concordant and discordant pairs.

Table 1
Correlation interpretation examples from different fields: Psychology, Medicine, Health and Sport, Technology. Here "str" means "strong".

+/-	Psy [40]	Med [41]	H&S [42]	Tech [43]	**This study**
1.0	Perfect	Very str	Perfect	Very str	**Deterministic**
.9–1	Strong	Very str	Distinct	Very str	**Very str**
.8-.9	Strong	Strong	Very high	Very str	**Very str**
.7-.8	Strong	Strong	Very high	Strong	**Strong**
.6-.7	Strong	Moderate	High	Strong	**Strong**
.5-.6	Moderate	Moderate	High	Moderate	**Moderate**
.4-.5	Moderate	Weak	Moderate	Moderate	**Moderate**
.3-.4	Weak	Weak	Moderate	Moderate	**Moderate**
.2-.3	Weak	Weak	Low	Weak	**Weak**
.1-.2	Weak	Weak	Low	Weak	**Weak**
0-.1	Weak	Negligible	Insubstantial	Weak	**Weak**
0	None	Negligible	Trivial	Weak	**None**

this study. It fits best if the underlying scale of both variables is similar and is calculated as:

$$\rho_k = \tau_B = \frac{N_c + N_d}{\sqrt{(N_c + N_d + N_x)(N_c + N_d + N_y)}} \tag{10}$$

where N_x is the number of pairs tied *only* in $\mathbf{X}$ and N_y is the number of tied pairs *only* in $\mathbf{Y}$, which follows from [39]. The interpretation of Kendall is different from both Pearson and Spearman as it represents a probability that the data are in the order that was observed. It can be interpreted as the % of data points that show a positive correlation by using $\tau - \frac{1-\tau}{2}$, e.g. $\tau = 0.5$ gives a positive correlation between 75% of all pairs.

2.4. Correlation strength

Correlation is not an exact science. As an outlook, an interpretation from a few different fields is shown in Table 1.

2.5. Overall correlation considerations

There are different significance tests for correlation, but p-values below 0.05 are typically seen as significant meaning less than a 5% chance that the correlation result is just pure coincidence. Important to notice is that the *p*-value is not directed at the variables' actual relationship, i.e. proving that the result is always true, but at the null hypothesis which states the opposite, i.e. that the result is not true. The correlation is also not directed at the rate of change. This means that a low *p*-value does not guarantee the presence of a large effect, and vice versa. Tiny effects can give rise to p-values < 0.05 if the sample size is large enough the same way that large effects can produce what is generally seen as unacceptable p-values of > 0.05 if the sample size is small [44,45]. A stronger correlation requires fewer data points however to show significance, which is beneficial for this study as stronger correlations, especially positive ones, are more important to point out as these cases include activities that increase their consumption at the same time as the grid load typically has a peak. The nature of this study also makes subsets with potentially fewer data points unavoidable, and significance might be difficult to achieve even though a trend exists between the variables. Common correlation pitfalls that also should be addressed in general are not considering that there might be other variables explaining the correlated variable's relationship (correlation does not imply causation), that another trend than the tested one exists (e.g. testing a linear trend on heavily non-linear data), and that the correlated data do not mix important features that carry explanatory information (Simpson's paradox). In this study the causality is known as the grid load is the aggregation of thousands of different electricity users, the data is plotted and inspected to not include any evident non-linear and non-monotonic trends, and important key features are used to split data into subsets that are analyzed separately. The used data also need to be of "interval" or "ratio" type, described in e.g. [46,47]. A brief summary is shown in Table 2 of the used correlation measures.

Table 2
Summary of the implemented correlation measures.

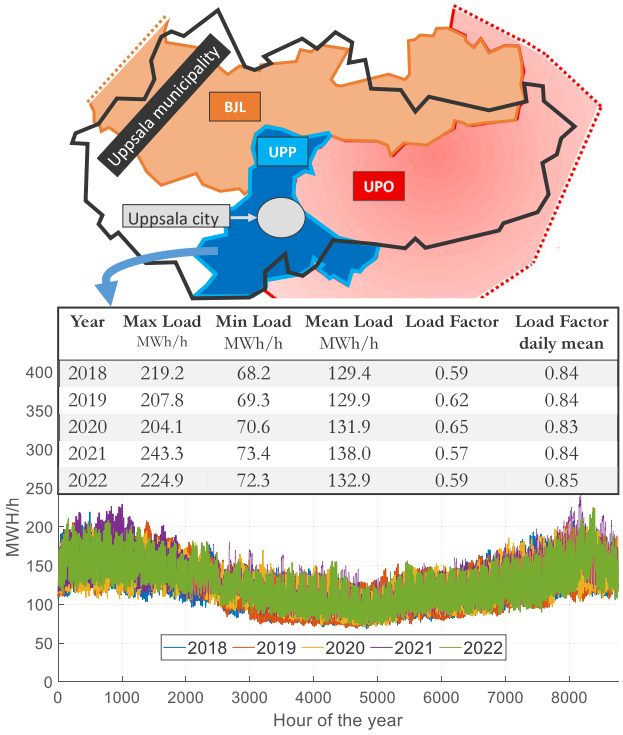
	Advantages	Disadvantages
Correlation measures	Can show relations in data Allow further analysis Quite simple to calculate	Can be misinterpreted Difficult to treat further algebraically
Pearson	Simple Most common, making comparisons easier	Cannot capture non-linear trends Sensitive to outliers
Spearman	More robust than Pearson Can capture some non-linearity	Cannot capture non-monotonic trends
Kendall	More robust than Spearman with smaller sample sizes Visually easier to interpret than the others	Cannot capture non-monotonic trends Requires more calculations

2.6. Re-scaling correlation coefficients

A final component was implemented to emphasize subsets where the user had a larger consumption difference. As the CCs are scaled to [−1,1] it does not provide any information on the magnitude of the consumption difference, i.e. the rate of change. Two users might both have the same positive correlation with the grid load, i.e. both the user and the grid increase their load simultaneously. This fact alone does however not say anything about the magnitude of the users' changes; if one user increases from 10 to 15 kWh/h and the other from 10 to 50 the CCs could be the same. However, if a user has a larger increase this is valuable information as that user will contribute more to the grid peaks. To account for this, a re-scaling factor was implemented using the user's consumption data split into five quantiles. The quote of the fifth and first quantile was used to re-scale the subsets CCs, enough to react to typically larger values but not too sensitive for occasional ones. For example, a vector containing a hundred evenly spaced values from 10 to 15 can be divided into six equally probable partitions with the five quantiles [10.82 11.66 12.50 13.34 14.18] which gives a re-scaling factor of 1.31. A larger spread in data results in a larger re-scaling factor.

3. Data pre-processing and subset creation

The creation of subsets is based on the following steps: (1) **data pre-processing**, aggregate and check time series data; (2) **splitting data into subsets**, accounting for the influence of time and temperature on



Year	Max Load MWh/h	Min Load MWh/h	Mean Load MWh/h	Load Factor	Load Factor daily mean
2018	219.2	68.2	129.4	0.59	0.84
2019	207.8	69.3	129.9	0.62	0.84
2020	204.1	70.6	131.9	0.65	0.83
2021	243.3	73.4	138.0	0.57	0.84
2022	224.9	72.3	132.9	0.59	0.85

Fig. 5. Uppsala Municipality and network areas with grid load data from UPP during 2018–2022. Load Factor is a measure of grid usage efficiency calculated as $\frac{\text{mean load}}{\text{max load}}$.

the electricity consumption; (3) **plot and inspect the data**, to assess trends and potential outliers. But first an introduction to the case study and the used data.

3.1. Studied case - Uppsala activity and grid characteristics

Uppsala municipality has over 14,000 employees and operates a broad portfolio of activities from schools and sports facilities to EV-parking garages [48]. In Fig. 5 UM and the largest "network areas" are shown, with data from UPP. Within each network area, a single company has a monopoly on grid operations. While UPP extends beyond Uppsala, a vast majority of its inhabitants live in Uppsala. Thus, an assumption is that UPP is a fair approximation of Uppsala's grid load. The city has two main transmission substations that provide a subscription limit of 293 MW, and between 2015 to 2019 it was exceeded approximately 2% of the time for a total of 802 h during 112 days; all during the winter [49]. As Fig. 5 shows, the exceeding peaks take place during a shorter time than the hourly data granularity can capture. The data however gives an indication of when the highest load hours occur, and another assumption is that the exceeding peaks coincide with the peaks of the hourly data.

The studied activities were a school, an office, a sports facility, and a parking garage with EV-chargers; all with hourly data from 2021. The school is a vocational high school, and as it focuses on technical education it likely uses more machines and utilities than the average school. The office houses Region Uppsala, a public entity responsible for healthcare and transportation in addition to regional development, and it is around 8000 m^2 with 350–400 office workers. The sports facility is a combined outdoor summer and winter stadium hosting a variety of activities, e.g. football during summer and ice skating during winter. It can artificially create ice which prolongs the winter season, but leads to high electricity consumption when used. The parking garage finished in late 2020 and is equipped with 60 EV-chargers. Its consumption data also includes peripheral utilities (lights, ventilation, and elevators), and an energy management system is under development to optimize the building's load behavior with respect to grid conditions [50].

3.2. Data pre-processing

First, data were checked for missing values and outliers. No unanimous framework exists for handling outliers, but common measures are interquartile ranges or standard deviations [51]. As grid peaks are the main interest of this study, removing outliers using these might remove points of interest and hence they were not used. Detecting outliers was left to the visual inspection of the data trend in Fig. 9. Secondly, Swedish national holidays were removed as these days usually have a non-typical consumption behavior.

3.3. Splitting data into subsets

As described in Section 1.1, the consumption data was split into subsets based on two influential features: time and temperature [21,52]. These splits needed to be reasonable — not too strict, nor too loose.

3.3.1. Subset creation using time

The time split criteria were selected by aggregating the 250 highest load hours per year from 2017–2021 and analyzing when these occurred, details shown in Fig. 6. Weekdays and weekends showed a clear difference and constituted the first split criterion. The second split criterion was based on the two most common daily peak hours, and two hours before and after the identified peaks were chosen to analyze the behavior in relation to it.

3.3.2. Subset creating using temperature

The grid load in UPP had a temperature dependency as shown in Fig. 7 **(a)**, and separating the data between daytime (07:00–19:00) and nighttime (19:00–07:00) narrowed the variance somewhat showing the effect of daily behaviors. To capture overall trends, the third split criterion was based on a seasonal classification inspired by the Swedish Meteorological and Hydrological Institute's (SMHI) that uses the classifications in Table 3. In Fig. 7 **(a)** the trend between the grid load and the temperature could be approximated by two linear trends meeting at approximately 13 °C, one to the left with a negative slope

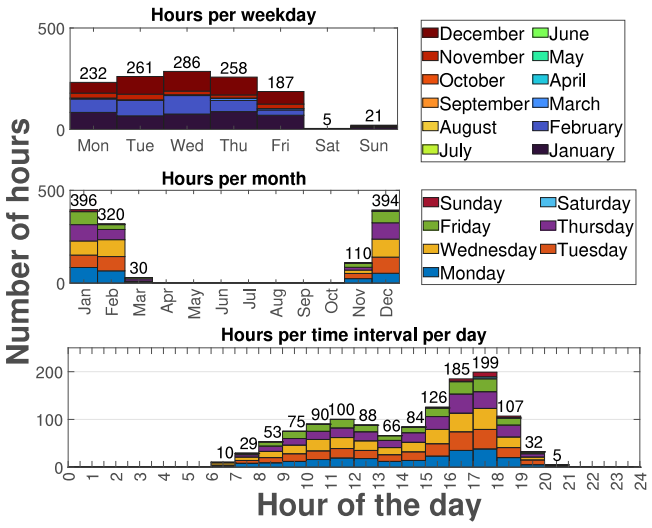


Fig. 6. The occurrence of the 250 highest load hours per year from UPP 2017–2021.

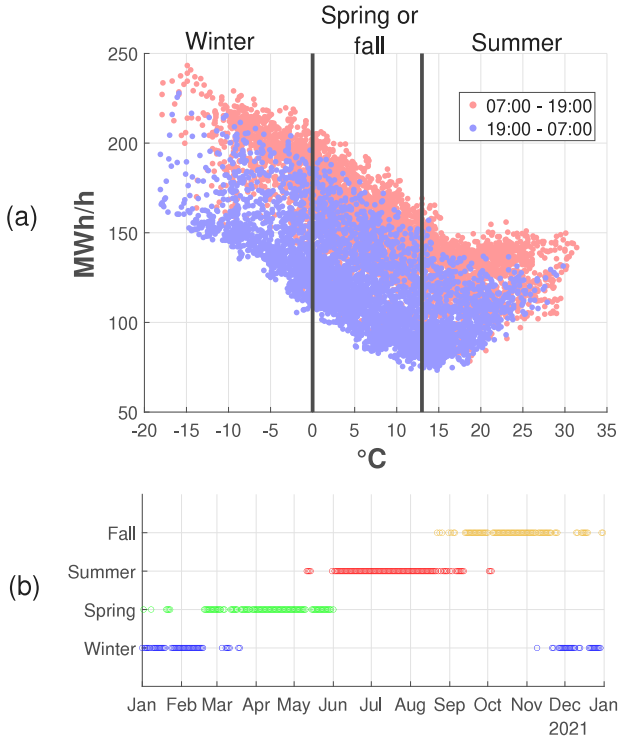


Fig. 7. **(a)** UPP grid load vs. ambient temperature with temperature limit distinctions used for season classification, and **(b)** resulting daily classification used for the subsets.

and one to the right with a positive slope. Thus, 13 °C was used instead of 10 °C as the limit between summer and spring/fall. This separated the temperature dependency into three sections with an overall linear trend. Also, instead of using the arithmetic mean as T_m, a balanced mean temperature was used calculated with Ekholm–Modén's equation:

$$T_m = \frac{aT_{07} + bT_{13} + cT_{19} + dT_{\min} + eT_{\max}}{100} \quad (11)$$

where the numerical indices are temperatures at the corresponding time of the day using wintertime, while $T_{\min}$ and $T_{\max}$ is the minimum and maximum temperature from 19:00 the previous day until 19:00 the current day (+1 h during summertime). The variables a, b, c, d, e are parameters that depend on the month and the longitude of the observation, rounded to the nearest whole meridian selecting the highest if tied. Uppsala's coordinates are 59°51'31.75" N and 17°38'20.00" E, and selecting 18° gave the values in Table 4 [53]. When SMHI uses Eq. (11) seven days in a row with $0 < T_m < 10$°C are required before spring is declared (five for fall); in this study each day was classified independently as the electricity consumption changes quicker with ambient temperature. The resulting seasonal classification by using Eq. (11) is shown in Fig. 7 **(b)**. In Fig. 6 the highest load hours were most frequent in January, February, November, and December, and most of these align quite well with the winter classification in Fig. 7 **(b)**.

Table 3
SMHI seasonal classifications where T_m = daily mean temperature.

Season	Date	Temperature
Winter	Dec-Feb	$T_m < 0$ °C
Spring	Mar-May	$0 < T_m < 10$°C[a]
Summer	Jun-Aug	$T_m > 10$ °C
Fall	Sep-Nov	$0 < T_m < 10$ °C[a]

[a] Spring if taking place between winter and summer, otherwise fall.

Table 4
Ekholm–Modén model coefficients.

	a	b	c	d	e
Jan	33	15	32	10	10
Feb	31	18	31	10	10
Mar	31	21	28	10	10
Apr	23	18	30	10	19
May	22	20	23	10	25
Jun	21	19	24	10	26
Jul	19	18	26	10	27
Aug	18	22	23	10	27
Sep	25	23	24	10	18
Okt	29	19	32	10	10
Nov	30	16	34	10	10
Dec	34	15	31	10	10

Table 5
Used subsets splitting criteria.

Feature	Split condition	Resulting sets
Time	Type of day	Weekday (Mon–Fri) Weekend (Sat–Sun)
Time	In proximity to grid peaks	11–12 ± 2 h 17–19 ± 2 h
Temperature	Ambient temperature	Winter (<0 °C) Spring (0–13 °C[a]) Summer (>13 °C) Fall (0–13 °C[a])

[a] Spring if taking place between winter and summer, otherwise fall.

3.3.3. Resulting subsets

The criteria used to create subsets are summarized in Table 5. Using all three gave a total number of $2 \cdot 2 \cdot 4 = 16$ subsets to capture all combinations.

3.4. Plot subsets and inspect overall data trend

It is vital to graph data to reveal any evident trends that might impact the correlation measures. The data set called *Anscombe's quartet* shows why inspecting the data is essential as it shows how data sets with nearly identical descriptive statistics can have very different appearances when graphed [54]. As the correlation measures test linear and monotonic relationships, data that would strongly exhibit other trends would give an inadequate interpretation of the resulting CCs. To illustrate, Pearson correlation on different data trends is shown in Fig. 8; in **(a)** on data with an increasing linear trend, and in **(b)** on data

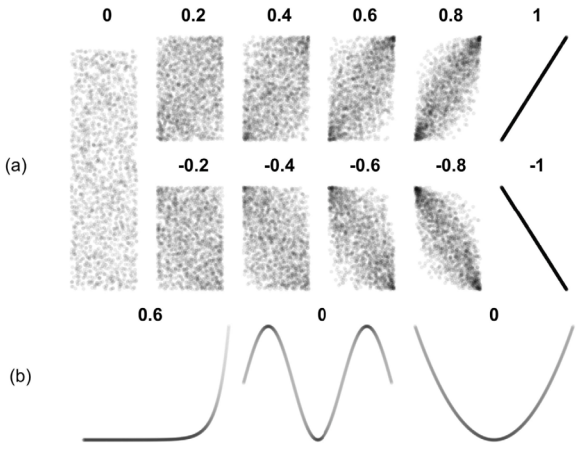


Fig. 8. Pearson correlation on different data trends, linear in (a) and non-linear in (b).

with clearly non-linear trends where the first trend is monotonic, which Spearman and Kendall can handle, while the others are non-monotonic and where all three measures would give unrepresentative results.

A visual inspection to identify any non-linear and non-monotonic trends was done in Fig. 9, of which none could be found; the trends instead resembled patterns in Fig. 8 **(a)**. The patterns in Fig. 8 **(b)**, and especially the last two, would represent more complex (and probably rare) electricity consumption patterns. Finally, an overview of the entire presented framework is summarized in Fig. 10.

4. Framework correlation analysis results

The result of the framework in Section 3 is shown in Fig. 11 using the sports facility as an example. First, correlating all data is shown in Fig. 11 **(a)** and using the 500 highest load hours in Fig. 11 **(b)**, showing two distinctively different trends just with this simple split. Can these results be used to conclude if the sports facility, in general, contributed to grid peak loads? Overall the sports facility seemed to consume less electricity at high grid load hours, but it was not clear if this was a consistent behavior during a specific time interval or not. Splitting the data using the type of day is shown in Fig. 11 **(c)** showing that the grid's load was lower on weekends, but the CCs were quite similar to Fig. 11 **(a)**. The next step, shown in Fig. 11 **(d)**, split the data using the vicinity to the two typical daily grid peaks. A slight difference emerged between the subsets, but overall the CCs were still similar. The data points not used for correlation were omitted to improve clarity. The data were then split a final time using the season classification shown in Fig. 11 **(e)**, and here a substantial difference appeared between the subsets where e.g. the trends between weekdays during the winter and the spring were clearly different — which was not visible in the previous split. The CCs, including re-scaled CCs, are shown for fall weekdays as an example of how the activity's consumption difference is emphasized, showing that there was a larger difference between typical values.

The resulting CCs for all activities are summarized in Tables 6–8. The CCs in Table 6, the re-scaled CCs in Table 7, and with the re-scaled CCs during winter weekdays where most of the highest grid load hours had occurred in Table 8. The CCs are colored between −1 (green) to +1 (red), while the re-scaled CCs are colored based on the lowest value (green) to the highest value (red). A bold font means that the correlation's p-value was above 0.05.

5. Discussion

The discussion starts with some overall aspects, and end with the specific case study and its local context. A p-value larger than 0.05, i.e. statistical insignificance, occurred in a few subsets but typically only for weaker CCs closer to 0 and never for the stronger CCs. Smaller CCs are however, apart from being more difficult to prove, less interesting as it is the users with larger positive correlations that tend to increase their consumption at the same time as the grid load increases, thus contributing more to the grid load peaks. Looking at the CCs in Table 6, they differed over the year in all activities. This is not at all reflected by the first row that correlated all data simultaneously. Take the office as an example, during weekdays it mostly showed a weak to moderate correlation except for the summer which was close to strong according to Table 1. The summer is however less interesting as the highest grid peak loads only have occurred during the winter months, as shown in Fig. 6. Looking at the winter months, and especially the weekdays, the school, the office, and the EV-parking garage all had a weak to moderate positive correlation. The sports facility instead had a larger and negative correlation in the vicinity of the 09–14 peak, which is favorable as the activity then typically reduced its consumption during these hours.

The average magnitude difference between the CCs in the same subsets was 0.12, with the smallest being 0.01 and the largest 0.20, showing that it can be meaningful to use more than one type of correlation and compare their results to decide if the data should be scrutinized in more detail. For instance, the sports facility had a larger difference during fall weekdays. Looking at this subset in Fig. 9 or

Table 6
CCs for all activities in all subsets.

USER -->			School			Office			Sport Facility			Parking Garage		
MEAN kWh/h -->			73.51			23.28			378.72			23.02		
CC -->			ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K
ALL DATA			0.37	0.36	0.25	0.55	0.56	0.41	0.40	0.62	0.44	0.20	0.24	0.16
9-14	Wi	Wd	0.24	0.23	0.18	0.22	0.16	0.10	-0.51	-0.51	-0.35	**0.07**	**0.09**	**0.06**
15-20			**0.12**	0.23	0.16	0.32	0.31	0.21	-0.18	-0.12	**-0.07**	0.21	0.24	0.17
9-14		We	0.33	0.35	0.23	-0.26	-0.25	-0.19	-0.25	-0.22	**-0.14**	**0.20**	**0.16**	**0.11**
15-20			0.39	0.40	0.28	-0.25	-0.25	-0.20	**-0.14**	**-0.01**	**0.00**	**0.08**	**0.08**	**0.05**
9-14	Sp	Wd	-0.22	**-0.22**	**-0.14**	0.39	0.42	0.30	0.64	0.51	0.35	**0.04**	**0.04**	**0.03**
15-20			**-0.18**	**-0.09**	**-0.06**	0.15	0.15	0.10	0.73	0.67	0.47	0.15	0.18	0.12
9-14		We	0.05	0.01	0.00	0.29	0.30	0.22	0.77	0.75	0.54	**0.13**	0.26	0.18
15-20			0.36	0.29	0.23	0.19	0.23	0.17	0.59	0.53	0.33	0.17	0.24	0.17
9-14	Su	Wd	0.46	0.58	0.42	0.72	0.69	0.52	0.26	0.45	0.31	0.10	0.14	0.10
15-20			0.37	0.50	0.36	0.68	0.70	0.52	0.19	0.33	0.23	**-0.01**	**0.07**	**0.05**
9-14		We	0.39	0.36	0.28	0.30	0.28	0.22	0.29	0.26	0.19	0.23	0.30	0.22
15-20			0.51	0.36	0.29	0.33	0.36	0.27	**-0.04**	**-0.05**	**-0.04**	**0.14**	0.21	0.16
9-14	Fa	Wd	**-0.05**	**-0.02**	**-0.02**	0.27	0.43	0.32	0.50	0.46	0.28	0.42	0.39	0.25
15-20			**0.07**	**0.01**	**0.01**	0.23	0.22	0.15	0.57	0.58	0.39	0.37	0.37	0.24
9-14		We	**-0.09**	**-0.08**	**-0.07**	0.23	0.20	0.16	0.68	0.59	0.39	0.35	0.34	0.23
15-20			0.25	0.21	0.16	**-0.19**	**-0.18**	**-0.13**	0.58	0.55	0.35	0.36	0.39	0.25

Table 7
CCs re-scaled, each activity compared separately.

USER -->			School			Office			Sport Facility			Parking Garage		
MEAN kWh/h -->			73.51			23.28			378.72			23.02		
CC -->			ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K
9-14	Wi	Wd	0.57	0.55	0.42	0.29	0.20	0.14	-1.32	-1.32	-0.90	**0.15**	**0.17**	**0.12**
15-20			**0.20**	0.39	0.27	0.63	0.60	0.42	-0.33	-0.22	**-0.13**	0.42	0.48	0.33
9-14		We	0.39	0.41	0.27	-0.30	-0.29	-0.23	-0.63	-0.56	**-0.36**	**0.42**	**0.33**	**0.24**
15-20			0.49	0.50	0.36	-0.29	-0.29	-0.23	**-0.44**	**-0.02**	**-0.01**	**0.13**	**0.13**	**0.09**
9-14	Sp	Wd	-0.39	-0.39	-0.26	0.44	0.47	0.34	2.54	2.03	1.38	**0.06**	**0.06**	**0.05**
15-20			-0.34	**-0.18**	**-0.12**	0.27	0.27	0.18	3.26	2.99	2.09	0.35	0.42	0.28
9-14		We	**0.06**	**0.01**	**0.00**	0.33	0.34	0.25	3.72	3.61	2.62	**0.17**	0.35	0.25
15-20			0.39	0.32	0.25	0.21	0.27	0.20	2.97	2.68	1.68	0.35	0.49	0.34
9-14	Su	Wd	0.79	0.99	0.72	0.91	0.87	0.66	0.33	0.58	0.40	0.16	0.22	0.15
15-20			0.74	1.00	0.73	1.22	1.26	0.93	0.26	0.44	0.31	**-0.03**	**0.17**	**0.11**
9-14		We	0.42	0.39	0.31	0.34	0.33	0.25	0.43	0.40	0.29	0.41	0.53	0.38
15-20			0.56	0.39	0.31	0.39	0.42	0.32	**-0.14**	**-0.15**	**-0.12**	**0.32**	0.49	0.36
9-14	Fa	Wd	**-0.08**	**-0.03**	**-0.03**	0.33	0.52	0.39	2.63	2.44	1.49	1.05	0.99	0.64
15-20			**0.14**	**0.02**	**0.01**	0.47	0.45	0.31	2.80	2.84	1.91	1.29	1.31	0.85
9-14		We	**-0.11**	**-0.09**	**-0.09**	0.27	0.23	0.19	4.11	3.60	2.38	1.50	1.49	1.00
15-20			0.27	0.23	0.17	**-0.22**	**-0.21**	**-0.15**	2.94	2.75	1.74	1.09	1.18	0.75

Table 8
CCs re-scaled, all activities compared jointly during winter weekdays.

USER -->			School			Office			Sport Facility			Parking Garage		
MEAN kWh/h -->			73.51			23.28			378.72			23.02		
CC -->			ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K	ρ_P	ρ_S	ρ_K
9-14	Wi	Wd	0.57	0.55	0.42	0.29	0.20	0.14	-1.32	-1.32	-0.90	**0.15**	**0.17**	**0.12**
15-20			**0.20**	0.39	0.27	0.63	0.60	0.42	-0.33	-0.22	**-0.13**	0.42	0.48	0.33

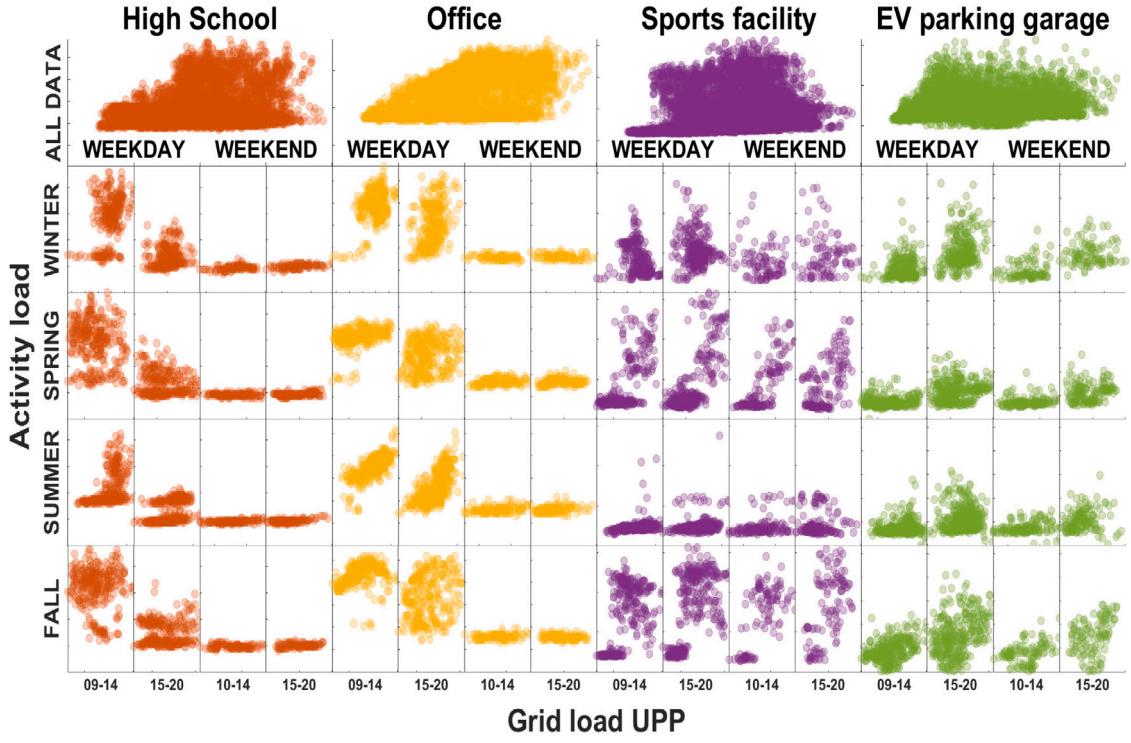


Fig. 9. Scatter plot of every subset with user consumption against grid load, showing no evident outliers or non-linear trends.

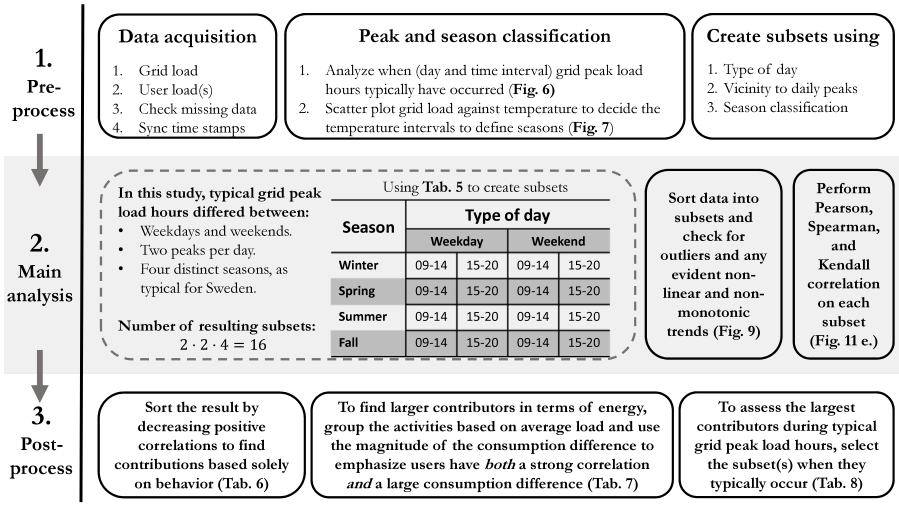


Fig. 10. Overview of the framework described in Section 3.

Fig. 11 **(e)**, there was a concentration of lower activity consumption values together with spread out significantly larger ones. The difference in CCs can be attributed to how the three measures handle such situations, and this is a delicate part because depending on how the data is split different results may occur. For instance, if the lower values were filtered away, the CCs would be different. These data points are however a part of how the sports facility behaves in the vicinity of high grid load hours. The motivation behind the creation of the subsets aims to capture well-distinguished and re-occurring instances of time during the year to assess a user's behavior typically associated with it.

The effect of re-scaling the CCs in Table 6 is shown by Table 7. As can be seen, some CCs increased by several hundred percent emphasizing a larger consumption difference in the associated subset. The CCs cannot point out what causes the difference, but the re-scaling can indicate where it exists. If it would be possible to change it – e.g. by shifting the consumption in time or using an energy storage – the users' grid load contribution would reduce and the associated CCs decrease.

Finally, the result could be further discussed using expert knowledge to decide how, if, and which users should be investigated in terms of possible changes to reduce their grid peak load contributions. This is not something the CCs can answer, they can only provide input on where time and resources strategically could be directed given there is a possibility for change. It could also be useful to take the activities' average consumption into account, shown under the activity type in Tables 6–8. If UM would analyze a larger number of their activities, the average consumption can be used to sort users into groups; and sort the users within each group based on their CCs values to find the largest contributors at each level. Of the studied activities, the sports

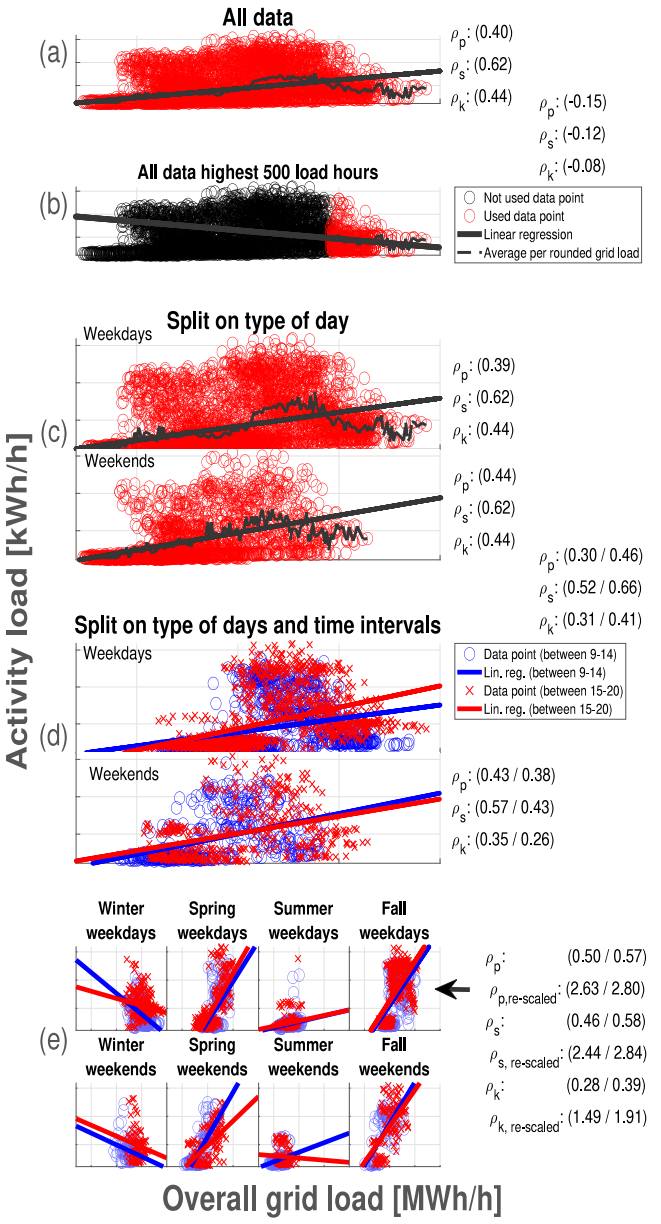


Fig. 11. Correlation on increasing split levels of the sports facility data. The limits of the axes are the same in all figures.

facility had by far the largest load, and also the largest CCs in several subsets which should be analyzed further. The number of users within the type of activity could also be an important factor; if there would be e.g. 50 offices but only two parking garages and all having similar positive CCs, it might be better to focus on the offices for an effective energy transition process as it likely is simpler to implement changes and spread information within the same type of activity. However, although fewer, the parking garages could have a behavior that is easier to change due to EV-charging typically having a high consumption for a shorter time. If the EV-charging could be shifted, without affecting the car users notably, it would be a low-hanging fruit to implement. The presented framework in this paper is a usable tool in these kinds of continued analyses and decision-making processes.

6. Conclusion

This paper presents a framework to assess if a user's electricity consumption contributes to the grid load peaks. The general idea is to correlate the time series data from one, or several, users with the grid load in subsets based on time and ambient temperature to analyze potentially different trends separately. Three correlation measures is used to complete each other and catch different types of trends. If both the user's consumption and the grid's load frequently increased at the same time the correlation would be positive, and the user seen as contributing to peaks and the risk of grid congestion; if the user's consumption instead decreased while the grid load increased, the correlation would be negative and the user could rather be seen as supporting the grid during peak hours. If there were no clear pattern the correlation would be closer to zero, and the user would not be contributing either way. This framework can be a useful and valuable tool for a larger organization such as a public entity or a grid company that handles many electricity-consuming types of activities and users to identify which contributes most to grid peak loads. Increased knowledge in this topic can be used as input to decision-making about smart grid interventions such as energy storage, local electricity production, and potential flexibility options, or if grid extension seems to be the only solution. Extending the grid is however costly in both capital and time, thus alternatives should be investigated prior.

A case study is done to investigate users from four types of activities: a school, an office, a sports facility, and a parking garage with EV chargers. First, as a base scenario, all activities' data are correlated with the grid load without splitting data into subsets where the result, the first row in Table 6, shows that all activities had a positive correlation with the grid. Sorting the data into subsets instead shows differences that were not visible before, and that there even were times of the year with moderate negative correlations, showing the importance of the data split process to avoid unrepresentative conclusions.

The framework this paper presents is not exclusive to analyzing the relationship between electricity consumption and the grid load, it can also be used on other types of variables changing over time. A couple of examples are how a user's consumption correlates with the production of different types of renewable electricity, giving an indication of which source to prefer. It could also be used in urban planning to geographically co-locate activities that typically are negatively correlated, increasing the likelihood of a more even grid load. This would lead to a more effective energy transition process without procuring and investing in more appliances or services than needed.

CRediT authorship contribution statement

Carl Flygare: Conceptualization, Methodology, Formal analysis, Writing. **Alexander Wallberg:** Investigation. **Erik Jonasson:** Methodology. **Valeria Castellucci:** Supervision. **Rafael Waters:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Thanks to Uppsala Municipality for supporting this Ph.D. project, and to Marcus Nystrand and Maria Thomtén at Uppsala Municipality for discussions. Also thanks to providers of data within Uppsala Municipality (Skolfastigheter, UPAB, and Arenor och Fastigheter), and Region Uppsala.

References

[1] International Energy Agency. Contribution of renewables to energy security. Tech. rep., 2007.
[2] International Energy Agency. World energy outlook. Tech. rep., 2022.
[3] Coles D, Wray B, Stevens R, Crawford S, Pennock S, Miles J. Impacts of tidal stream power on energy system security: An Isle of Wight case study. Appl Energy 2023;334. http://dx.doi.org/10.1016/j.apenergy.2023.120686.
[4] Fant C, Boehlert B, Strzepek K, Larsen P, White A, Gulati S, et al. Climate change impacts and costs to U.S. Electricity transmission and distribution infrastructure. Energy 2020-03;195:116899. http://dx.doi.org/10.1016/j.energy.2020.116899.
[5] Hoang AT, Nižetić S, Olcer AI, Ong HC, Chen WH, Chong CT, et al. Impacts of COVID-19 pandemic on the global energy system and the shift progress to renewable energy: Opportunities, challenges, and policy implications. Energy Policy 2021;154. http://dx.doi.org/10.1016/j.enpol.2021.112322.
[6] International Energy Agency. A 10-point plan to reduce the European union' s reliance on Russian natural gas. In: IEA publications, no. March. 2022, p. 1–12, URL https://www.iea.org/reports/a-10-point-plan-to-reduce-the-european-unions-reliance-on-russian-natural-gas.
[7] Energimyndigheten. Scenarier över sveriges energisystem 2020. Tech. rep., 2021, URL www.energimyndigheten.se.
[8] Svenskt Näringsliv. Kraftsamling elförsörjning: Scenarioanalys 2050. Tech. rep., 2022.
[9] Energiföretagen. Efterfrågan på fossilfri el: Analys av högnivåscenario. Tech. rep., 2021.
[10] IVA. Om vätgas och dess roll i elsystemet - syntesrapport från IVAs projekt vätgasens roll i ett fossilfritt samhälle. Tech. rep., 2022.
[11] NEPP. Insikter och vägval i energiomställninge - slutrapport från NEPP:s andra etapp. Tech. rep., 2020.
[12] Svenska Kraftnät. System development plan 2022–2031: Towards a doubled use of electricity. Tech. rep., 2022.
[13] Zhong J, Bollen M, Rönnberg S. Towards a 100% renewable energy electricity generation system in Sweden. Renew Energy 2021;171:812–24. http://dx.doi.org/10.1016/j.renene.2021.02.153.
[14] Bjärstig T, Mancheva I, Zachrisson A, Neumann W, Svensson J. Is large-scale wind power a problem, solution, or victim? A frame analysis of the debate in Swedish media. Energy Res Soc Sci 2022;83. http://dx.doi.org/10.1016/j.erss.2021.102337.
[15] Öhrlund I. Demand side response: Exploring how and why users respond to signals aimed at incentivizing a shift of electricity use in time. [PhD dissertation], Uppsala University; 2020.
[16] Fung W, Lam K, Hung W, Pang S, Lee Y. Impact of urban temperature on energy consumption of Hong Kong. Energy 2006;31:2623–37. http://dx.doi.org/10.1016/j.energy.2005.12.009.
[17] Kim HT, Jin YG, Yoon YT. An economic analysis of load leveling with battery energy storage systems (BESS) in an electricity market environment: The Korean case. Energies 2019;12. http://dx.doi.org/10.3390/en12091608.
[18] Jesper M, Pag F, Vajen K, Jordan U. Heat load profiles in industry and the tertiary sector: Correlation with electricity consumption and ex post modeling. Sustainability (Switzerland) 2022;14. http://dx.doi.org/10.3390/su14074033.
[19] Bell WP, Wild P, Foster J, Hewson M. Wind speed and electricity demand correlation analysis in the Australian national electricity market: Determining wind turbine generators' ability to meet electricity demand without energy storage. Econ Anal Policy 2015;48:182–91. http://dx.doi.org/10.1016/j.eap.2015.11.009.
[20] Richardson DB, Harvey LD. Strategies for correlating solar PV array production with electricity demand. Renew Energy 2015;76:432–40. http://dx.doi.org/10.1016/j.renene.2014.11.053.
[21] Humeau S, Wijaya TK, Vasirani M, Aberer K. Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households. In: 2013 sustainable internet and ICT for sustainability. IEEE; 2013, p. 1–6. http://dx.doi.org/10.1109/SustainIT.2013.6685208.
[22] SWECO. Elnätsutmaningen. Tech. rep., 2019.
[23] Energimyndigheten. Nuläget på elmarknaden: Oktober 2022. Tech. rep., 2022.
[24] Olauson J, Bergkvist M. Correlation between wind power generation in the European countries. Energy 2016;114:663–70. http://dx.doi.org/10.1016/j.energy.2016.08.036.
[25] Olausson J, Ayob MN, Bergkvist M, Carpman N, Castellucci V, Goude A, et al. Net load variability in nordic countries with a highly or fully renewable power system. Nat Energy 2016 1:12 2016;1:1–8. http://dx.doi.org/10.1038/nenergy.2016.175.
[26] Lindberg O, Lingfors D, Arnqvist J. Analyzing the mechanisms behind temporal correlation between power sources using frequency separated time scales: A Swedish case study on PV and wind. Energy 2022-11;259:124817. http://dx.doi.org/10.1016/j.energy.2022.124817.
[27] Milanović JV. Probabilistic stability analysis: The way forward for stability analysis of sustainable power systems. Phil Trans R Soc A 2017;375. http://dx.doi.org/10.1098/rsta.2016.0296.
[28] Lee J, Cho Y. National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? Energy 2022-01;239:122366. http://dx.doi.org/10.1016/j.energy.2021.122366.
[29] Alipour P, Mukherjee S, Nateghi R. Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region. Energy 2019-10;185:1143–53. http://dx.doi.org/10.1016/j.energy.2019.07.074.
[30] Zeng P, Jin M. Peak load forecasting based on multi-source data and day-to-day topological network. IET Gener Transm Distrib 2018-03;12(6):1374–81. http://dx.doi.org/10.1049/iet-gtd.2017.0201.
[31] Chen T, Lehr JM, Lavrova O, Martínez-Ramón M. Distribution feeder-level day-ahead peak load forecasting methods and comparative study. IET Gener Transm Distrib 2018-07;12(13):3270–8. http://dx.doi.org/10.1049/iet-gtd.2017.1745.
[32] Takido K, Valeriano OCS, Ryo M, Tanume K, Ushio T, Kubota T. Spatiotemporal evaluation of the gauge-adjusted global satellite mapping of precipitation at the Basin scale. J Meteorol Soc Jpn II 2016;94(2):185–95. http://dx.doi.org/10.2151/jmsj.2016-010.
[33] Zhao J, Chen C, Huang H, Xiang C. Unifying Uber and taxi data via deep models for taxi passenger demand prediction. Pers Ubiquitous Comput 2020;1–13. http://dx.doi.org/10.1007/s00779-020-01426-y.
[34] Achen CH. Measuring representation: Perils of the correlation coefficient. Source: Am J Political Sci 1977;21:805–15.
[35] Freedman D, Pisani R, Purves R. Statistics: Fourth international student edition. Emersion: Emergent village resources for communities of faith series, W.W. Norton & Company; 2007, URL https://books.google.se/books?id=vWeQDAAAQBAJ.
[36] Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. Amer Statist 1988;42:59–66.
[37] Spearman rank correlation coefficient. In: The concise encyclopedia of statistics. Springer New York; 2008, p. 502–5. http://dx.doi.org/10.1007/978-0-387-32833-1_379.
[38] Kendall rank correlation coefficient. In: The concise encyclopedia of statistics. Springer New York; 2008, p. 278–81. http://dx.doi.org/10.1007/978-0-387-32833-1_211.
[39] Agresti A. Analysis of ordinal categorical data. Wiley series in probability and statistics, second ed. Hoboken, N.J: Wiley; 2010.
[40] Dancey CP, Reidy J. Statistics without maths for psychology. Harlow, England: Pearson Education; 2014.
[41] Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg 2018;126:1763–8. http://dx.doi.org/10.1213/ANE.0000000000002864.
[42] Hopkins WG. A new view of statistics. 1997.
[43] Cantão MP, Bessa MR, Bettega R, Detzel DH, Lima JM. Evaluation of hydro-wind complementarity in the Brazilian territory by means of correlation maps. Renew Energy 2017;101:1215–25. http://dx.doi.org/10.1016/j.renene.2016.10.012.
[44] Nahm FS. What the P values really tell us. Korean J Pain 2017;30:241–2. http://dx.doi.org/10.3344/kjp.2017.30.4.241.
[45] de Mariscal EG, Guerrero V, Sneider A, Jayatilaka H, Phillip JM, Wirtz D, et al. Use of the p-values as a size-dependent function to address practical differences when analyzing large datasets. Sci Rep 2021;11:20942. http://dx.doi.org/10.1038/s41598-021-00199-5.
[46] Tan P-N, Steinbach M, Karpatne A, Kumar V. Introduction to data mining. Pearson Education Limited; 2019.
[47] Khamis H. Measures of association: How to choose? J Diagn Med Sonogr 2008;24:155–62. http://dx.doi.org/10.1177/8756479308317006.
[48] Uppsala kommun. Våra verksamheter. 2022, URL https://vardochomsorg.uppsala.se/karriar/verksamheter. [Last accessed 7 March 2023].
[49] Etherden N, Ruwaida Y, Johansson S. D4.5 - Report on lessons learned, bug fixes and adjustments in products and routines within the Swedish demo. Tech. rep., CoordiNet; 2020.
[50] Wallberg A, Flygare C, Waters R, Castellucci V. Peak shaving for electric vehicle charging infrastructure& - A case study in a parking garage in uppsala, Sweden. World Electr Veh J 2022;13:152. http://dx.doi.org/10.3390/WEVJ13080152.
[51] Cousineau D, Chartier S. Outliers detection and treatment: a review. Int J Psychol Res 2010;3(1):58–67. http://dx.doi.org/10.21500/20112084.844.
[52] Meng Q, Mourshed M. Degree-day based non-domestic building energy analytics and modelling should use building and type specific base temperatures. Energy Build 2017;155:260–8. http://dx.doi.org/10.1016/j.enbuild.2017.09.034.
[53] Koefficienterna i ekholm-modéns formel. 2022, URL https://www.smhi.se/kunskapsbanken/meteorologi/koefficienterna-i-ekholm-modens-formel-1.18371. [Last accessed 30 October 2022].
[54] Anscombe F. Graphs in statistical analysis. Amer Statist 1973;27(1):17–21. http://dx.doi.org/10.1080/00031305.1973.10478966.