Impact of Weather Factors on the Electricity Consumption of Estonia

Bachelor's Thesis

Student: Marti Kütt
Student code: 222805 YAFB
Supervisor: Jaan Kalda, Department of Cybernetics, Tenured Full Professor
Study programme: Applied Physics

Tallinn 2025

Ilmastikufaktorite mõju Eesti elektritarbimisele

Bakalaureusetöö

Üliõpilane: Marti Kütt
Üliõpilaskood: 222805 YAFB
Juhendaja: Jaan Kalda, Küberneetika Instituut, täisprofessor tenuuris
Õppekava: Rakendusfüüsika

Tallinn 2025

Author’s declaration

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Marti Kütt

The thesis meets the requirements of a bachelor’s thesis.
Supervisor: Jaan Kalda

Table of contents

List of Figures

List of abbreviations and terms

CDS	Climate Data Store

Introduction

Understanding the complex relationship between weather conditions and electricity consumption patterns is essential for effective energy planning and grid management in Estonia's variable climate. This thesis provides a novel county-level assessment that reveals critical regional differences in how weather influences energy demand across Estonia.

Estonia presents a particularly interesting case study due to its distinct seasonal variations, coastal-inland climate differences, and ongoing energy transition. As the country works toward synchronization with Continental Europe's power grid and increases its renewable energy capacity, quantifying specific weather impacts becomes increasingly valuable for both short-term operations and long-term planning.

Additionally, this thesis addresses a key methodological limitation identified in [1], where the analysis was constrained by reliance on single weather station data and solar radiation estimates derived from solar altitude angles and cloud cover parameters. To overcome this spatial and measurement limitation, the present study employs ERA5-Land reanalysis data, which provides comprehensive meteorological coverage across all Estonian counties. This approach is expected to reduce measurement noise and enable more robust quantification of weather-electricity consumption relationships at the county level.

The primary objective of the present work is to develop multiple linear regression models that precisely quantify weather effects on electricity consumption patterns across Estonian counties for both business and residential consumers. This approach allows for isolation of specific weather variables' contributions to consumption variance, revealing regional energy profiles and efficiency patterns.

1. Literature review

This chapter examines methodologies for analyzing weather impacts on electricity consumption, with supporting insights drawn from the field of load forecasting. Section 1.1 reviews forecasting approaches across different time horizons, which helps identify which modeling techniques are most relevant for weather impact analysis. Section 1.2 builds on this foundation by detailing the variables affecting electricity demand. Section 1.3 then explores specific methodologies for quantifying weather impacts on consumption. Sections 1.4 and 1.5 provide the theoretical foundation for multiple linear regression, including the mathematical framework and coefficient of determination, which form the basis for this study's methodology.

1.1 Load forecasting

Load forecasting or the prediction of future electricity consumption quickly becomes a challenging task due to the convoluted and stochastic nature of the data. Forecasting types can roughly be separated by the time horizon of the forecasts and the type of model or approach used for prediction. [2]

Short term forecasting usually spans from minutes to a few days, and has a key role in grid reliability and dispatch analysis. Medium term forecasting ranges from weeks to months, being important in adequacy assessment and evaluating the financial attributes of the energy system. Long term forecasting expands the timescale to years and is needed for capturing long term trends and production planning. [2]

The methodological approaches vary considerably across these time horizons. Short-term forecasting commonly employs similar-day approaches, regression methods, time series analysis (like Auto-Regressive Moving Average and Auto-Regressive Integrated Moving Average models), neural networks, fuzzy logic, and expert systems with each offering different capabilities for capturing immediate patterns. Medium and long-term forecasting typically relies on end-use approaches that analyze specific consumer usage patterns and appliance adoption, or econometric approaches that combine economic theory with statistical techniques to model relationships between consumption and influencing factors. [3]

Simpler statistical methods like multiple linear regression are often preferable to complex machine learning or hybrid models, especially in utility forecasting. While all forecasts face inherent uncertainty, simpler approaches provide greater defensibility through interpretability, traceability, and reproducibility. Though complex models might marginally improve accuracy in some cases, this benefit is frequently outweighed by the "black box" nature that makes them difficult to explain and defend. When transparency is paramount, the clarity of simpler statistical methods offers significant practical advantages. [3]

1.2 Variables influencing electricity consumption

European Network of Transmission System Operators for Electricity points out in their demand forecasting methodology temperature, irradiance and wind speed as some of the key factors used for load forecasting [4]. In addition to this it is noted that holidays or special days are treated separately. The approach is further verified in [2], where it is argued that "consumer's electricity demand is significantly influenced by weather conditions, with heating and cooling needs causing peak consumption during extreme temperatures". Other key factors outlined are wind chill index, measuring cold stress in winter, and temperature humidity index, quantifying heat discomfort in summer, as weather factors and weekdays as temporal factors.

Understanding weather-related influences requires examining the underlying physical mechanisms that drive building energy consumption. Building walls act as thermal buffers due to their mass and thermal properties, creating thermal inertia that delays the indoor response to outdoor temperature changes. Thermal inertia is a complex phenomenon that significantly affects electricity demand patterns. While constructing methods with high thermal mass - often using large quantities of concrete or bricks - might be beneficial for reducing overheating risks, thermal mass isn't always beneficial with regard to comfort and energy consumption. An exponential relationship between the quantity of thermal capacity and internal diurnal temperature fluctuations implies that adding additional thermal mass to a heavyweight building has a negligible effect. Furthermore, buildings with large amounts of thermal mass need more time to reach cooling or heating set point temperatures in case of intermittent building use, potentially causing thermal discomfort for occupants and resulting in increased energy consumption. [5]

Temperature effects are further complicated by humidity considerations. Using wet-bulb temperature, which represents the lowest temperature achievable through evaporative cooling when air is adiabatically saturated with water vapor, we can directly quantify the presence of water vapor in the air. This parameter is crucial for determining cooling potential during warm periods. Solar radiation adds another dimension by heating building surfaces directly, converting electromagnetic energy to sensible heat and creating surface temperatures that can exceed ambient air temperature by 20-40℃. [6]

These physical phenomena directly translate to energy demands, as buildings require energy to maintain thermal comfort by counteracting heat losses in winter and heat gains in summer. Heating needs arise from conduction through walls, convection at surfaces, radiation to cold surfaces, and ventilation heat losses, while cooling needs result from solar gains, internal heat sources, and outdoor air infiltration. Since heating and cooling represent a significant portion of primary energy consumption in developed countries, minimizing these demands through improved building design is both an environmental necessity and economic priority. [6]

In terms of Estonia, a comprehensive study by Elering [7] employed multiple linear regression analysis to establish a base demand model and predict consumption growth from economic advancement, societal digitalization, and increased automation. The model utilized 11 years (2010-2021) of hourly electricity consumption data from 110 kV substations as the dependent variable, with 37 independent variables including Fourier transforms to capture seasonal patterns and dummy variables for categorical effects. To accurately model temperature sensitivity, the researchers converted ambient air measurements to heating and cooling degree hours, incorporating up to three days of heating degree hours to account for thermal inertia. Similar to findings in [1], the Elering model confirmed the significant impact of temperature variations on consumption patterns in Estonia.

1.3 Measuring the effects of weather on electricity consumption

Multiple regression analysis stands out as a viable method for quantifying weather impacts on electricity consumption across various studies. A benefit of regression approaches is that it enables clear and simple isolation of specific meterological effects while controlling for other variables.

Research in the Jiangtsu province in China utilised a two-way fixed effects regression model to analyze daily electricity consumption patterns [8]. The approach effectively controlled for both regional and temporal effects simultaneously. The meterological factors included wind speed variables, sunshine duration, precipitation, temperature variables, precipitation and cloud cover. The regression coefficients extracted from the model provided quantitative measures of each factor's influence across each subregion in the province.

Temperature consistently emerges as one of the most significant determinants of electricity consumption. National-level research in the US employed population-weighed cooling degree days to quantify temperature impacts. This approach revealed that "during a summer week, temperature-related energy output accounts for approximately 17% of the total output," increasing to 20% during heat waves. [9]

Similarly, research in Dhaka determined that "temperature accounted [for] about 75% of the total variance" in electricity consumption, with a reduction of 1℃ potentially saving 81 MV of electricity [10]. The Bangladesh study utilized multiple linear regression incorporating "ambient temperature, relative humidity, wind speed, wind direction, and precipitation" [10], employing Pearson's correlation to assess relationship strength.

More advanced research has examined broader climate phenomena, such as El Niño effects on electricity consumption [11]. This work introduced a composite "Body Feeling Temperature" measure to account for complex meteorological interactions, using "Pearson Analysis to measure the correlation between weather and electricity" and extracting "weather-used electricity from whole society electricity using least square method" [11].

1.4 Multiple linear regression

The linear regression model has the form:

$$Y = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \tag{1.1}$$

where we have an input vector $X^T = (X_1, X_2, \ldots, X_p)$ and want to predict a real-valued output Y. The model either assumes the regression function $E(Y|X)$ to be linear or that the linear model is a reasonable approximation. Here β_j represents unknown coefficients, and X_j can be sourced from either quantitative inputs, transformations of quantitative inputs, interactions between variables or numeric encodings of inputs. No matter the source of X_j, the model is linear in the coefficients. The linear model with $p > 1$ is called the multiple linear regression model. [12]

Typically a set of training data $(x_1, y_1) \ldots (x_N, y_N)$ is provided to estimate the parameters β. An observation of feature measurements for the ith case may be given in vector form as $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$. Coefficients may be then estimated using the least squares method, in which coefficients are picked to minimize the residual sum of squares:

$$RSS(\beta) = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 . \tag{1.2}$$

[12]

Denoting the $N \times (p+1)$ matrix as $\mathbf{X}$ with each row as an input vector (with a 1 in the first position), and similarly letting $\mathbf{y}$ be the N-vector of outputs in the training set, we can write the residual sum of squares as:

$$RSS(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta). \tag{1.3}$$

Which is a quadratic function in the $p+1$ parameters. Differentiating with respect to β yields:

$$\begin{aligned} \frac{\partial RSS}{\partial \beta} &= -2\mathbf{X}^T (\mathbf{y} - \mathbf{X}\beta) \\ \frac{\partial^2 RSS}{\partial \beta \partial \beta^T} &= 2\mathbf{X}^T \mathbf{X}. \end{aligned} \tag{1.4}$$

Assuming that $\mathbf{X}$ has full column rank, we set the first derivative to zero

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\beta) = 0 \tag{1.5}$$

and obtain the unique solution

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{1.6}$$

The predicted values at input vector x_0 can be given by $\hat{y}_0 = (1 : x_0)^T\hat{\beta}$ and the fitted values at the training inputs are:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{1.7}$$

[12]

We assume the deviations of Y around its expectation to be additive and Gaussian. Thus

$$Y = E(Y|X_1, \ldots, X_p) + \varepsilon = \beta_0 + \sum_{j=1}^{p} X_j\beta_j + \varepsilon, \tag{1.8}$$

where the error $\varepsilon \sim N(0, \sigma^2)$. [12]

Hypothesis testing can then be performed to determine which coefficients significantly contribute to the model. For individual coefficients, we test if $\beta_j = 0$ using the t-statistic

$$t = \hat{\beta}_j/SE, \tag{1.9}$$

where SE equals $\hat{\sigma}$ times the square root of the jjth element of $(\mathbf{X}^T\mathbf{X})^{-1}$. The $P > |t|$ value quantifies the probability of observing a t-statistic at least as extreme as the one calculated, assuming the null hypothesis is true. Lower p-values indicate stronger evidence against the null hypothesis. Typically, if $P < 5\%$, then $\hat{\beta}_j$ is considered statistically significant, while $P < 1\%$ indicates the coefficient is highly significant. [13]

1.5 The coefficient of determination

After fitting a multiple linear regression model, we can decompose the total variance in the response variable as:

$$\text{var}(Y) = \text{var}(\mathbf{X}\hat{\beta}) + \text{var}(\varepsilon), \tag{1.10}$$

where var$(\mathbf{X}\hat{\beta})$ represents the "explained variance" or variance captured by the fitted values and var(ε) represents the "unexplained" or residual variance. [13]

The coefficient of determination, denoted R^2, is defined as [13]:

$$R^2 = \frac{\text{var}(\mathbf{X}\hat{\beta})}{\text{var}(Y)}. \tag{1.11}$$

Here R^2 represents the proportion of the total variance in Y that is explained by the regression model. It ranges from 0 to 1, where:

- $R^2 = 0$ indicates the model explains none of the variance in Y;
- $R^2 = 1$ indicates the model explains all the variance in Y.

We may use R^2 also as a measure of goodness of fit. Higher values indicate the model fits the data well (small residuals relative to the variation in Y), while lower values suggest a poor fit. [13]

2. Methodology

This chapter presents the data and methodology used in our analysis of electricity consumption patterns in Estonia. Section 2.1 provides an overview of the datasets, describing their collection methods and original format. Section 2.2 details the pre-processing steps required to prepare the data for model development and offers a closer insight into the feature engineering. Sections 2.4 and 2.5 present the methodology and theoretical basis for fitting the multiple linear regression model.

All data manipulation, model fitting, and visualizations were executed using the Python programming language and its relevant libraries.

2.1 Datasets

2.1.1 Electricity consumption dataset

The electricity consumption data has been provided to Tallinn University of Technology on behalf of Elektrilevi OÜ for study and research purposes under the agreement JV-ARI-18/27941. This information is not publicly accessible.

The dataset consists of two main components: business and private electricity consumption data, both covering the period from January 1, 2019 to December 5, 2021 with 1 070 measurement dates. The data spans all 15 counties in Estonia, with measurements collected from 3 084 villages for business consumers and 3 875 villages for private consumers. For business consumption, the dataset includes 3 993 records representing 69 228 measurement points. For private consumption, there are 5 870 records representing 577 171 measurement points.

2.1.2 Weather dataset

The weather dataset was obtained from the Copernicus Climate Data Store database. Copernicus CDS provides comprehensive information about climate variables on a global, continental and regional scale. It contains data types including satellite observations, in-situ measurements, climate model projections and seasonal forecasts. [14]

ERA5-Land hourly data from 1950 to present was chosen for the specific dataset. ERA5-Land has been produced by replaying the land component of the European Centre for Medium-Range Weather Forecasts ERA5 climate reanalysis. Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset. Its high spatial resolution of $0.1° \times 0.1°$ captures local weather conditions with precision, while maintaining temporal consistency [15]. The dataset provides comprehensive climate parameters directly influencing electricity consumption. The specific variables requested are provided in Table 1.

Table 1. ERA5 Land Variables

Variable	Description	Unit
`t2m`	2-meter temperature	K
`d2m`	2-meter dewpoint temperature	K
`sp`	surface pressure	Pa
`u10`	10-meter eastward wind component	m/s
`v10`	10-meter northward wind component	m/s
`ssrd`	surface solar radiation downward	J/m^2
`tp`	total precipitation	m

The data was obtained in `grib` format via the CDS Application Programming Interface with a Python-based calling function that collected the relevant variables sampled at an hourly interval for each Estonian county. The counties were defined by their bounding box coordinates visualized in Figure 1. In order to keep the request sizes reasonable, each month for years 2019 to 2021 was requested and downloaded separately for an individual county. Then the `grib` files were converted into `csv` format for ease of use. Converted files contained the following information: time, latitude and longitude with the respectful measurement for each of the variables.

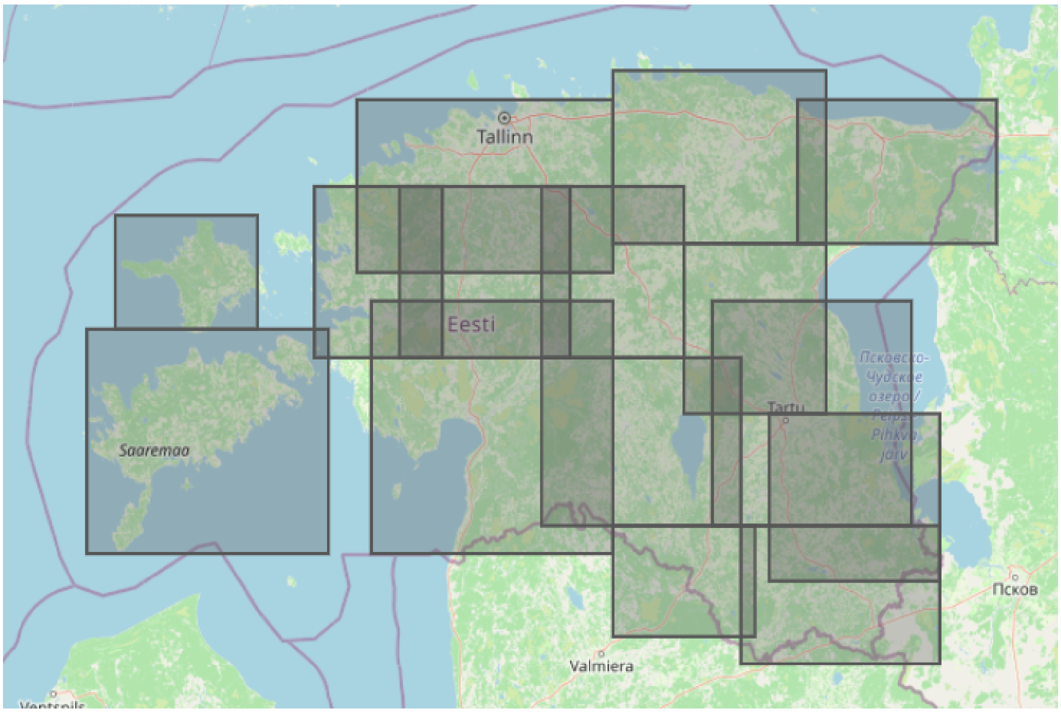

Figure 1. Bounding boxes for Estonian counties

2.2 Preprocessing and feature engineering

2.2.1 Approach to electricity consumption data

Consumption values were aggregated by sum for each county. In contrast to conventional approaches, outliers – representing significant drops or spikes in consumption – were deliberately retained rather than filtered out. These anomalies capture real-world scenarios such as electricity grid disruptions that operational models must accommodate, or could serve diagnostic purposes. Additionally, removing outliers risked eliminating valuable data points potentially correlated with extreme weather events. By preserving these variations, the model training baseline maintains greater fidelity to actual operational conditions, albeit with increased data variability.

The private consumption dataset exhibits substantial quality issues, with all counties showing over 20% missing values. The most affected regions include Võru County (57.4%), Hiiu County (42%), and Saare County (39.6%). Both private and business datasets display notable gaps, particularly on April 18 and 20, 2021, when recording issues affected 30-70% of measurements across counties. Business consumption data maintains good overall quality with no significantly affected counties, but still shows similar anomalous spikes in 2021.

Given these data quality challenges, the decision was made to restrict the analysis to data prior to March 11, 2020 - immediately before a state of emergency was declared in Estonia due to COVID-19 [16]. This choice was informed by research [1] demonstrating that linear regression models can experience substantially larger residuals during the pandemic period, if the effects are not accounted for. This approach also conveniently avoids the problematic data points concentrated in 2021.

2.2.2 Approach to weather data

Temperature variables, solar radiation and surface pressure were aggregated by mean for all coordinates to get a common hourly representation for the whole county being examined. Temperature and dewpoint temperature were then converted to degrees Celsius and in conjunction with surface pressure were used to calculate wet-bulb temperature using the code provided in [17]. Solar radiation and precipitation were converted into an hourly differential form since they are represented as integrated variables over the whole day in the ERA5-Land dataset. Another benefit of working with hourly data was that day and night periods could be differentiated for more precise feature engineering. The day hours were chosen as ranging from 04:00 to 00:00, making the rest of the 24-hour period night hours.

Thermal inertia was modeled using an exponentially weighted moving average in order to simulate how building mass delays temperature response as:

$$T_d(t) = \sum_{i=0}^{L} w_i T(t-i), \tag{2.1}$$

where T_d is the delayed hourly temperature at time t, L is the length of the window for past temperatures to be considered. This means that we can take into account both the immediate and delayed effects of dry- and wet-bulb temperature. The weights follow an exponential decay pattern:

$$w_i = \frac{e^{-\frac{i}{\tau}}}{\sum_{j=0}^{L} e^{-\frac{j}{\tau}}}, \tag{2.2}$$

with τ denoting the thermal time constant.

When outdoor temperature changes, the effects through building walls are not felt immediately unlike immediate effects through windows and ventilation, as pointed out in [5]. The response curve effectively rises to a maximum and then falls as time increases. As a consequence of this the immediate temperatures are overestimated in the model, meaning that immediate effects were included separately to counteract this.

The parameters for business consumers were chosen as $L = 120$ h, $\tau = 29$ h and for private consumers as $L = 96$ h, $\tau = 16$ h. The physical interpretation of these constants is that larger, either industrial or office, buildings in the business sector generally present heavier thermal mass, thus leading to the choice of a longer time window and slower decay rate. Conversely, smaller parameter values for private sector buildings represent lighter thermal mass.

The thermal mass parameters and subsequent model parameters were empirically tuned using data from Harju County to optimise overall model fit. Despite this empirical tuning, the parameter values remain within reasonable ranges and preserve meaningful interpretability of the regression results.

The same approach was used for solar radiation, with smaller values for τ chosen as 10 h and 6 h for business and private consumers respectively. In order to take into account the total thermal effects of immediate and delayed solar radiation in combination with temperature, a unit conversion had to be done since the original solar radiation variable represents energy flux in J/m^2. The solar factors s_1, s_2, s_3 for unit conversion were chosen as $s_1 = 4 \cdot 10^{-6}$, $s_2 = 3.5 \cdot 10^{-6}$, $s_3 = 0.6 \cdot 10^{-6}$ for private consumers and as $s_1 = 8 \cdot 10^{-6}$, $s_2 = 1 \cdot 10^{-6}$, $s_3 = 0.4 \cdot 10^{-6}$ for business consumers. Factor s_1 was used for combination with dry-bulb temperature, and was chosen with the largest magnitude to account for the heating effect of solar radiation during colder periods, s_2 was used for combination with wet-bulb temperature and chosen to be slightly smaller, to represent a slightly decreased importance of solar heating in warm, humid periods and s_3 the smallest for delayed effects. The factors have implicit units of ℃·m²/J, thus allowing to interpret the values as effective temperature increase for every joule of solar energy per square meter.

The combined thermal effect of temperature and solar radiation or effective temperature was then found as:

$$T_{\text{eff}}(t) = T_d(t) + s_k S(t) + s_3 S_d(t), \tag{2.3}$$

where S is the immediate solar heating effect and S_d represents the thermal mass heating from accumulated solar energy, $k = 1, 2$ denotes either dry- or wet-bulb constant. Utilizing the effective temperatures for both wet-and dry-bulb variables we can then quantify the heating and cooling needs of consumers, which reflect the energy needed to maintain comfort when outdoor conditions are insufficient. Heating need was calculated as:

$$H(t) = (T_t(t) - T_{\text{eff}})\Theta(T_t(t) - T_{\text{eff}}), \tag{2.4}$$

where T_t is a dynamic threshold, chosen as 10.5℃ during the day and 9℃ during night for businesses and as 14℃ or 15℃ for private consumption. Here Θ represents the Heaviside function.

The choice of thresholding parameters reflects the fact that business buildings operate with lower heating thresholds due to internal heat generation from occupants and equipment, superior insulation standards, and economic incentives to minimize energy costs, while private buildings prioritize comfort with higher thresholds. The day-night differential accounts for reduced occupancy and metabolic heat production during off-hours in commercial buildings, whereas private buildings maintain more consistent temperature requirements throughout the day.

Cooling need can be similarly found using effective wet- and dry-bulb temperature as:

$$C_{\text{wet}}(t) = (T_{\text{wb,eff}} - T_{h,\text{wet}}(t))\Theta(T_{\text{wb,eff}} - T_{h,\text{wet}}(t)) \tag{2.5}$$

and

$$C_{\text{dry}}(t) = (T_{\text{db,eff}} - T_{h,\text{dry}}(t))\Theta(T_{\text{db,eff}} - T_{h,\text{dry}}(t)), \tag{2.6}$$

where $T_{\text{wb,eff}}$, $T_{\text{db,eff}}$ are the effective wet-and dry-bulb temperature, $T_{h,\text{wet}}(t)$ and $T_{h,\text{dry}}(t)$ are the respective cooling thresholds that vary by day and night. The wet-bulb thresholds are chosen as 16.0℃ for both building types during day and 18.0℃ for businesses or 17.0℃ for private consumers during night, while dry-bulb thresholds are set at 24.0 ℃ or 25.0 ℃ during day and 25.0 ℃ for both types during night.

The choice of cooling thresholds reflects the fact that business buildings employ lower dry-bulb cooling thresholds during operational hours due to higher internal heat loads from equipment, lighting, and occupant density, while private buildings prioritize comfort with slightly higher thresholds. The wet-bulb thresholds remain similar across building types as humidity-related comfort is more universal, but the day-night differential accounts for relaxed comfort standards during off-hours when occupancy is reduced and economic efficiency becomes more important than optimal comfort.

Furthermore, the temperature differentials are calculated to better quantify immediate heating and cooling needs. They may be formulated as:

$$\Delta T_H(t) = \begin{cases} T_{\text{room}} - T_{\text{db,eff}}(t) & \text{if } H(t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

and

$$\Delta T_C(t) = \begin{cases} T_{\text{db,eff}}(t) - T_{\text{room}} & \text{if } C(t) > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2.8}$$

where $\Delta T_H(t)$ represents the temperature deficit requiring heating and $\Delta T_C(t)$ represents the temperature excess requiring cooling.

Solar contributions are explicitly taken into account during periods where heating is needed as:

$$S_{\text{help,imm}}(t) = \begin{cases} S(t) & \text{if } H(t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

and

$$S_{\text{help,d}}(t) = \begin{cases} S_d(t) & \text{if } H(t) > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2.10}$$

where $S_{\text{help,imm}}(t)$ and $S_{\text{help,d}}(t)$ represent the immediate and delayed solar radiation contributions during heating periods, respectively. The reasoning being that solar radiation effectively reduces heating requirements by providing extra energy.

Heat loss is also enhanced by wind speed, which is taken into account via the wind chill effect during periods when heating is needed. Mathematically, it can be described as:

$$W(t) = \begin{cases} (T_{\text{room}} - T_{\text{eff}}(t)) \cdot v(t)^2 & \text{if } H(t) > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2.11}$$

where T_{room} is the target indoor temperature, $T_{\text{eff}}(t)$ is the effective outdoor dry-bulb temperature and $v(t)$ is the wind speed.

The square of wind speed is selected because wind-driven heat loss follows a quadratic relationship with wind velocity according to Bernoulli's principle, which describes how dynamic pressure increases with the square of fluid velocity. As wind speed increases, the pressure differential across building surfaces scales as $\Delta P \propto v^2$, driving increased air infiltration through cracks, gaps, and other openings in the building envelope. Poiseuille's law then governs how this pressure differential translates to flow rate as $Q \propto \Delta P \propto v^2$. [6]

All the mentioned variables were then aggregated by mean to obtain daily values for modeling energy consumption. This aggregation converts the hourly data into daily summaries while preserving the key relationships between weather conditions and energy demand established through the thermal modeling approach. Daily aggregation aligns with the natural periodicity observed in electricity consumption patterns and provides a suitable temporal resolution for modeling.

The lighting demand for buildings is modeled using weighted daily aggregations that account for solar radiation deficits during periods of building occupancy. The daily lighting need is formulated as:

$$L(d) = \sum_{h=0}^{23} w_h \cdot \max(0, S_{\text{max}} - \min(S_h, S_{\text{max}})), \tag{2.12}$$

where $L(d)$ represents the daily lighting need, w_h are the hourly weights reflecting building usage patterns, which were normalised at each hour, $S(t)$ is the solar radiation at hour $t = h$, and S_{max} is the maximum meaningful solar radiation threshold, set at 150 W/m^2. The capping operation $\min(S_h, S_{\text{max}})$ ensures that solar radiation values above the threshold do not contribute to reduced lighting needs, while the deficit $S_{\text{max}} - S_h$ quantifies the amount of artificial lighting required.

Similarly, precipitation exposure during active hours is calculated as a weighted daily sum

$$P(d) = \sum_{h=0}^{23} w_h \cdot \max(0, P_h(t)), \tag{2.13}$$

where $P(d)$ represents the weighted daily precipitation and $P_h(t)$ is the hourly precipitation. The weighting scheme, depicted in Figure 2, captures the varying importance of precipitation based on building usage patterns, reflecting when weather exposure most significantly impacts occupant comfort and building operations. The lower weights, which are taken as 0 at some intervals in the day, are a consequence of the fact that during those hours the effect of lighting combines with other solar radiation related effects, meaning that a better fit for the model was achieved when weights were lowered.

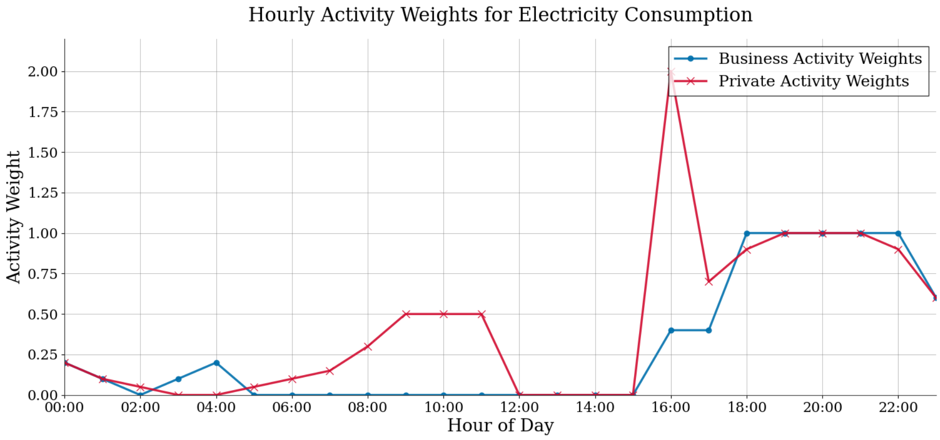

Figure 2. Daily weights w_h (before normalization) for lighting and precipitation

The total daily solar radiation available for solar energy applications was calculated as:

$$S_{\text{total}}(d) = \max_h(S_h), \tag{2.14}$$

representing the peak solar radiation value within the day, which serves as an indicator of photovoltaic potential for solar panels. This allows the quantification of reduced electricity consumption due to solar panels.

2.2.3 Approach to temporal data

Temporal variables represented days of the week using one-hot encoding, where each day was encoded as a binary vector with a one in the position corresponding to that day and zeroes elsewhere. The holiday dates were imported using the Python library `holidays` for ease of processing [18].

In the model implementation, holidays were treated as a special case: when a holiday occurred on a regular working day, it was encoded as if it were a Sunday to capture the similar consumption patterns. This encoding strategy effectively reflected the reduced business consumption characteristic of both weekends and holidays. For holidays that already fell on weekends, no additional encoding was done as they naturally exhibited weekend consumption patterns.

2.3 Implementation of models

We denote c_i $(i = 1, ..., N)$ as consumption on the i-th day as the response variable and let f_{ij} $(j = 1, ..., p)$ denote the factors affecting consumption. The $p = 18$ variables included for analysis were:

Temporal factors:

1. - 7. $f_{i1}, \ldots, f_{i7}$: One-hot encoded weekdays

Weather and thermal factors:

8. $f_{i8} = H$: Heating need
9. $f_{i9} = C_{\text{wet}}$: Wet-bulb cooling need
10. $f_{i10} = C_{\text{dry}}$: Dry-bulb cooling need
11. $f_{i11} = W$: Wind chill effect
12. $f_{i12} = \Delta T_H$: Temperature differential during heating periods
13. $f_{i13} = \Delta T_C$: Temperature differential during cooling periods
14. $f_{i14} = S_{\text{help,imm}}$: Immediate solar assistance during heating
15. $f_{i15} = S_{\text{help,d}}$: Delayed solar assistance during heating
16. $f_{i16} = L$: Lighting need
17. $f_{i17} = P$: Weighted precipitation exposure
18. $f_{i18} = S_{\text{total}}$: Total daily solar radiation

We formulate the general expression for our multiple linear regression model as:

$$\mathbf{C} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.15}$$

thus for a single observation

$$c_i = \sum_{j=1}^{p} f_{ij}\beta_j + \varepsilon_i. \tag{2.16}$$

The fit intercept is not used because it lacks physical meaning. When all consumption factors are zero, the consumption should logically be zero rather than some baseline offset, making a zero-intercept model more interpretable.

2.4 Model fitting and evaluation of variables

After initial preprocessing the variables were combined into one `csv` file, containing the load data and all of the features derived for both business and private consumers for each county to be analyzed.

Counties were grouped based on geographical proximity and consumption patterns to keep analysis regional. The four counties with the largest electricity consumption – Harju, Tartu, Lääne-Viru, and Pärnu – were analyzed individually due to their significant contribution to overall national consumption. The remaining counties were clustered into geographical regions: Western Counties (Hiiu, Saare, Lääne), Central Counties (Rapla, Järva, Viljandi), Southern Counties (Valga, Põlva, Võru), and Eastern Counties (Ida-Viru, Jõgeva). Features for counties were then weighed in proportion to each county's contribution to the overall consumption in the group for analysis.

The multiple linear regression models were developed for each county or region and client type, with model quality of fit measured by R^2 values. The range of dates from January 5 to June 20 2019 were used in order to avoid larger holidays and to have a balanced sample size from the hot and cold period. The specific modeling approach utilized was Ordinary Least Squares regression from the `statsmodels` library [19], which conveniently outputs coefficients, their standard errors, t-statistics, p-values, and confidence intervals for statistical inference. In order to quantify how many kWh of consumption each weather related coefficient accounted for, the coefficients were multiplied by the root mean square value of the respective feature and divided by the average daily consumption for a percentage representation. Utilizing percentage indicators here is useful, since each regions consumption and coefficient values can vary greatly.

Additionally, the percentage residuals were calculated as the relative error divided by the actual value to inspect the model's performance in the time domain. This helps track any problematic periods and provides a more intuitive comparison of results. Finally the results were collected for analysis. The full pipeline is visualised in Figure 3.

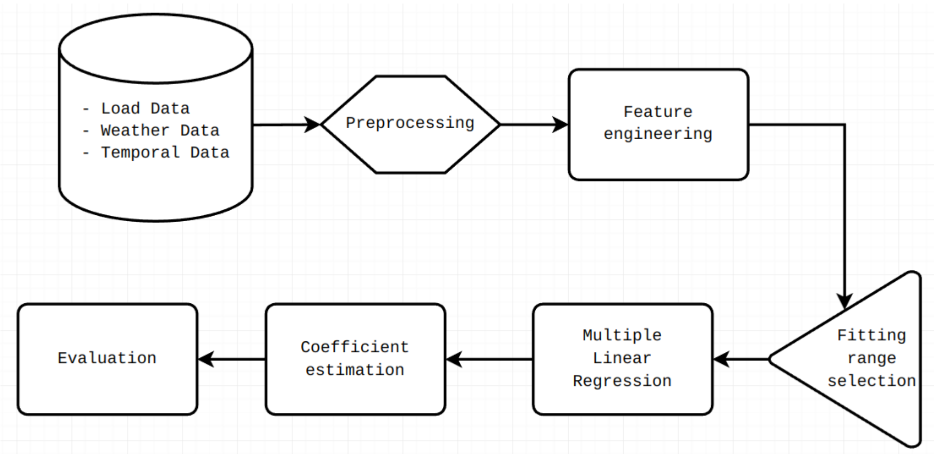

Figure 3. Schematic description of the model fitting and weather variable evaluation pipeline

3. Results and discussion

This section presents the building profile analysis across Estonian counties and regions, examining how weather variable impacts on electricity consumption reveal building construction quality, heating system characteristics, and energy efficiency patterns. Through quantification of weather variables' percentage impacts and statistical significance, distinctive regional consumption patterns and sector-specific responses are identified.

3.1 Harju County

Both regression models for Harju County demonstrate exceptional explanatory power with identical R^2 values of 0.993 for both business and private consumption. This indicates that 99.3% of the variance in electricity consumption can be explained by the variables included in the models, suggesting very strong predictive capability. This is confirmed by the relatively stable residuals as seen in Figure 4.

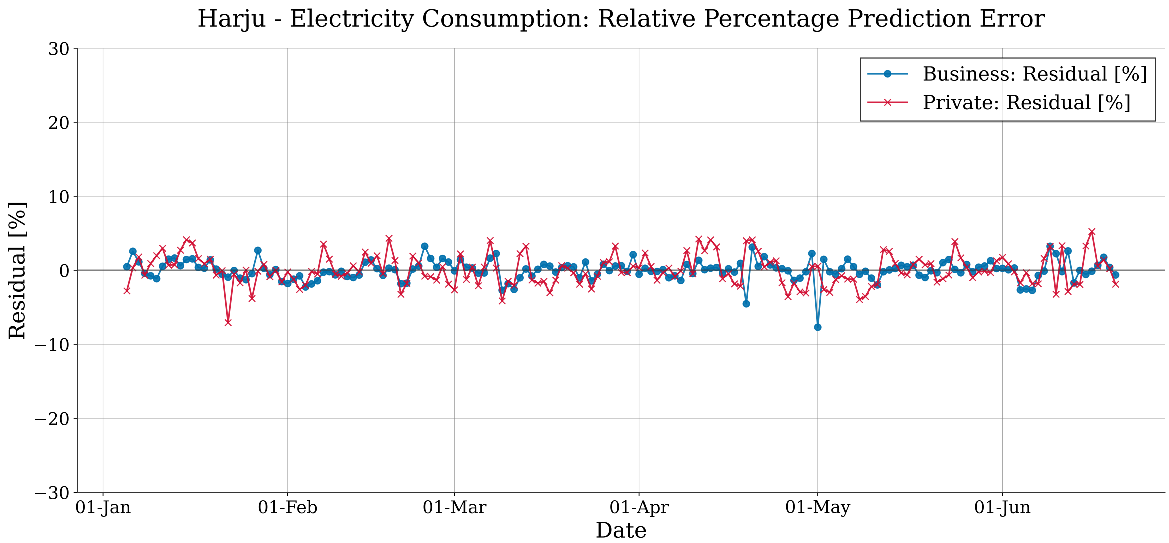

Figure 4. Percentage residuals for Harju business and private consumers

3.1.1 Business consumers

Weather effects on business electricity consumption in Harju are characterized by:

- **Temperature differentials during heating periods** (6.91%, $t = 5.06$, $P < 0.001$) are the most significant weather factor, indicating substantial heat loss through windows and ventilation systems in commercial buildings.
- **Dry bulb cooling needs** (3.17%, $t = 1.26$) show positive correlation with consumption but with lower statistical significance ($P = 0.211$), suggesting businesses may run cooling regardless of perceived temperature comfort.

- **Temperature differentials during cooling periods** (-3.16%, $t = -1.16$, $P = 0.249$) have a negative correlation, though with limited statistical significance.
- **Lighting requirements** (1.57%, $t = 4.43$, $P < 0.001$) significantly contribute to business electricity demand, indicating potential for energy savings through lighting optimization.
- **Wind chill** (0.84%, $t = 1.78$, $P = 0.076$) shows a modest effect, suggesting relatively good building envelope sealing in business structures.

3.1.2 Private consumers

Residential electricity consumption shows different responses to weather conditions:

- **Heating needs** (16.63%, $t = 6.66$, $P < 0.001$) represent the most significant weather variable for private consumption, indicating residential heating systems are a primary energy driver.
- **Cooling preferences** show an interesting pattern with dry-bulb temperature cooling needs positive (12.95%, $t = 3.18$, $P = 0.002$) but wet-bulb cooling needs negative (-3.43%, $t = -3.66$, $P < 0.001$), suggesting residents respond more to perceived temperature than thermometer readings.
- **Temperature differentials during cooling periods** (-11.99%, $t = -2.81$, $P = 0.006$) have significant negative correlation, indicating residents may reduce electricity use during hot weather by opening windows rather than using air conditioning.
- **Solar radiation** contributes positively ($S_{\text{help, imm}}$: 1.77%, $t = 3.41$, $P < 0.001$), suggesting solar heat gain through windows affects residential energy use.
- **Total solar radiation** (1.48%, $t = 1.37$, $P = 0.173$) shows positive but less significant impact, indicating potential for solar energy utilization.

3.2 Tartu County

Both regression models for Tartu County demonstrate strong explanatory power with R^2 values of 0.991 for business and 0.990 for private consumption. This indicates that approximately 99% of the variance in electricity consumption can be explained by the variables included in the models, suggesting excellent predictive capability. Again, the quality of fit is confirmed by the generally stable residuals seen in Figure 5.

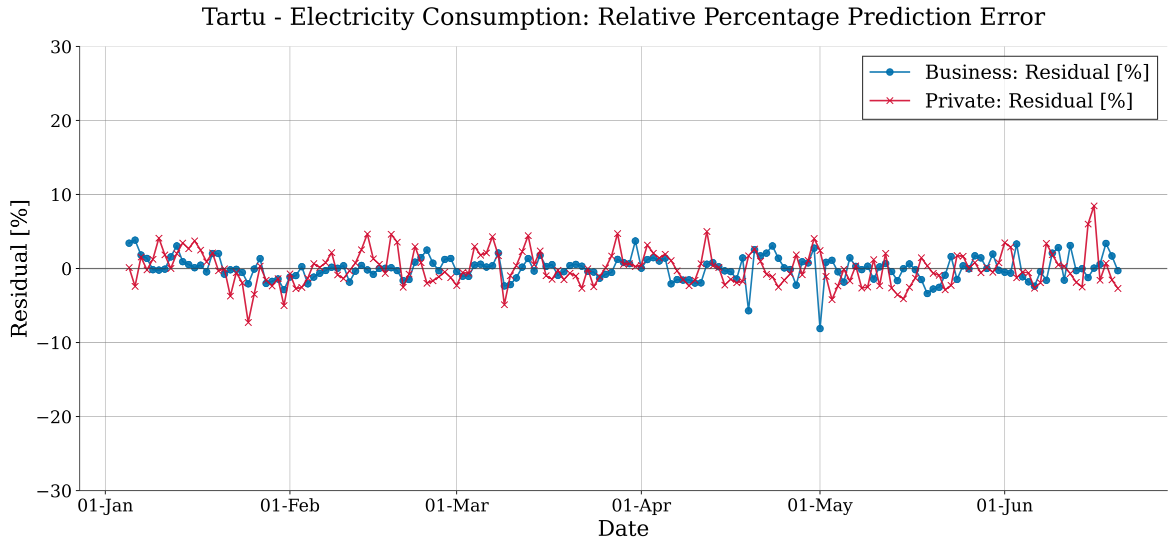

Figure 5. Percentage residuals for Tartu business and private consumers

3.2.1 Business consumers

Weather effects on business electricity consumption in Tartu are characterized by:

- **Temperature differentials during heating periods** (3.06%, $t = 2.80$, $P < 0.01$) Indicate moderate heat losses during heating periods.
- **Solar radiation** (-1.36%, $t = -2.47$, $P < 0.05$) has a significant negative impact, suggesting businesses may reduce electricity usage during periods of high solar radiation due to natural lighting or passive solar heating.
- **Delayed solar radiation effects** (-0.98%, $t = -2.78$, $P < 0.01$) negatively influence consumption, indicating accumulated heat through windows reduces energy needs.
- **Wet-bulb cooling needs** (0.70%, $t = 2.93$, $P < 0.01$) have a small but statistically significant positive impact, indicating businesses respond to humidity-related discomfort.
- **Direct heating variables** (H: 1.58%, $t = 1.31$; ΔT_H: 2.41%, $t = 1.45$) show notable impacts but lack statistical significance, suggesting more complex heating responses in Tartu's business consumers.

3.2.2 Private consumers

Residential electricity consumption shows different responses to weather conditions:

- **Heating needs** (13.47%, $t = 5.02$, $P < 0.001$) represent the most significant weather variable for private consumption, indicating residential heating systems are a primary energy driver with much stronger influence than in the business consumers.
- **Temperature differentials during heating periods** (8.47%, $t = 2.55$, $P < 0.05$) substantially affect residential consumption, suggesting significant heat loss through windows and ventilation in residential buildings.

- **Cooling preferences** show interesting patterns with wet-bulb cooling needs (-2.60%, $t = -2.18$, $P < 0.05$) having a significant negative correlation, while dry-bulb cooling needs show positive but non-significant impact (3.94%, $t = 0.88$, $P = 0.38$). This might indicate cooling via natural ventilation.
- **Lighting requirements** (1.52%, $t = 2.45$, $P < 0.05$) significantly contribute to private electricity demand, showing residents' dependence on artificial lighting.
- **Temperature differentials during cooling periods** show substantial negative impact (-2.75%, $t = -0.52$) but lack statistical significance, suggesting varied cooling behaviors among residents.

3.3 Lääne-Viru County

The regression models for Lääne-Viru County show a notable contrast in explanatory power with R^2 values of 0.862 for business and 0.990 for private consumption. This indicates that while the model captures residential consumption patterns very effectively, it has more limited success in explaining business consumption variability, suggesting more complex or unpredictable commercial energy usage patterns in this county. This is further confirmed by the residuals in Figure 6. While the residuals for private consumers are relatively stable, a periodic error can be seen in the business residuals. This suggests some underlying behaviour that is not being captured by the model, implying the need for incorporating cultural or economic features. The larger residuals near the beginning of May are possibly a result of spring holiday.

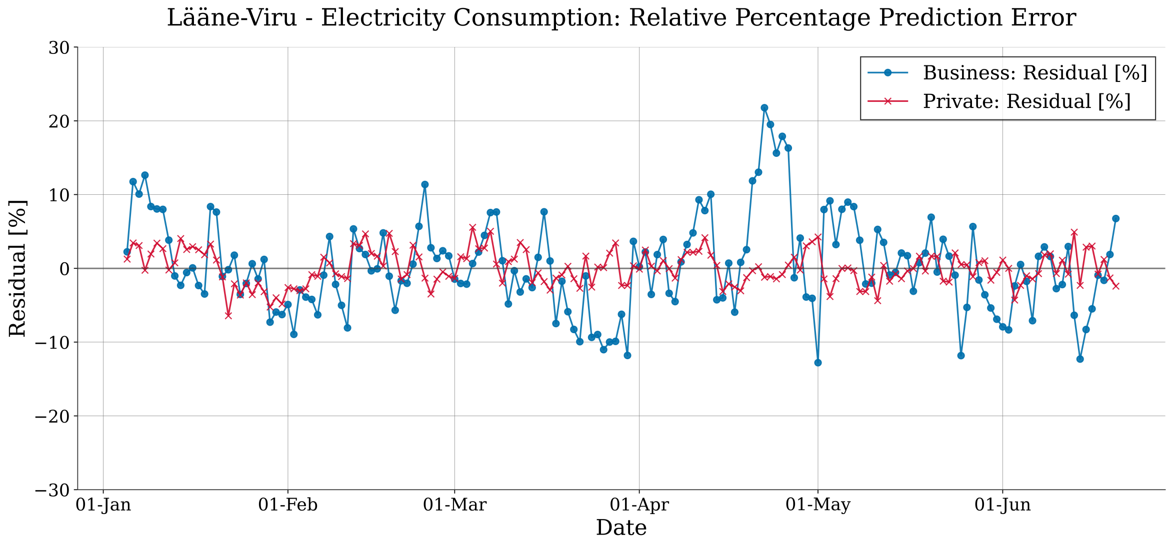

Figure 6. Percentage residuals for Lääne-Viru business and private consumers

3.3.1 Business consumers

Weather effects on business electricity consumption in Lääne-Viru show distinctive patterns:

- **Cooling needs based on dry-bulb temperature** (23.06%, $t = 1.93$, $P = 0.055$) have an exceptionally strong impact, nearly reaching statistical significance, suggesting businesses in this region are highly responsive to temperature-based cooling requirements.
- **Temperature differentials during cooling periods** (-22.01%, $t = -1.72$, $P = 0.087$) show a strong negative relationship, possibly caused by the compensatory behaviour of immediate effects.
- **Direct solar radiation** during heating periods (2.75%, $t = 3.79$, $P < 0.001$) is the only strictly statistically significant weather variable, suggesting solar heat gain through windows plays an important role in commercial buildings.
- **Lighting requirements** (-3.17%, $t = -1.97$, $P = 0.051$) show a borderline significant negative correlation, contrary to expectations, possibly indicating businesses substitute artificial lighting with natural light during periods of higher electricity demand.
- **Wind chill effects** (1.53%, $t = 1.81$, $P = 0.072$) show a positive borderline significant impact, suggesting building envelope issues that increase energy consumption during windy conditions.

3.3.2 Private consumers

Residential electricity consumption shows clearer and more consistent patterns:

- **Heating needs** (22.47%, $t = 7.20$, $P < 0.001$) demonstrate an exceptionally strong and highly significant impact, revealing heating as the dominant driver of residential electricity consumption.
- **Wet-bulb temperature cooling needs** (-3.64%, $t = -3.04$, $P < 0.01$) have a significant negative impact, indicating residents may reduce electricity usage during humid conditions, possibly by opening windows rather than using air conditioning.
- **Lighting requirements** (2.38%, $t = 3.64$, $P < 0.001$) significantly contribute to residential consumption, demonstrating the importance of artificial lighting in homes.
- **Precipitation** (-0.49%, $t = -1.95$, $P = 0.053$) shows a borderline significant negative effect, suggesting residents may reduce electricity consumption during rainy periods, possibly due to behavioral changes.
- **Direct solar radiation** (-1.01%, $t = -1.78$, $P = 0.078$) has a borderline significant negative impact, indicating solar heat gain may reduce reliance on electric heating in residential buildings.

3.4 Pärnu County

Both regression models for Pärnu County demonstrate strong explanatory power with R^2 values of 0.983 for business and 0.992 for private consumption. This indicates that over 98% of the variance in electricity consumption can be explained by the variables included in the models, suggesting excellent predictive capability in this coastal county. The residuals are generally stable, showing a small dip for business consumers near spring holiday as seen in

Figure 7.

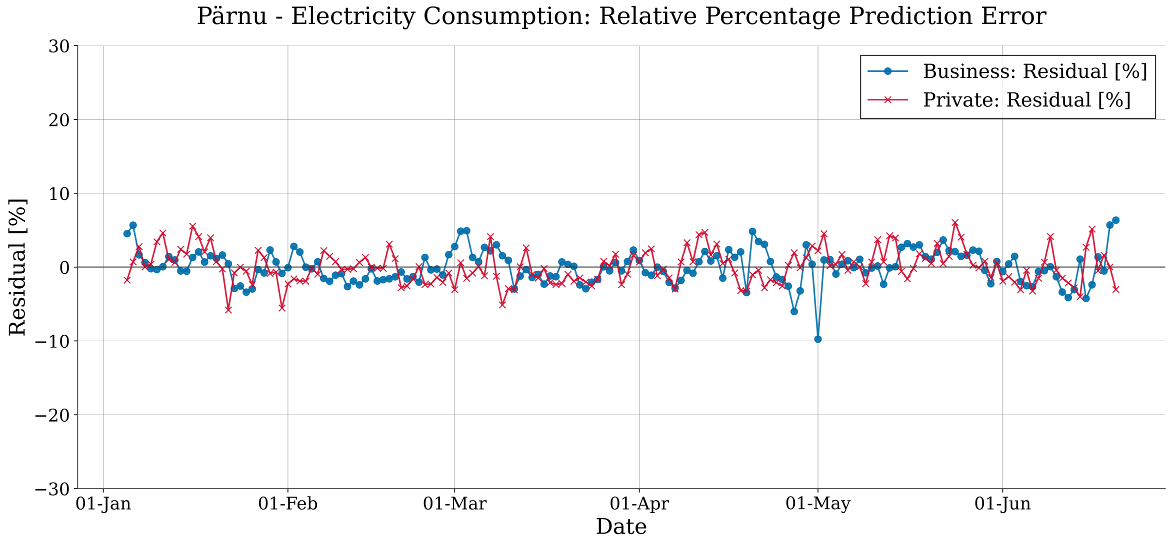

Figure 7. Percentage residuals for Pärnu business and private consumers

3.4.1 Business consumers

Weather effects on business electricity consumption in Pärnu are characterized by:

- **Temperature differentials during heating periods** (5.10%, $t = 3.33$, $P < 0.01$) show a strong influence, indicating considerable heat loss through windows and ventilation.
- **Lighting requirements** (3.37%, $t = 5.89$, $P < 0.001$) have a substantial and highly significant impact, suggesting businesses in Pärnu are particularly dependent on artificial lighting, possibly related to the county's northern latitude and coastal weather patterns.
- **Heating needs** (2.67%, $t = 1.65$, $P = 0.101$) show notable but non-significant impact, indicating that while heating is important, it may be overshadowed by other factors such as lighting and temperature differentials.
- **Wet-bulb cooling needs** (1.02%, $t = 3.10$, $P < 0.01$) have a smaller but statistically significant positive impact, suggesting humidity-related comfort plays a role in business electricity consumption.
- **Direct solar radiation during heating** (0.58%, $t = 2.08$, $P < 0.05$) shows a small but significant positive effect, indicating solar heat gain through windows contributes to business energy needs.

3.4.2 Private consumers

Residential electricity consumption shows different patterns:

- **Temperature differentials during heating periods** (14.60%, $t = 4.35$, $P < 0.001$) have an exceptionally strong and significant impact, suggesting substantial heat loss through windows and ventilation in residential buildings of this coastal county.

- **Heating needs** (9.15%, $t = 3.40$, $P < 0.001$) show strong significance but with less impact than temperature differentials, indicating that while base heating is important, the quality of insulation may be a more critical factor.
- **Lighting requirements** (3.79%, $t = 6.18$, $P < 0.001$) are highly significant, revealing the importance of artificial lighting in residential settings in this northern coastal area.
- **Delayed solar radiation effects** (-2.83%, $t = -3.30$, $P < 0.01$) have a significant negative impact, suggesting accumulated heat through windows helps reduce energy consumption in residential buildings.
- **Total solar radiation** (2.13%, $t = 2.06$, $P < 0.05$) has a positive significant effect, possibly suggesting solar energy is used as a viable energy alternative in this region.

3.5 Central Counties

Both regression models for Central Counties demonstrate excellent explanatory power with R^2 values of 0.991 for business and 0.990 for private consumption. This indicates that 99% of the variance in electricity consumption can be explained by the variables included in the models, suggesting very strong predictive capability in these inland regions. The generally stable residuals confirm the quality of fit, as seen in Figure 8.

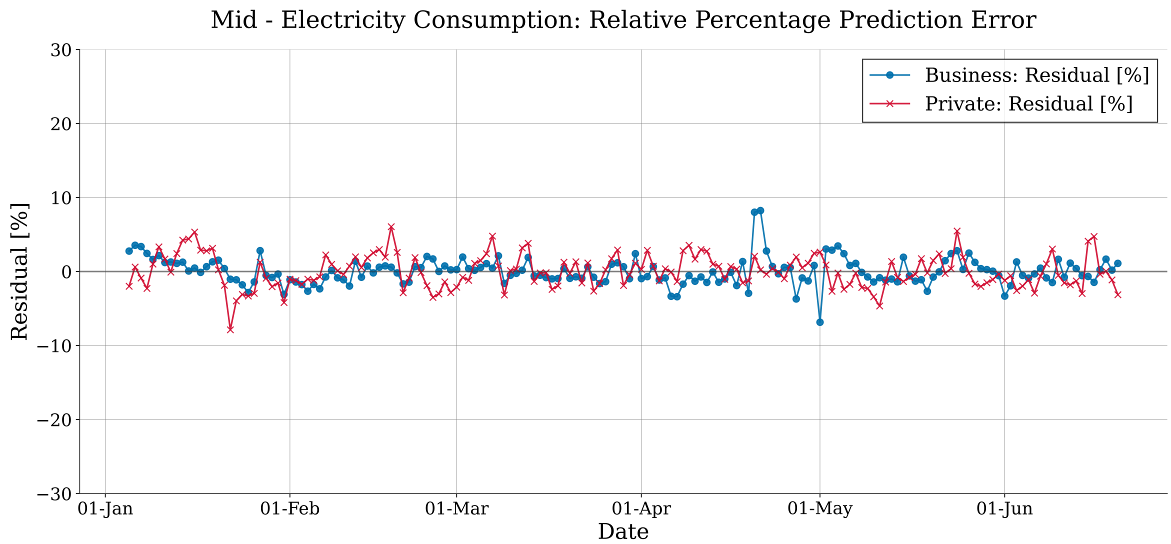

Figure 8. Percentage residuals for Central business and private consumers

3.5.1 Business consumers

Weather effects on business electricity consumption in Central Counties show a distinctive pattern:

- **Temperature differentials during cooling periods** (-9.02%, $t = -2.25$, $P < 0.05$) have a strong negative impact, again suggesting the compensatory behaviour of the immediate effects.

- **Dry-bulb cooling needs** (7.63%, $t = 2.09$, $P < 0.05$) show a significant positive correlation, indicating that temperature-based cooling requirements substantially drive electricity consumption in business buildings.
- **Heating needs** (2.70%, $t = 2.21$, $P < 0.05$) are significant but have less impact than cooling factors, which is notable given Estonia's northern climate and suggests well-insulated commercial buildings in these central regions.
- **Lighting requirements** (1.93%, $t = 4.39$, $P < 0.001$) are highly significant contributors to business electricity demand despite their relatively modest impact.
- **Direct solar radiation** during heating periods (0.88%, $t = 4.21$, $P < 0.001$) shows modest but highly significant impact, indicating solar heat gain consistently affects business energy consumption.

3.5.2 Private consumers

Residential electricity consumption shows different priorities:

- **Heating needs** (18.64%, $t = 6.77$, $P < 0.001$) demonstrate exceptionally strong and highly significant impact, dominating residential electricity consumption in these inland counties.
- **Lighting requirements** (3.70%, $t = 5.82$, $P < 0.001$) are also highly significant contributors, reflecting the importance of artificial lighting in residential buildings.
- **Wet-bulb cooling needs** (-2.37%, $t = -2.43$, $P < 0.05$) have a significant negative correlation, suggesting residents may reduce electricity usage when humidity-based discomfort increases, possibly through natural ventilation.
- **Precipitation** (-0.53%, $t = -2.24$, $P < 0.05$) has a small but significant negative impact, indicating weather-related behavioral changes affecting electricity consumption.
- **Dry-bulb cooling needs** (3.89%, $t = 0.86$, $P = 0.39$) show substantial but non-significant impact, suggesting greater variability in how residents respond to temperature-based cooling requirements.

3.6 Southern Counties

Both regression models for Southern Counties demonstrate strong explanatory power with R^2 values of 0.984 for business and 0.985 for private consumption. This indicates that over 98% of the variance in electricity consumption can be explained by the variables included in the models, suggesting excellent predictive capability in these southern regions. The residuals are generally stable but show a period of higher errors near the beginning of summer for business consumers, seen in Figure 9.

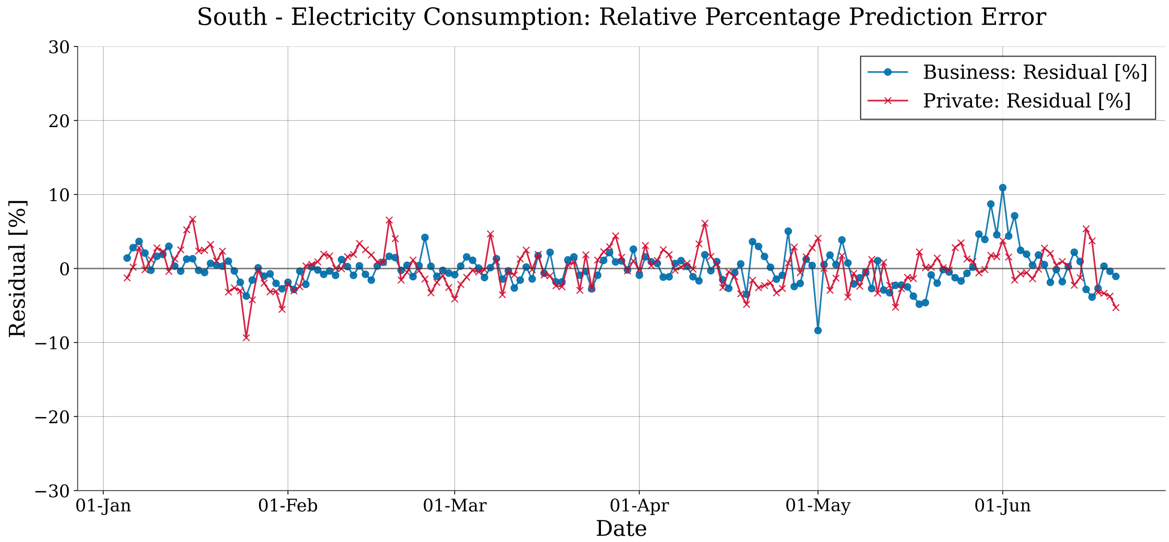

Figure 9. Percentage residuals for Southern business and private consumers

3.6.1 Business consumers

Weather effects on business electricity consumption in Southern Counties reveal distinctive patterns:

- **Temperature differentials during cooling periods** (14.69%, $t = 3.14$, $P < 0.01$) show an exceptionally strong positive impact, indicating air conditioning or cooling systems are heavily used in response to temperature increases.
- **Dry-bulb cooling needs** (-12.38%, $t = -2.88$, $P < 0.01$) have a significant but counterintuitively negative correlation, which is again explained by the possibility of the immediate effects dominating and counteracting the latent effects.
- **Temperature differentials during heating** (7.54%, $t = 5.21$, $P < 0.001$) shows a strong positive impact, showing heat losses during heating periods.
- **Total solar radiation** (-1.83%, $t = -2.83$, $P < 0.01$) has a negative impact, suggesting businesses effectively utilize natural light or passive solar heating, reducing electricity consumption.
- **Lighting requirements** (1.51%, $t = 2.85$, $P < 0.01$) significantly contribute to business electricity demand but with relatively modest impact compared to temperature-related factors.

3.6.2 Private consumers

Residential electricity consumption follows more conventional patterns:

- **Heating needs** (18.92%, $t = 6.64$, $P < 0.001$) are by far the most significant weather factor, confirming the primacy of heating systems in residential electricity consumption.
- **Lighting requirements** (2.68%, $t = 3.98$, $P < 0.001$) show substantial and highly

significant impact, underscoring the importance of artificial lighting in homes.

- **Wet-bulb cooling needs** (-1.91%, $t = -1.99$, $P < 0.05$) have a significant negative impact, suggesting residents may reduce electricity usage during humid conditions, possibly through natural ventilation.
- **Direct solar radiation during heating** (-1.26%, $t = -2.46$, $P < 0.05$) negatively affects consumption, indicating solar heat gain through windows helps reduce energy needs in residential buildings.
- **Temperature differentials during cooling** (3.37%, $t = 0.52$) and **dry-bulb cooling needs** (-2.32%, $t = -0.39$) show substantial but non-significant impacts, suggesting more complex and variable cooling behaviors among residents. The opposing directions can be explained due to the interleaved nature of the variables.

3.7 Eastern Counties

The regression models for Eastern Counties show a notable contrast in explanatory power with R^2 values of 0.816 for business and 0.982 for private consumption. This indicates that while the model effectively captures residential consumption patterns, it has more limited success in explaining business consumption variability, suggesting more complex or heterogeneous commercial energy usage patterns in this region than elsewhere in Estonia. This is verified by the residuals, showing higher variation for private consumers and highly unstable behaviour for businesses, seen in Figure 10. This is mostly caused by the effects of Ida-Virumaa, which highlights the need for additional, possibly economic, features and better handling of holiday effects.

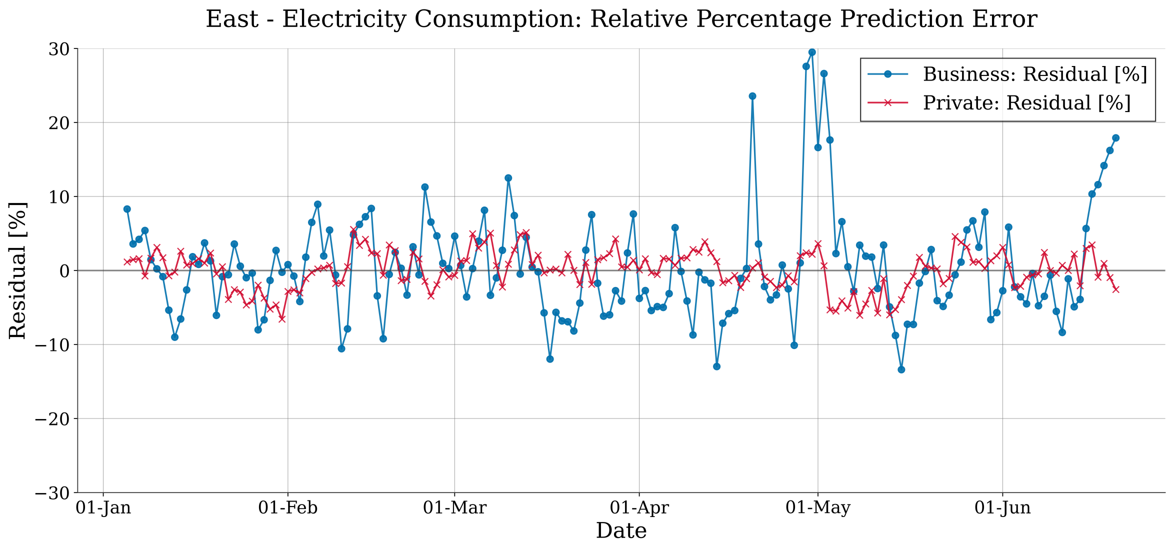

Figure 10. Percentage residuals for Eastern business and private consumers

3.7.1 Business consumers

Weather effects on business electricity consumption in Eastern Counties reveal several distinctive patterns:

- **Temperature differentials during heating periods** (-16.67%, $t = -2.30$, $P < 0.05$) show a strong significant negative impact, compensating the latent heating needs.
- **Heating needs** (10.26%, $t = 2.07$, $P < 0.05$) have a substantial positive impact, confirming the importance of base heating requirements.
- **Lighting requirements** (7.91%, $t = 4.56$, $P < 0.001$) demonstrate an exceptionally strong and highly significant effect, indicating greater dependence on artificial lighting in eastern business buildings than in other regions.
- **Cooling needs** (C_{dry}: 13.63%, $t = 1.11$, $P = 0.271$; ΔT_H: -13.16%, $t = -0.99$, $P = 0.324$) show large but non-significant impacts. The opposing directions appear again due to compensation.
- **Direct solar radiation** during heating (1.70%, $t = 2.16$, $P < 0.05$) has a significant positive effect, indicating solar heat gain through windows contributes to business energy needs.

3.7.2 Private consumers

Residential electricity consumption shows more conventional and consistent patterns:

- **Heating needs** (11.18%, $t = 3.45$, $P < 0.001$) are the dominant weather factor for residential consumption, though with somewhat less impact than in other regions of Estonia.
- **Temperature differentials during heating** (6.01%, $t = 1.49$, $P = 0.137$) show substantial but non-significant impact, suggesting variable insulation quality among residential buildings.
- **Wet-bulb cooling needs** (-2.88%, $t = -2.11$, $P < 0.05$) have a significant negative effect, indicating residents may reduce electricity usage during humid conditions through natural ventilation rather than electrical cooling.
- **Lighting requirements** (2.29%, $t = 3.29$, $P < 0.01$) significantly contribute to residential consumption with an impact similar to other Estonian regions.
- **Direct solar radiation** (-1.20%, $t = -2.14$, $P < 0.05$) shows a significant negative effect, suggesting effective utilization of passive solar heating in residential buildings.

Summary

This study analyzed how weather variables influence electricity consumption patterns across Estonian counties for both business and residential sectors, revealing insights about building characteristics and energy efficiency. The summary results are visually presented in Appendix 2. All multiple linear regression models demonstrated strong explanatory power at $R^2 \in [0.982, 0.993]$ except for business models in Lääne-Viru at $R^2 = 0.862$ and Eastern counties at $R^2 = 0.816$. The lower model performance in these regions is confirmed by unstable residual patterns, with Eastern counties showing particularly erratic business consumption behavior, largely attributed to Ida-Virumaa's economic characteristics and uncaptured holiday effects.

Heating vs Cooling Patterns: A distinctive regional divide emerged in business sector consumption. Southern and Central counties show cooling dominance (heating/cooling ratios 0.36 and 0.53), while Eastern counties uniquely exhibit heating dominance (ratio 1.24). Private consumption consistently shows heating dominance across all regions, with ratios ranging from 2.39 to 3.94.

Temperature Differentials: Temperature differentials during heating periods significantly impact consumption across most counties, with exceptionally strong effects in residential buildings in coastal Pärnu (14.60%) and strong impact in Eastern business buildings (-16.67%). This reveals building envelope quality issues in most regions.

Heating Needs: Direct heating requirements dominate residential consumption in all regions (11.18-22.47%), while showing more varied impact in businesses (1.58-10.26%), suggesting more diverse heating strategies in commercial buildings.

Lighting Requirements: Lighting consistently affects electricity consumption across all counties with higher significance in businesses (1.51-7.91%) than in residences (1.52-3.79%), with Eastern businesses showing the greatest dependence on artificial lighting.

Cooling Preferences: Businesses and residences respond differently to cooling needs. Business consumption in Lääne-Viru and Southern counties shows strong responsiveness to cooling metrics (up to 23.06%), while residential consumption typically shows negative correlation with wet-bulb cooling needs, indicating natural ventilation is preferred over air conditioning in homes.

The results reveal several opportunities for optimisations:

- **Building envelope improvements** could yield significant savings, particularly in regions with high temperature differential impacts.
- **Lighting optimization** presents consistent savings opportunities across all regions, especially in Eastern business buildings. As a solution LED lighting retrofits could

provide immediate energy savings.

- **Heating system efficiency** should be prioritized for the residential sector in all regions, with Lääne-Viru showing the highest potential benefits.
- **Cooling strategy assessments** for businesses in Southern and Central counties could identify efficiency opportunities, given their cooling-dominant consumption patterns.
- **Solar energy potential** exists in several regions, particularly Pärnu, where solar radiation significantly impacts consumption. However, the generally non-significant effects in 2019 suggest residents could employ more solar panels to better capitalize on available solar resources.
- **Economic and cultural modeling** should be incorporated comprehensively for regions showing poor weather-based predictive performance, particularly Eastern counties, to capture industrial patterns and holiday effects.
- **Enhanced weather data aggregation strategies** could utilize adaptive geographical weighting based on past consumption correlations, ensuring weather inputs better reflect the actual demand centers within each region.
- **Parameter tuning** should be done for each region separately.

Acknowledgements

The author of the thesis would like to acknowledge the supervision, support and contributions of his supervisor Jaan Kalda, who provided invaluable guidance throughout this research. Professor Kalda's comprehensive contributions – from refinements to the feature engineering to insightful interpretations of results – were instrumental in developing this interdisciplinary work. The author also thanks his parents for wise recommendations and copious amounts of proofreading, and his girlfriend for motivation.

Additionally, it is noted here that Claude AI was used for language correction purposes and Sematic Scholar provided an essential source for research information.

References

[1] J. Belikov and J. Kalda. “Impact of COVID-19 Pandemic on Energy Demand in Estonia”. In: *2022 IEEE Power & Energy Society General Meeting (PESGM)*. Denver, CO, USA, 2022, pp. 1–5. DOI: 10.1109/PESGM48719.2022.9916825.

[2] Naqash Ahmad et al. “Load Forecasting Techniques for Power System: Research Challenges and Survey”. In: *IEEE Access* 10 (2022), pp. 71054–71090. DOI: 10.1109/ACCESS.2022.3187839.

[3] Tahreem Anwar et al. “Introduction to Load Forecasting”. In: *International Journal of Pure and Applied Mathematics* 119.15 (2018), pp. 1527–1538. ISSN: 1314-3395.

[4] ENTSO-E (TF TRAPUNTA). *Demand Forecasting Methodology*. Rep. Version 1.1. European Network of Transmission System Operators for Electricity, 2019. URL: https://eepublicdownloads.entsoe.eu/clean-documents/sdc-documents/MAF/2020/Demand_forecasting_methodology_V1_1.pdf.

[5] Stijn Verbeke and Amaryllis Audenaert. “Thermal inertia in buildings: a review of impacts across climate and building use”. In: *Renewable and Sustainable Energy Reviews* 82.3 (2018), pp. 2300–2318. ISSN: 1364-0321. DOI: 10.1016/J.RSER.2017.08.083.

[6] Hugo S.L.C. Hens. *Building Physics - Heat, Air and Moisture: Fundamentals and Engineering Methods with Examples and Exercises*. A Wiley Company. Berlin: Ernst & Sohn Verlag für Architektur und technische Wissenschaften GmbH und Co. KG, 2007. ISBN: 978-3-433-01841-5.

[7] Elering. *Electricity Demand Scenarios*. Tech. rep. Elering AS, Oct. 2022. URL: https://elering.ee/sites/default/files/2022-10/Study%20-%20Electricity%20demand%20scenarios.pdf.

[8] Liang Xie et al. “Analysis of Meteorological Factors on Power Supply-Demand Balance: A Case Study of Jiangsu Province”. In: *2024 5th International Conference on Smart Grid and Energy Engineering (SGEE)*. IEEE. Nov. 2024, pp. 268–271. DOI: 10.1109/SGEE64306.2024.10865933.

[9] Douglas M Le Comte and Henry E Warren. “Modeling the impact of summer temperatures on national electricity consumption”. en. In: *J. Appl. Meteorol.* 20.12 (Dec. 1981), pp. 1415–1419.

[10] Arif Istiaque and Shahidul Islam Khan. “Impact of ambient temperature on electricity demand of Dhaka city of Bangladesh”. In: *Energy Power Eng.* 10.07 (2018), pp. 319–331.

[11] B Miao, J Y Lin, and C Liu. “Research on relation between El Nino climate and summer electricity consumption”. In: *IOP Conf. Ser. Earth Environ. Sci.* 108 (Jan. 2018), p. 052006.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. Springer Series in Statistics. New York, NY: Springer, 2009. ISBN: 978-0-387-84857-0. DOI: 10.1007/978-0-387-84858-7.

[13] David A. Freedman. *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press, 2009.

[14] Copernicus Climate Change Service. *Climate Data Store (CDS)*. 2025, [Online] Available https://climate.copernicus.eu/the-climate-data-store, Accessed Apr. 15, 2025.

[15] J. Muñoz Sabater et al. *ERA5-land post-processed daily-statistics from 1950 to present*. Accessed: Apr. 14, 2025. 2024. DOI: 10.24381/cds.e9c9c792.

[16] ERR. *Valitsus kuulutas välja eriolukorra*. 2020, [Online] Available https://www.err.ee/1063213/valitsus-kuulutas-valja-eriolukorra, Accessed Apr. 26, 2025.

[17] smartlixx. *WetBulb: Python code to calculate wet bulb temperature from dry bulb temperature and dew point using Davies-Jones (2008) method*. 2024, [Online] Available https://github.com/smartlixx/WetBulb, Accessed May 13, 2025.

[18] Vacanza Team. *Holidays: A Python library for working with holidays and calendars*. 2023, [Online] Available https://pypi.org/project/holidays/, Accessed May 15, 2025.

[19] statsmodels Developers. *statsmodels.regression.linear_model.OLS*. 2025, [Online] Available https://www.statsmodels.org/dev/generated/statsmodels.regression.linear_model.OLS.html, Accessed May 20, 2025.

Abstract

The objective of this thesis was to create multiple regression models to accurately assess the impact of weather factors on electricity consumption of business and private customers in Estonian counties. Using the models developed in this work, the effect of each observed weather factor on electricity consumption was described through percentage impact estimates of the model coefficients.

All regression models showed high coefficients of determination $R^2 \in [0.982, 0.993]$, except for commercial models in Lääne-Viru County at $R^2 = 0.862$ and Eastern Estonian counties (Ida-Viru, Jõgeva) at $R^2 = 0.816$. Several important conclusions were drawn based on the impact estimates of weather coefficients in the developed models.

In all Estonian counties, heating demand has the greatest impact on residential electricity consumption (11.18%–22.47%). For commercial customers, the impact of heating demand varies (1.58%–10.26%), indicating more diverse heating methods. For commercial customers in Southern and Central Estonia, cooling demand has the greatest impact on electricity consumption, whereas for Eastern Estonian commercial customers, heating demand has the strongest influence.

Cooling demand affects commercial and residential electricity consumption differently. Western-Viru and Southern Estonian commercial enterprises show high sensitivity to cooling parameters (up to 23.06%). Among residential consumers, the correlation with wet-bulb temperature-based cooling demand is negative, suggesting that natural ventilation is preferred over air conditioning for cooling.

Temperature differences during the heating season significantly affect electricity consumption in most counties. Particularly strong effects were observed among Pärnu residential consumers (14.60%) and Eastern Estonian commercial enterprises (-16.67%). The results may indicate problems with building thermal insulation.

Lighting demand affects electricity consumption uniformly across all counties. The impact of lighting demand is greater for commercial customers (1.51%–7.91%), while for residential customers it is smaller (1.52%–3.79%). Eastern Estonian commercial enterprises show the greatest dependence on lighting.

The county-specific weather sensitivity profiles described in this work enable the implementation of targeted energy efficiency measures and more effective grid management and energy planning, taking into account Estonia's regional characteristics.

Annotatsioon

Lõputöö eesmärgiks oli luua mitmese regressiooni mudeleid ilmastikufaktorite täpse mõju hindamiseks Eesti maakondade äri- ning eraklientide elektritarbimisele. Töö tulemusena loodud mudelite abil kirjeldati iga vaadeldud ilmastikufaktori mõju elektritarbimisele, kasutades mudeli koefitsentide protsentuaalseid mõjuhinnaguid.

Kõik regressioonimudelid näitasid kõrget determinatsioonikordajat väärtusega $R^2 \in [0.982, 0.993]$, välja arvatud ärimudelid Lääne-Viru maakonnas tulemusega $R^2 = 0,862$ ja Ida-Eesti maakondades (Ida-Viru, Jõgeva) tulemusega $R^2 = 0,816$. Loodud mudelite ilmastikukoefitsentide mõjuhinnangute põhjal tehti mitmed olulised järeldused.

Kõigis Eesti maakondades on suurim mõju eraklientide elektritarbimisele kütmisvajadusel (11,18-22,47%). Äriklientide puhul on kütmisvajaduse mõju varieeruv (1,58-10,26%), viidates mitmekesisematele kütmismeetoditele. Lõuna- ja Kesk-Eesti äriklientide elektritarbimisele on suurim mõju jahutusvajadusel, seevastu Ida-Eesti äriklientide elektritarbimist mõjutab enim kütmisvajadus.

Jahutusvajadus mõjutab äri- ja eraklientide elektritarbimist erinevalt. Lääne-Viru ja Lõuna-Eesti äriettevõtted näitavad kõrget tundlikkust jahutusparameetritele (kuni 23,06%). Eratarbijate seas on korrelatsioon märja termomeetri näidu põhise jahutusvajdusega negatiivne, mis viitab, et konditsioneeriga jahutamisele eelistatakse looduslikku ventilatsiooni.

Temperatuurierinevused küttehooajal mõjutavad märkimisväärselt elektritarbimist enamikes maakondades. Eriti tugevad mõjud ilmnesid Pärnu eratarbijate seas (14,60%) ja Ida-Eesti äriettevõtete seas (-16,67%). Tulemused võivad viidata probleemidele ehitiste soojustusisolatsioonis.

Valgustusvajadus mõjutab elektritarbimist ühtlaselt kõigis maakondades. Valgustusvajaduse mõju on suurem äriklientide puhul (1,51-7,91%), erakilentide puhul on see väiksem (1,52-3,79%). Suurim sõltuvus valgustusest on Ida-Eesti äriettevõtetel.

Töös kirjeldatud maakondlikud ilmastikutundlikkuse profiilid võimaldavad sihipäraste energiatõhususmeetmete rakendamist ning tõhusamat võrguhaldust ja energiaplaneerimist, arvestades Eesti piirkondlikke iseärasusi.

Appendices

Appendix 1 – Non-Exclusive License for Reproduction and Publication of a Graduation Thesis

[1]

I Marti Kütt

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis "Impact of Weather Factors on the Electricity Consumption of Estonia", supervised by Jaan Kalda
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

23.05.2025

[1] The non-exclusive license is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive license, the non-exclusive license shall not be valid for the period.

Appendix 2 – Variable impacts

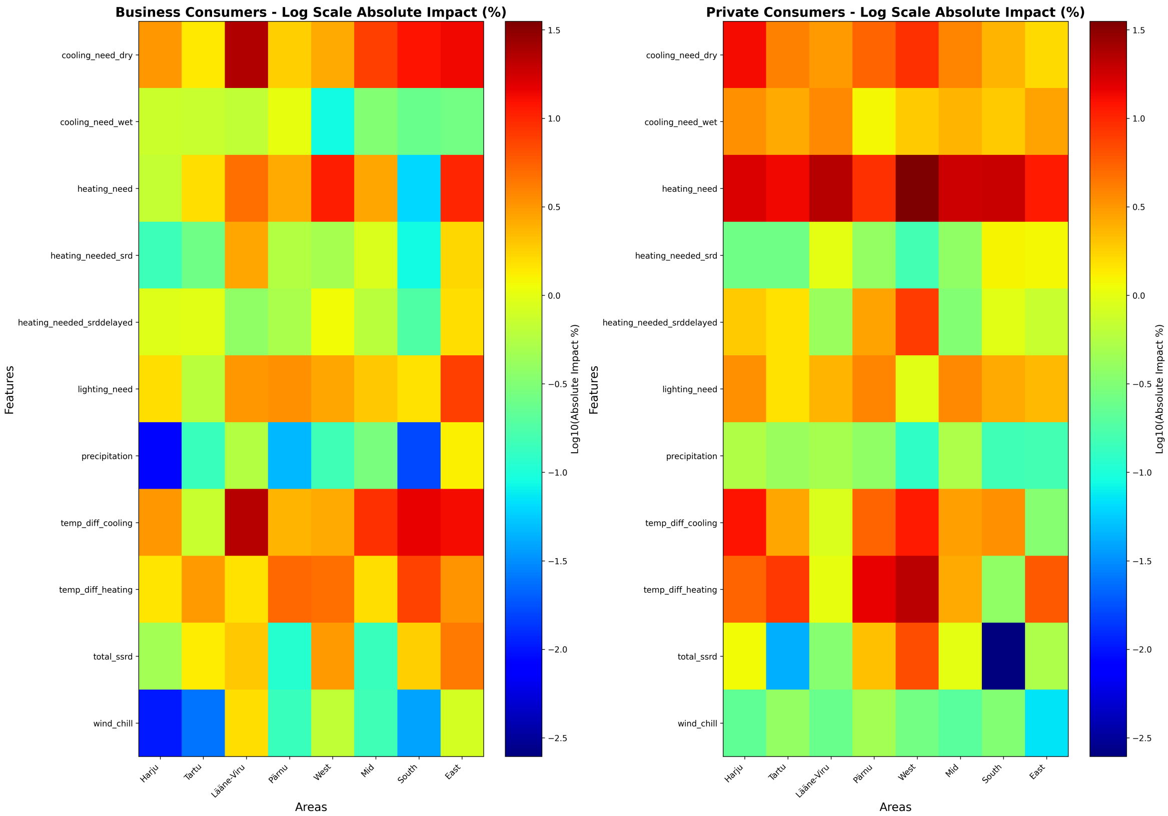

Figure 11. Log-scale percentage national impacts of weather variables for business and private consumers