Forecasting electrical power consumption using individual load profiles

— Project Report —
Case Studies II

Abdul Muqsit Farooqi

Project Supervisors:
Prof. Dr. Christine Müller
Dr. Mirko Jakubzik

September 2, 2025

TU Dortmund

Contents

1 Introduction

With the increasing demand for sustainable energy and the rising costs associated with power generation and storage, forecasting electrical power consumption has become a critical area of research. In particular, predicting consumption at the household level can support more efficient energy planning and distribution. Accurate household-level forecasts can reduce dependence on fossil fuels, optimize the integration of renewable energy sources, and mitigate unnecessary energy storage costs. This project focuses on forecasting electrical power consumption using individual load profiles of households equipped with heat pumps but without photovoltaic systems.

The dataset comprises 33 time series, each representing a different household. Each data file includes the total power consumption by the heat pump (`pumpe_tot`), total household consumption (`haushalt_tot`), and ambient temperature (`temperature_total`), indexed by timestamp. The data was provided by the instructors of Fallstudien II at TU Dortmund.

The goals of this project are as follows:

- Generate smoothed daily power consumption profiles using both short-term (hourly) and longer-term (weekday-based) smoothing methods.
- Investigate the influence of household, weekday, and season on power consumption and assess model simplification possibilities, including interaction effects.
- Cluster households based on their estimated consumption effects and classify unknown households into these clusters.
- Forecast one hour ahead power consumption using both general and specific cluster load profiles with kalman and particle filters.
- Compare forecasting accuracy and smoothing techniques with and without personalized load profiles.

The project proceeds in multiple analytical stages. First, a multivariate analysis of variance (MANOVA) is used to explore the effects of weekday and season, with and without interaction terms, followed by AIC-based model comparisons. Smoothed consumption profiles are derived and used in cluster analyses to group households. Three households are then classified into the resulting clusters, and their consumption is forecast using kalman and particle filters. Forecast accuracy is assessed using the root mean square error (RMSE).

In section 2, the dataset is explained briefly the quality of the dataset and to structure preprocessing of data are discussed separately for each task. In section 3, statistical methods are explained that are used in fulfilling the task of fitting the model, cluster analysis, power consumption of each household, and forecasting electrical power consumption using load profiles of households with heat pumps. In section 4, graphical plots such as QQ-plot, line plot, scatter plots, and models, different tests and filters are used to interpret the results. Lastly, all the results are summarized in section 5.

2 Description of the Problem

Understanding the dynamics of residential electricity consumption is essential for efficient energy planning and integrating renewable energy into the grid. This project investigates electricity consumption behavior across 33 households, each equipped with a heat pump but without a photovoltaic (PV) system. The primary aim is to model, analyze, and forecast short-term power usage at the household level using individual load profiles.

The project focuses on evaluating the effects of behavioral patterns (weekday, seasonality) and specific household consumption on daily load curves. In particular, it explores the use of smoothing techniques, multivariate statistical models, clustering, and state-space forecasting methods such as kalman and particle filters to improve forecasting performance.

2.1 Objectives of the Project

The statistical and analytical objectives of the project are:

- Develop smoothed daily electricity consumption profiles using hourly and weekday-based smoothing techniques.
- Examine the effects of household identity, weekday, and seasonality on power usage patterns.
- Evaluate whether simpler models (e.g., without interaction terms) can adequately explain consumption variation.
- Cluster similar households based on estimated effects and classify new ones accordingly.
- Predict next hour electricity consumption using both general and specific cluster load profiles via kalman and particle filters.
- Compare forecasting errors across various smoothing and modeling strategies.

2.2 The Data Material

The dataset consists of 33 observational time series, each corresponding to one household over a calendar year (2019). Each household's data is stored in a separate file and includes the following variables:

- `index`—Timestamp of the observation (date and hour),
- `pumpe_tot`—Power consumption of the heat pump (in watts),
- `haushalt_tot`—Total household electricity consumption (in watts),
- `temperature total`—Ambient temperature at the time of measurement (in degrees Celsius).

Only the variable `haushalt_tot` is used in the analysis. Each day's power consumption is considered as a 24-dimensional vector, forming the basis for multivariate modeling.

The dataset is comprehensive and regularly recorded at an hourly resolution, covering an entire year. However, several issues were addressed in preprocessing:

- **Missing values:** Some records lacked complete 24-hour entries and were removed or interpolated depending on context.
- **Extreme outliers:** Occasional measurement spikes or flatlined segments were identified and smoothed.
- **Alignment:** Timestamp alignment across households ensured consistency in daily and weekly comparisons.

Overall, the data quality is high, making it well-suited for time series analysis and statistical modeling. After preprocessing, the cleaned dataset supports the project's objectives of evaluating and forecasting household electricity demand in a structured and interpretable way.

3 Statistical Methods

In this section, various statistical methods are discussed for analyzing the data. For the calculation and graphical representation of statistical measures, R software version 4.2.3 (R Core Team 2023) is used with the packages `ggplot2` for data visualization (Wickham 2016), `dplyr` (Wickham 2022) for data manipulation and transformation, `clusterSim` (Walesiak and Dudek 2020) for clustering, and similarity-based ordering, `clValid` (Brock et al. 2008) for Cluster validation, `lubridate` (Grolemund and Wickham 2011) for simplifying date-time manipulation, `factoextra` (Kassambara 2017) for visualize PCA results, and `tidyverse` (Wickham et al. 2019) for data science and data wrangling.

3.1 Moving Average

The moving average is a core and widely utilized nonparametric method in time series analysis, primarily aimed at trend estimation and smoothing. Instead of constructing a parametric model for the underlying process, it focuses on capturing long-term trends and mitigating short-term variations in the data. This approach is particularly valuable for detecting overall patterns without assuming a specific model structure.

A commonly used form is the two-sided (or centered) moving average, defined for a time series $\{X_i\}$ as (Brockwell and Davis 2016, p. 21):

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j}$$

where $2q + 1$ is the window size and q is a nonnegative integer.

This method computes the average of the $2q+1$ data points centered at time t. It assumes that the trend component m_t is nearly linear within the interval $[t - q, t + q]$ and that the mean of the error terms in this window is approximately zero.

Under these assumptions, the moving average provides an estimate of the local trend (Brockwell and Davis 2016, p. 22):

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j}, \quad q+1 \leq t \leq n-q.$$

This smoothing method is especially useful when the signal-to-noise ratio is low and the underlying trend is fairly smooth. At the edges of the time series, where $t \leq q$ or $t > n - q$, the standard moving average cannot be directly applied due to a lack of sufficient neighboring observations.

Moving averages can be adapted into one-sided filters for forecasting or modified to prioritize recent observations, as in weighted and exponential moving averages. Although not designed for formal statistical inference, they serve as valuable exploratory tools facilitating pattern recognition, aiding in deseasonalization, and preparing datasets for more complex modeling approaches (Brockwell and Davis 2016).

3.2 Multivariate Analysis of Variance (MANOVA)

Two-way Multivariate Analysis of Variance (MANOVA) extends the one-way MANOVA by incorporating two categorical independent variables (factors) and their interaction. It facilitates simultaneous hypothesis testing across multiple dependent variables, accounting for both main and interaction effects.

The two-way MANOVA model can be expressed as (Rencher 2002, p. 186):

$$\mathbf{Y}_{ijk} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{\varepsilon}_{ijk}$$

where (Rencher 2002, p. 186):

- $\mathbf{Y}_{ijk}$ is the $p \times 1$ vector of responses for the k-th observation in the i-th level of factor A and j-th level of factor B,
- $\boldsymbol{\mu}$ is the overall mean vector $(p \times 1)$,
- $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_j$ are the main effect vectors for factor A and factor B, respectively,
- $\boldsymbol{\gamma}_{ij}$ is the interaction effect vector for level i of factor A and level j of factor B,
- $\boldsymbol{\varepsilon}_{ijk} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ is the random error vector.

Hypothesis Testing in Two-Way MANOVA The null hypotheses for the main and interaction effects are (Rencher 2002, p. 188):

$$\begin{aligned} H_0^A &: \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \cdots = \boldsymbol{\alpha}_a \\ H_0^B &: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_b \\ H_0^{AB} &: \boldsymbol{\gamma}_{ij} = \mathbf{0}, \quad \forall i, j \end{aligned}$$

These are tested using statistics based on the eigenvalues λ_i of $\mathbf{E}^{-1}\mathbf{H}$, where $\mathbf{H}$ and $\mathbf{E}$ are the hypothesis and error sum of squares and cross products matrices.

Test statistic:

- **Wilks' Lambda (Λ)** (Rencher 2002, p. 161):

$$\Lambda = \prod_{i=1}^{s} \frac{1}{1+\lambda_i} = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|}$$

Assumptions (Rencher 2002, p. 198)

- **Multivariate Normality**: Within each cell (i.e., each combination of factor levels), the response vectors follow a multivariate normal distribution.
- **Homogeneity of Covariance Matrices**: The covariance matrices are equal across all cells:

$$\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Sigma}, \quad \forall i, j$$

- **Independence of Observations**: Observations are independently and randomly sampled within each cell.

Properties and Interpretation (Rencher 2002)

- Two-way MANOVA detects both main and interaction effects across multiple response variables.
- It reduces Type I error inflation by evaluating all dependent variables simultaneously.
- If significant effects are found, follow-up analyses may include univariate ANOVAs or discriminant analyses to identify the contributing variables.

Two-way MANOVA is a robust extension of the MANOVA framework suitable for experiments involving two categorical factors and multiple interrelated dependent variables. It allows for the assessment of the individual and combined impacts of the factors, offering a comprehensive multivariate perspective on group differences (Rencher 2002).

3.3 Shapiro-Wilk Test

The Shapiro-Wilk goodness of fit test is used to determine if a random sample, Y_i for $i = 1, 2, \ldots, n$, drawn from a normal Gaussian probability distribution with true mean and variance, μ_i and σ^2, respectively (Shapiro and Wilk 1965, p. 592). That is $Y \sim N(\mu_i, \sigma^2)$.

Thus, we test the following hypothesis:

$$H_0 : \text{The random sample was drawn from a normal population, } N(\mu_i, \sigma^2)$$
$$\textbf{Vs}$$
$$H_0 : \text{The random sample does not follow } N(\mu_i, \sigma^2)$$

The Shapiro-Wilk test statistic is used to test the hypothesis, which is given by (Shapiro and Wilk 1965, p. 592):

$$W = \frac{(\sum_{i=1}^{n} a_i y_{(i)})^2}{\sum_{i=1}^{n} (y_i - \bar{y})}$$

where $y_{(i)}$ are the ordered sample values and a_i are constants that are generated by the expression (Shapiro and Wilk 1965, p. 593),

$$(a_1, a_2, \ldots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

where (Shapiro and Wilk 1965, p. 592)

$$(m = m_1, m_2, ..., m_n)^T$$

represents the expected values of the order statistics derived from independent and identically distributed standard normal random variables, $N(0,1)$, and V denotes the covariance matrix of these order statistics (Shapiro and Wilk 1965).

3.4 Akaike Information Criterion

The Akaike Information Criterion (AIC) is a statistical measure used to compare and select models by balancing goodness of fit with model complexity. It penalizes models with more parameters to avoid overfitting, favoring the model with the lowest AIC value as the best choice. The AIC is computed using the following formula (Heiberger and Holland 2015, p. 639):

$$\text{AIC} = -2ln(\hat{L}) + 2m$$

where m denotes the number of parameters estimated in the model and $\hat{L}$ represents the maximum log-likelihood value, which serves as an indicator of the model's goodness of fit (Heiberger and Holland 2015).

3.5 Clustering Method

Clustering is a central technique in unsupervised learning that groups a set of objects into clusters based on their similarities or differences. Among the most popular methods is the k-means algorithm, which partitions data into K distinct, non-overlapping clusters by minimizing the total within-cluster variability. Each cluster is represented by a centroid, and each object is assigned to the cluster with the closest centroid, typically measured using the Euclidean distance. Mathematically, for a set of points $\Theta_H \in \mathbb{R}^p$, k-means minimizes the total within-cluster sum of squares:

$$\sum_{k=1}^{K} \sum_{H \in C_k} \|\Theta_H - \Theta_k\|^2 .$$

Here, Θ_k denotes the centroid of cluster C_k. This approach is especially effective when the clusters are roughly spherical and of similar size.

3.5.1 Silhouette Method for Cluster Validation

The silhouette method, introduced by (Rousseeuw 1987), provides a graphical and quantitative means of assessing the quality of clustering results. It evaluates how similar each object is to its own cluster compared to other clusters, allowing for the detection of well-clustered, borderline, or misclassified points. The silhouette score $s(i)$ for an object i is computed as follows (Rousseeuw 1987, p. 55):

- Let $a(i)$ be the average dissimilarity of object i to all other points within its own cluster A.

- For every other cluster $C \neq A$, compute the average dissimilarity $d(i, C)$ of i to all objects in C.
- Let $b(i) = \min_{C \neq A} d(i, C)$, representing the lowest average dissimilarity to any neighboring cluster.

Then, the silhouette value for object i is defined by (Rousseeuw 1987, p. 56):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

This score lies between -1 and 1 (Rousseeuw 1987, p. 56):

- $s(i) \approx 1$ indicates that the object is matched well to its own cluster and poorly matched to neighboring clusters.
- $s(i) \approx 0$ suggests that the object lies near a cluster boundary.
- $s(i) < 0$ implies potential misclassification.

Plotting the silhouette scores of all data points, arranged by their assigned clusters, allows for a visual assessment of how well the clusters are formed in terms of cohesion and separation. The average silhouette width across all observations provides a summary measure of overall clustering quality. Evaluating this average for different values of k helps determine the optimal number of clusters. High and consistently wide silhouette values suggest clearly defined and well-separated natural groupings within the data.

Silhouette values are independent of the clustering algorithm itself and depend only on the final group assignments and the selected distance measure. Therefore, they provide a robust and versatile approach for assessing and interpreting cluster validity (Rousseeuw 1987).

3.6 The Kalman Filter

The kalman filter is a recursive algorithm used to estimate the state of a linear dynamic system from a series of noisy observations. It is commonly applied in time series analysis, control theory, and econometrics for real-time prediction, smoothing, and filtering.

A general state-space model consists of:

- **State equation** (Kalman 1960, p. 262):

$$\mathbf{X}_{t+1} = \mathbf{F}\mathbf{X}_t + \mathbf{V}_t$$

- **Observation equation** (Kalman 1960, p. 262):

$$\mathbf{Y}_t = \mathbf{G}\mathbf{X}_t + \mathbf{W}_t$$

Where $\mathbf{X}_t$ is the unobserved state vector at time t, $\mathbf{Y}_t$ is the observed vector at time t, $\mathbf{F}$ is the state transition matrix, and $\mathbf{G}$ is the observation matrix. The observation noise $\mathbf{W}_t$ is assumed to be distributed as $\mathcal{N}(\mathbf{0}, \mathbf{R})$, and the process noise $\mathbf{V}_t$ is independently distributed as $\mathcal{N}(\mathbf{0}, \mathbf{Q})$.

Kalman Prediction Equations (Kalman 1960, p. 271)

$$\hat{X}_{t+1} = F_t \hat{X}_t + \Theta_t \Delta_t^{-1} (Y_t - G_t \hat{X}_t)$$
$$\Omega_{t+1} = F_t \Omega_t F_t' + Q_t - \Theta_t \Delta_t^{-1} \Theta_t'$$

where $\hat{X}_t$ denotes the estimate of the state at time t, and $\hat{X}_{t+1}$ is the prediction of the state at time t. The matrix F_t is the state transition matrix, and G_t is the observation matrix. The observed measurement at time t is denoted by Y_t. The matrix Ω_t represents the covariance of the prediction error at time t, Ω_t is the predicted error covariance at time $t+1$, and Q_t is the covariance matrix of the process noise V_t in the state equation. Θ_t, Δ_t are intermediate matrices used to compute the Kalman gain.

Kalman Filter Update Equations (Kalman 1960, p. 274)

$$P_t X_t = P_{t-1} X_t + \Omega_t G_t' \Delta_t^{-1} (Y_t - G_t \hat{X}_t)$$
$$\Omega_{t|t} = \Omega_t - \Omega_t G_t' \Delta_t^{-1} G_t \Omega_t'$$

where $P_t X_t$ denotes the updated (posterior) expectation of the state vector X_t given all observations up to time t, i.e., $\mathbb{E}[X_t \mid Y_1, \ldots, Y_t]$. The quantity $P_{t-1} X_t$ represents the prior (predicted) expectation of X_t based on observations up to time $t-1$, before incorporating Y_t. The symbol $\hat{X}_t$ is an alternate notation for $P_{t-1} X_t$, representing the predicted state prior to the update step. The vector Y_t is the observation at time t, and G_t is the observation matrix that maps the latent state X_t to the observation Y_t. The matrix Ω_t represents the prediction error covariance before incorporating Y_t. The updated (posterior) error covariance matrix is denoted by $\Omega_{t|t}$.

Properties (Kalman 1960)

- Provides the Best Linear Unbiased Estimate (BLUE) under Gaussian assumptions.
- Posterior distribution of $\mathbf{X}_t$ remains Gaussian.
- Optimal in the Minimum Mean Square Error (MMSE) sense.

Assumptions and Justification (Kalman 1960)

- **Linearity:** Both state and observation equations must be linear.
- **Gaussian Noise:** $\mathbf{W}_t$ and $\mathbf{V}_t$ are zero-mean white noise with known covariances $\mathbf{Q}_t$ and $\mathbf{R}_t$.
- **Initial Conditions:** Require an initial estimate $\hat{\mathbf{X}}_{0|0}$ and covariance matrix $\mathbf{P}_{0|0}$.

The kalman filter is a powerful technique for dynamic linear modeling in time series. Its recursive nature and optimal properties under Gaussian assumptions make it suitable for real-time applications in diverse fields, including engineering, economics, and environmental sciences (Kalman 1960).

3.7 Particle Filter

The particle filter, or Sequential Monte Carlo (SMC) method, is a recursive simulation-based technique for estimating the states of dynamic systems, particularly when the model is non-linear and/or the noise is non-Gaussian. It approximates the filtering distribution (Cappé et al. 2007),

$$\pi_t(x_t \mid y_{0:t}),$$

which is the posterior distribution of the latent state x_t given observations $y_{0:t}$, using a set of particles $\{x_t^{(i)}\}_{i=1}^N$ and associated weights $\{\omega_t^{(i)}\}_{i=1}^N$ (Cappé et al. 2007, pp. 3, 5).

Algorithm Steps The basic particle filtering procedure is as follows:

1. **Initialization (for $t = 0$):** (Cappé et al. 2007, p. 6)

$$\tilde{x}_0^{(i)} \sim q_0(x_0 \mid y_0), \quad \tilde{\omega}_0^{(i)} = \frac{g(y_0 \mid \tilde{x}_0^{(i)})\pi_0(\tilde{x}_0^{(i)})}{q_0(\tilde{x}_0^{(i)} \mid y_0)}$$

2. **Recursive update (for $t = 1, \ldots, T$):** (Cappé et al. 2007, p. 6)

   - *Resampling (optional):* Draw ancestor indices j_i from $\{1, \ldots, N\}$ based on weights $\omega_{t-1}^{(j)}$, and set $x_{t-1}^{(i)} = \tilde{x}_{t-1}^{(j_i)}$.
   - *Propagation:*
   $$\tilde{x}_t^{(i)} \sim q_t(x_t \mid x_{t-1}^{(i)}, y_t)$$
   - *Weight update:*
   $$\tilde{\omega}_t^{(i)} = \omega_{t-1}^{(i)} \cdot \frac{g(y_t \mid \tilde{x}_t^{(i)}) f(\tilde{x}_t^{(i)} \mid x_{t-1}^{(i)})}{q_t(\tilde{x}_t^{(i)} \mid x_{t-1}^{(i)}, y_t)}$$
   - *Normalization:*
   $$\omega_t^{(i)} = \frac{\tilde{\omega}_t^{(i)}}{\sum_{j=1}^N \tilde{\omega}_t^{(j)}}$$

The filtered estimate of a function $h(x_t)$ is (Cappé et al. 2007, p. 5):

$$\hat{h}_t = \sum_{i=1}^{N} \omega_t^{(i)} h_t(\tilde{x}_t{}^{(i)})$$

They can handle complex distributions but may become computationally expensive in high-dimensional settings, requiring careful tuning of the number of particles and the proposal distribution (Cappé et al. 2007).

4 Statistical Analysis

In this section, all statistical methods are used to make a meaningful analysis of the dataset.

4.1 Data Pre-Processing via Smoothing

There are missing values in the dataset, and to have better quality of the data, the observations were then smoothed by not taking missing values into account, while smoothing rows that contained NA values were then ignored when applying the smoothing function. The sample size is big enough that help with not having biased results, but the variation can be seen in the sample data. The problem with a big sample size is that we need to improve the quality of the data by smoothing data that have no NA values, and if the data have too much variation, it can also lead to misleading or bad results. In the project, data is firstly changed to hours from minutes by taking the average of the minutes of the same hour, then there are two parts of smoothing: A (smoothing with adjacent hours) and B (smoothing with adjacent weekdays).

Table 1: Summary Statistics for Selected Households (Original and Smoothed Data)

Household (Type)	CV	Min	Max	Mean	Median
Household 3 (Original)	0.8124	74.72	2312.67	232.58	163.99
Household 3 (Smoothed)	0.5837	86.59	1690.67	231.62	181.03
Household 23 (Original)	0.4899	135.07	2331.99	356.25	292.18
Household 23 (Smoothed)	0.3347	135.07	1551.14	357.08	329.54
Household 40 (Original)	0.6740	138.77	2318.08	465.62	344.88
Household 40 (Smoothed)	0.5400	138.77	2318.08	463.98	402.75

For the above summary, smoothing the household power consumption data reduces variability by lowering the coefficient of variation and softening extreme values. While the mean remains largely unchanged, the median often increases slightly, reflecting a more balanced distribution. Overall, smoothing helps highlight underlying usage patterns by reducing noise and outlier impact, making the data more stable and interpretable.

4.2 Data Pre-Processing Cluster Load Profiles

A representative cluster-level load profile was computed using the function for each day of the year and each hour. This function aggregates smoothed hourly estimates for all households within the same clusters. The profile is computed as the average for each cluster k, hour h, and day d:

$$\widehat{L}_{k,h}(d) = \frac{1}{|S_k|} \sum_{i \in S_k} \left(\beta_{0,h}^{(i)} + \beta_{1,h}^{(i)} \cdot \sin\left(\frac{2\pi(d-\delta)}{365} \right) \right), \; d \in \{1, \ldots, 365\} \; \textit{is the calendar day}$$

Where S_k is the set of households assigned to cluster k, $\beta_{0,h}^{(i)}$ and $\beta_{1,h}^{(i)}$ are the coefficients from the seasonal model for the household i at hour h, and δ is a calendar offset used to align seasonal trends (specifically, $\delta = 31 + 28 + 21 = 80$).

Load profiles are computed separately for weekend and non-weekend. Function ignores days that do not belong to the specified type by setting their profile values to NA.

4.3 Data Pre-Processing for Filters

The function checks for the NA values in each column of days; if the whole column has NA values, then it ignores it and moves on to the next day, but if there are some NA values in a column, then it takes one value from the previous and one value from the next and then takes the mean to fill the NA values.

4.4 Global Test using MANOVA Method

4.4.1 MANOVA Results: Model with Interaction Terms

To explore both main and interaction effects of *Household*, *Weekday*, and *Season*, a MANOVA model was fit using Wilks' Lambda. The results are presented in Table 2.

Table 2: Results of MANOVA with Interaction Terms

Source	DF	Wilks Λ	approx F	num DF	den DF	P-Value
Household	29	0.08639	36.380	696	179132	$< 2.2 \times 10^{-16}$
Weekday	6	0.86263	10.046	144	56492	$< 2.2 \times 10^{-16}$
Season	1	0.85693	67.234	24	9665	$< 2.2 \times 10^{-16}$
Household:Weekday	174	0.35686	2.440	4176	230034	$< 2.2 \times 10^{-16}$
Household:Season	29	0.44857	11.384	696	179132	$< 2.2 \times 10^{-16}$
Weekday:Season	6	0.96356	2.500	144	56492	$< 2.2 \times 10^{-16}$
Household:Weekday:Season	174	0.53762	1.457	4176	230034	$< 2.2 \times 10^{-16}$

As shown in Table 2, all main effects and interaction terms are statistically significant at the 0.05 level. The significant three-way interaction implies that the influence of one factor depends on the levels of the others, indicating complex multivariate relationships among *Household*, *Weekday*, and *Season*.

4.4.2 MANOVA Results: Model without Interaction Terms

To assess only the main effects, a separate MANOVA was conducted, excluding all interaction terms. The results are summarized in Table 3.

Table 3: Results of MANOVA without Interaction Terms

Source	DF	Wilks Λ	approx F	num DF	den DF	P-Value
Household	29	0.11682	32.886	696	186226	$< 2.2 \times 10^{-16}$
Weekday	6	0.88489	8.624	144	58731	$< 2.2 \times 10^{-16}$
Season	1	0.87125	61.871	24	10048	$< 2.2 \times 10^{-16}$

Table 3 confirms that the main effects of *Household*, *Weekday*, and *Season* are all statistically significant, even without considering interaction terms. However, when compared to the interaction model, the non-interaction model may fail to capture more nuanced relationships between the predictors.

The MANOVA analyses demonstrate that *Household*, *Weekday*, and *Season* significantly influence the multivariate outcomes. The interaction model provides a more detailed understanding by capturing complex interdependencies between factors, whereas the non-interaction model offers a simpler yet statistically robust explanation of main effects.

4.5 Model Selection Based on Interaction Effects

When comparing models, it is important to balance model complexity with explanatory power. Including interaction terms may offer a more accurate representation of relationships among predictors, but at the cost of increased complexity and risk of overfitting. To evaluate model performance, the Akaike Information Criterion (AIC) is employed. A lower AIC value indicates a more parsimonious model with better goodness of fit relative to its complexity.

Figure 1 displays the AIC values for models with and without interaction terms across multiple dependent variables. Table 4 summarizes the number of cases in which each model type was preferred based on AIC differences.

From the visualization, it can be observed that the AIC values for models including interaction terms (particularly model B) are generally lower than those for their non-interaction counterparts. This suggests that the interaction models better explain the variation in the dependent variables. Light color shows lower AIC values.

Table 4: Model choice based on difference in the AIC values

sum(aic inter A - aic nointer A > 0)	sum(aic inter B - aic nointer B < 0)
15 models where no interaction is better	29 models where interaction is better

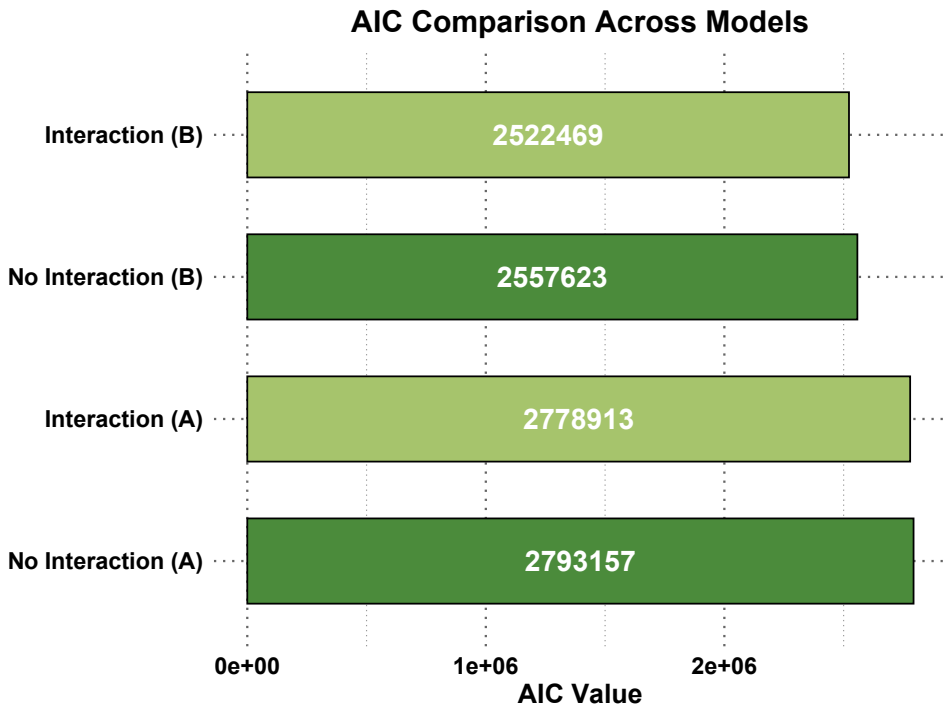

Figure 1: AIC values Interaction & Non-Interaction

Based on these findings, interaction models are favored in the majority of cases. Although simpler models without interactions are preferred in some situations (15 out of 30), the inclusion of interaction terms improves model fit more frequently (29 out of 30). Therefore, while considering interpretability and complexity, the interaction model is often the more appropriate choice when aiming to capture nuanced effects between factors.

4.6 Cluster Validation and Evaluation

To determine the most suitable number of clusters and to evaluate the quality of the clustering results, both the **silhouette coefficient** and the **Dunn index** were computed. These metrics offer insight into the *cohesion* (intra-cluster similarity) and *separation* (inter-cluster dissimilarity) of clusters formed during weekdays and weekends across Parts A and B.

4.6.1 Optimal Number of Clusters (Silhouette Method)

The silhouette coefficient ranges between -1 and 1, with higher values indicating better-defined clusters. Figure 15 illustrates the average silhouette widths for cluster counts from 1 to 10.

From the figures, the optimal number of clusters was:

- **2 clusters** for weekdays (Parts A and B).

- **3 clusters** for weekends (Parts A and B).

4.6.2 Cluster Visualization via PCA

To further understand the clustering patterns, the clusters were projected into principal component space. Figures 16, 17, 2, and 3 display the spatial separation between clusters for both trained and new households.

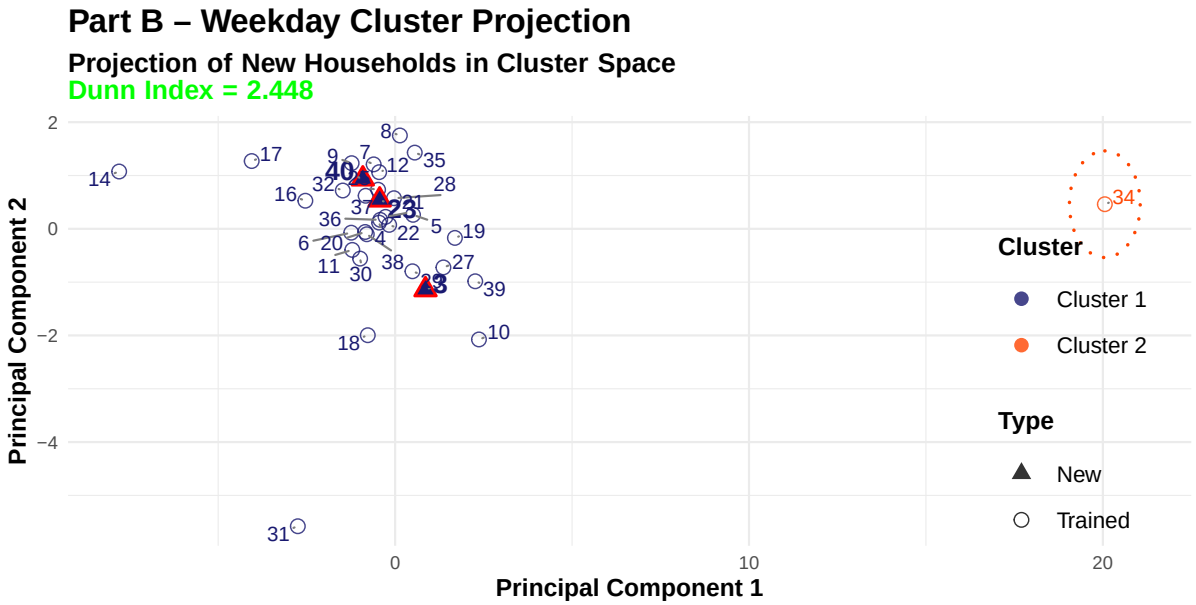

Figure 2: Part B – Weekday Cluster Projection

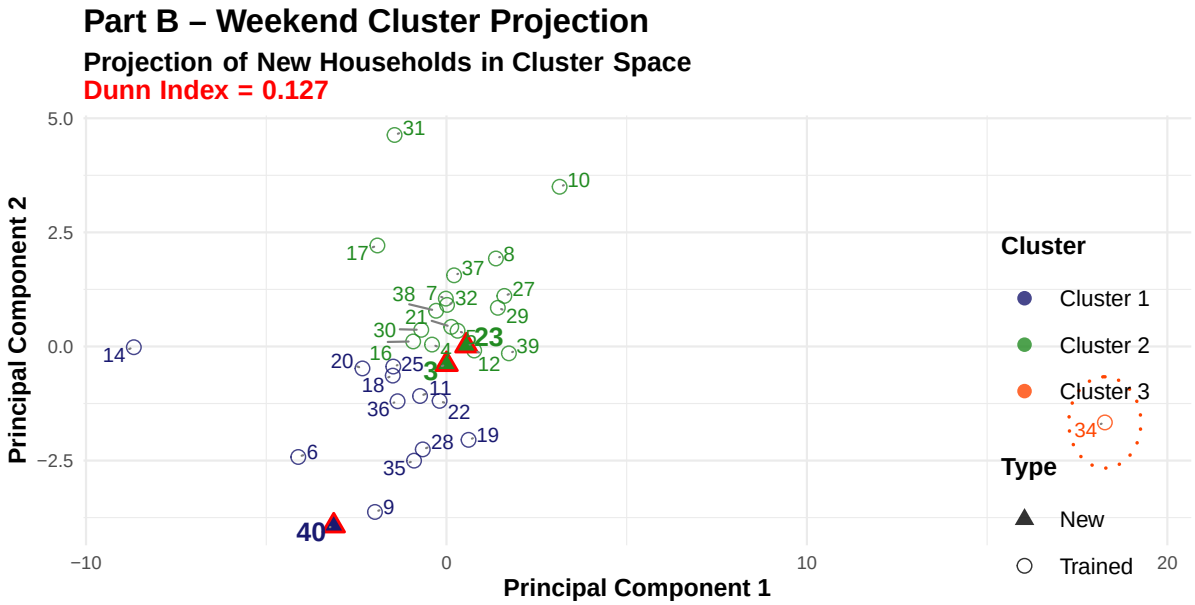

Figure 3: Part B – Weekend Cluster Projection

The PCA plots reaffirm the silhouette findings: weekday clusters are well-separated and compact, while weekend clusters are more scattered and overlapping.

In all the cluster projections for both weekend and weekday, household 34 behaves as an outlier, as there are many outliers in household 34. In Figure 14, outliers for all households are given; household 34 is symbolized as a "star."

4.6.3 Dunn Index Evaluation

The Dunn index measures the ratio of the smallest inter-cluster distance to the largest intra-cluster distance. A higher value indicates better clustering performance. In all Figures 16, 2, 17 and 3, computed Dunn indices for each case are highlighted as green where the Dunn Index is greater than 1 and red where the Dunn Index is lower than 1.

The highest Dunn index is observed for **Part B – Weekday** (2.476568), indicating very well-defined clusters. Both weekend scenarios exhibit lower values (below 1), suggesting less distinct cluster separation due to increased variability in household behavior on weekends.

4.7 Load Profiling Analysis

To understand typical power consumption behaviors, smoothed load profiles were constructed using regression-based seasonal modeling. These profiles were clustered into representative groups (clusters) for weekends and weekdays separately. Each curve in the visualizations represents the mean consumption over 24 hours with a ±1 standard deviation (SD) range, offering insight into intra-cluster variability.

4.7.1 Weekend Consumption Patterns

Figure 18, 19, 4, 5, 20, and 21 display load profiles for the weekend across different clusters and households. The following patterns are observed:

- **Bimodal Shape:** Most households show two distinct peaks—one in the late morning and another in the early evening—suggesting characteristic daily activities like cooking or appliance use during these times.
- **High Variation in HH 40:** For Household 40 (e.g., Cluster 1 HH 40A Day 55), there is significant deviation between the smoothed household data (black dotted line) and the cluster-based load profile (solid line). These differences lead to wider confidence intervals and suggest outlier behavior within this cluster.
- **Negative Intervals:** In some profiles, particularly where variation is high, the ±1 SD interval dips below zero. While physically implausible, this is a statistical artifact of the modeling process and reflects extreme variance rather than actual negative consumption.
- **Good Model Fit in HH 3 and HH 23:** In contrast, households 3 and 23 (e.g., Cluster 2 HH 3A Day 34 and Cluster 2 HH 23A Day 48) show relatively good alignment with cluster-level predictions, indicating successful classification.

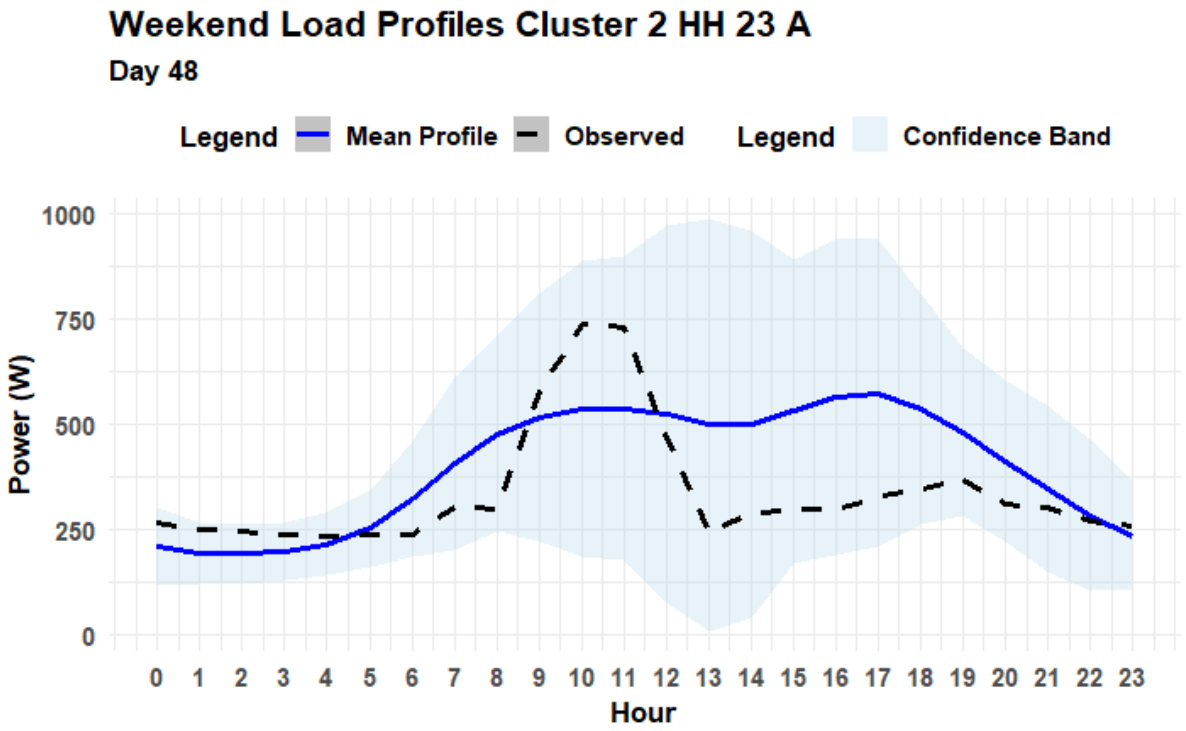

Figure 4: Weekend - Household 23 A Load Profile

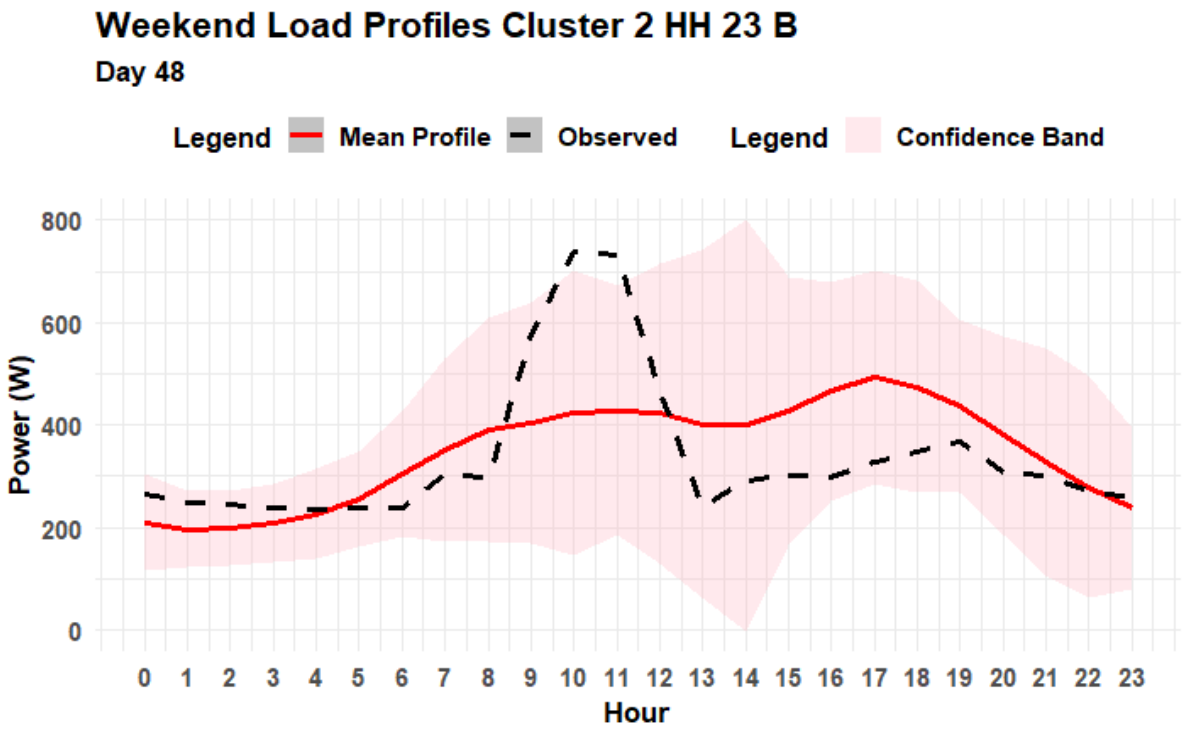

Figure 5: Weekend - Household 23 B Load Profile

4.7.2 Weekday Consumption Patterns

The weekday profiles (Figure 22, 23, 6, 7, 24, and 25) reflect different patterns compared to weekends:

- **Unified Behavior Across Households:** All households included are part of the same cluster, implying more homogeneous weekday behavior likely tied to standard workday routines.
- **Monotonic Load Increase:** Power consumption generally rises throughout the day, reaching a peak between 17:00 and 20:00, consistent with return from work activities.
- **Narrow Variance in Early and Late Hours:** The ± 1 SD interval is tightest during early morning (00:00–05:00) and late night hours (22:00–24:00), indicating more consistent low activity during these periods across users.
- **Model Accuracy:** Visual fits remain consistent, although slight deviations (e.g., Cluster 2 HH 40A Day 56) persist for higher variance households.

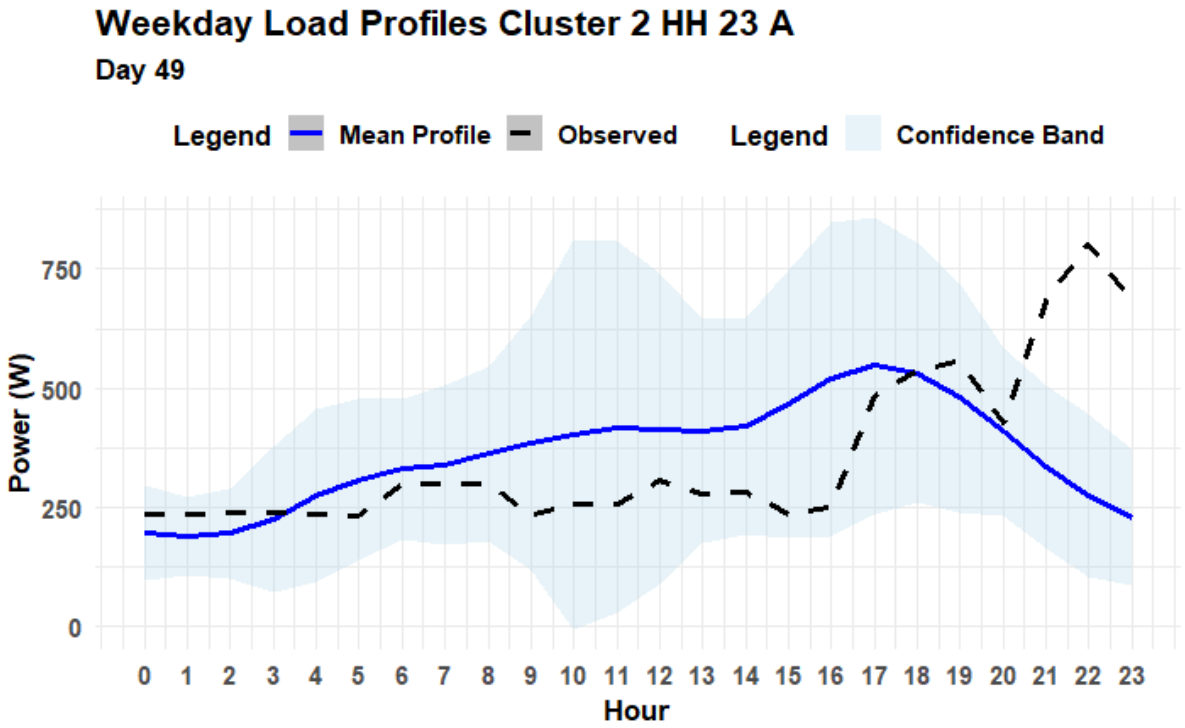

Figure 6: Weekday - Household 23 A Load Profile

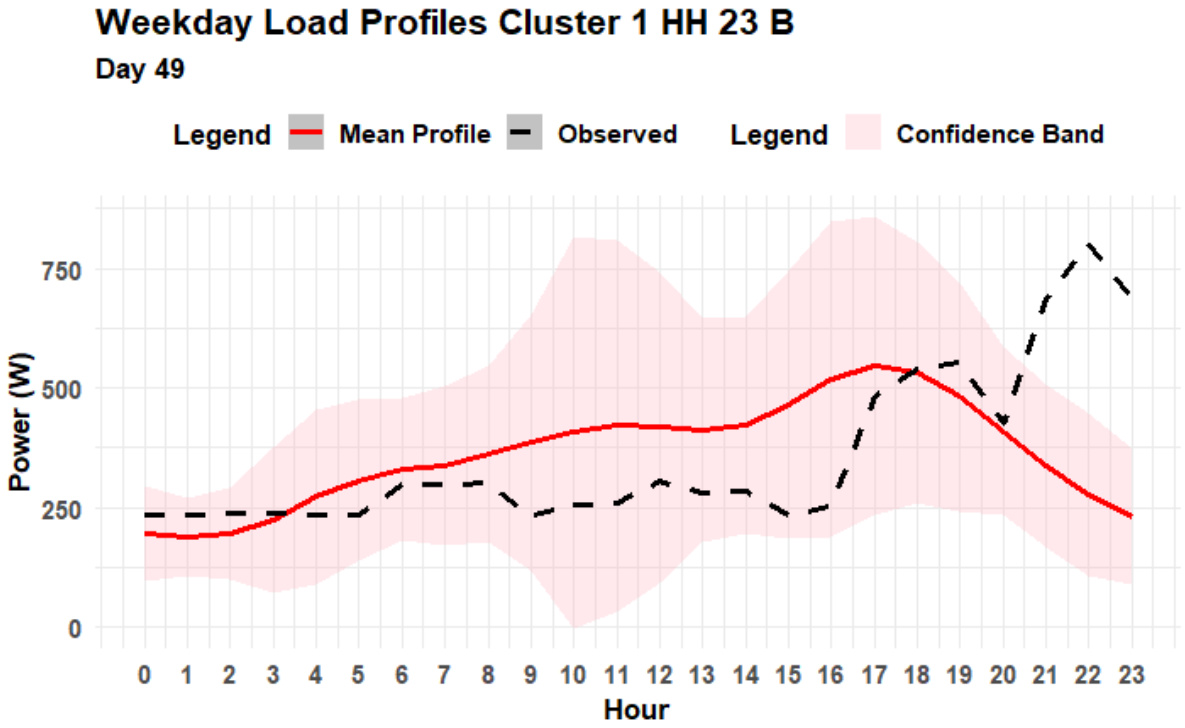

Figure 7: Weekday - Household 23 B Load Profile

4.8 Kalman Filter

4.8.1 Visual Comparison: Actual vs Predicted

Figures 26, 27, 8, 9, 28, and 29 illustrate the comparison between actual and predicted power consumption for Households 3, 23, and 40 respectively. Each figure shows both the load predictions (top) and the corresponding Q-Q plots of standardized residuals (bottom).

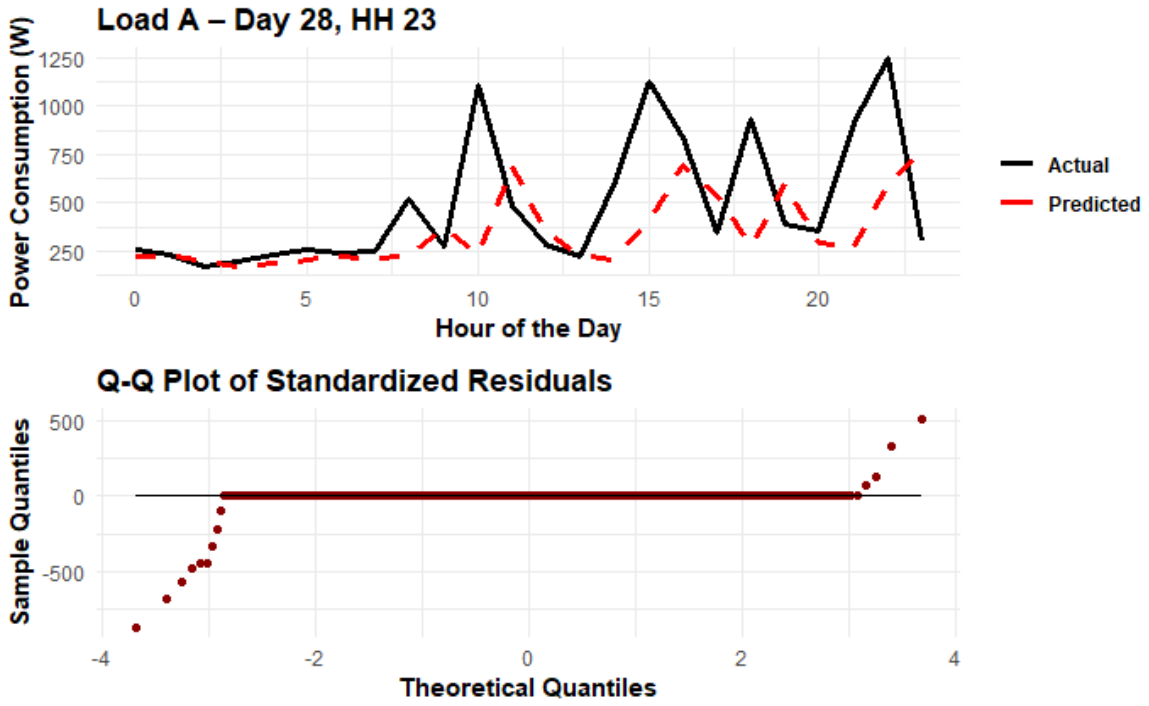

Figure 8: Kalman Filter Load A Predictions with Residuals - HH 23, Day 28

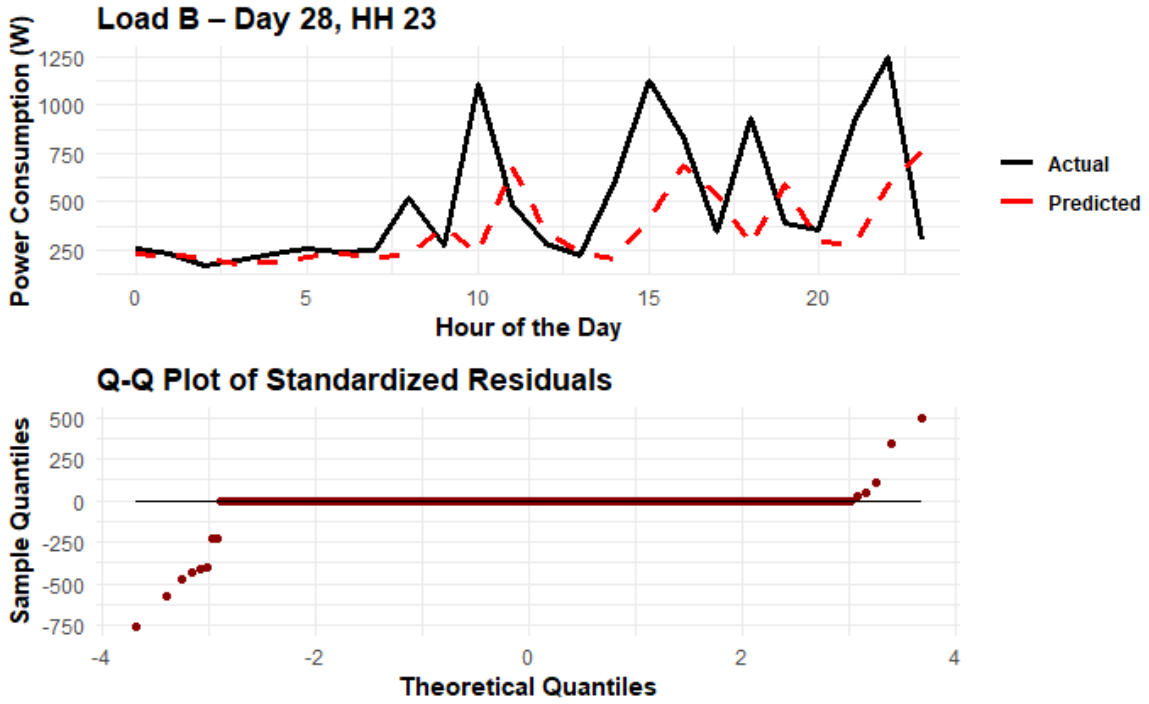

Figure 9: Kalman Filter Load B Predictions with Residuals - HH 23, Day 28

For household 23 in figure 8, 9, predictions are closely aligned with actual consumption, including peak timing and amplitude. Q-Q plots are flatter, suggesting residuals are more centralized and performance is statistically better.

For household 3 in figure 26, 27, load A and B predictions follow the general consumption pattern but underestimate high peaks. Q-Q plots show long tails, indicating non-normality and potential outliers in residuals.

For Household 40 in figure 28, 29, predictions are erratic and fail to capture consumption spikes accurately. Residual Q-Q plots show skewed and heavy-tailed distributions, reflecting data inconsistencies and missing values.

No Load Models

For no load models, flat red prediction lines across all households in Figure 30, 10, and 31 highlight poor model responsiveness. Q-Q plots show extreme non-normality and variance, confirming model inadequacy.

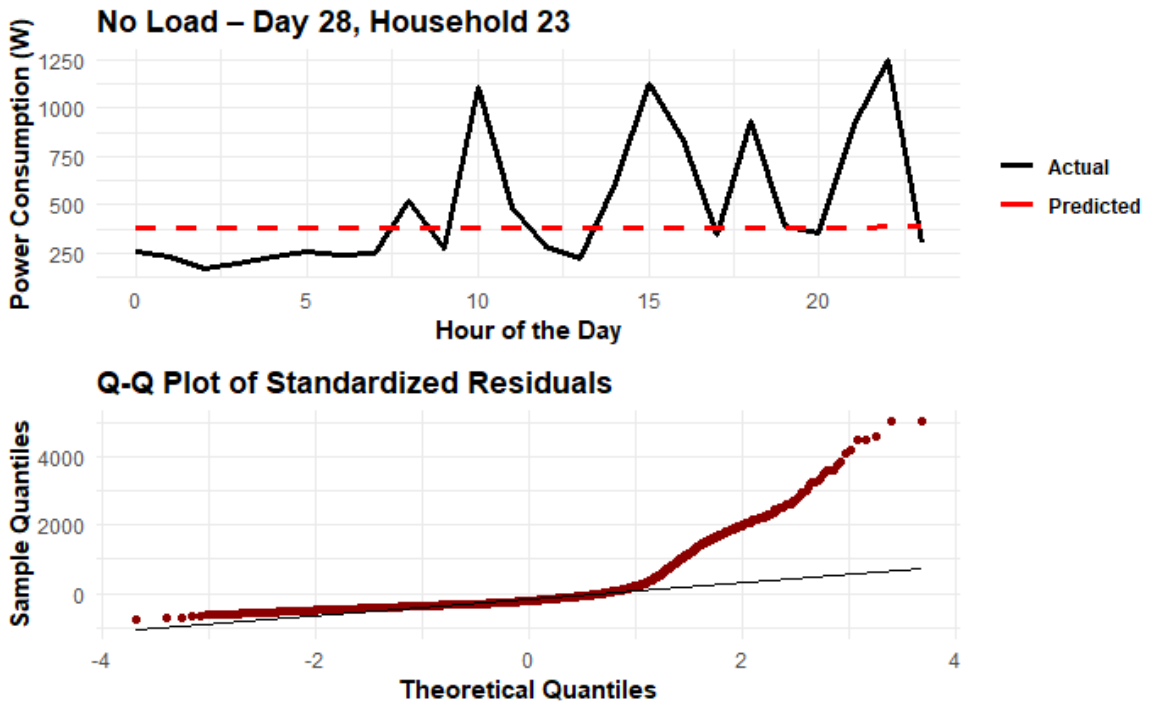

Figure 10: Prediction and Q-Q Residual Plot for No Load Model (HH 23)

Table 5: Summary of Ljung-Box (Independence) and Shapiro-Wilk (Normality) Tests

Household	Model	Ljung-Box p	Independence	Shapiro-Wilk p	Normality
3	Kalman A	2.22e-09	Rejected	4.97e-78	Rejected
	Kalman B	8.14e-07	Rejected	1.39e-76	Rejected
	No Load	0	Rejected	3.40e-77	Rejected
23	Kalman A	0.961	Accepted	1.23e-91	Rejected
	Kalman B	0.971	Accepted	1.23e-91	Rejected
	No Load	0	Rejected	2.45e-70	Rejected
40	Kalman A	0.0038	Rejected	2.62e-54	Rejected
	Kalman B	0.017	Rejected	1.28e-55	Rejected
	No Load	0	Rejected	3.67e-54	Rejected

Table 6: Kalman Filter MSE Comparison Across Households and Models

Household	Kalman A	Kalman B	No Load
3	53,578.61	**53,153.61**	65,026.09
23	53,237.97	**52,793.20**	56,172.84
40	110,721.60	**109,495.20**	151,963.00

4.8.2 Key Findings – Interpretation and Case Observations

- **Household 23**
  - Strongest statistical performance.
  - Residuals pass the independence test.
  - Accurate predictions visually and numerically.
- **Household 3**
  - Moderate prediction accuracy.
  - Residuals show correlation and deviate from normality.
- **Household 40**
  - Highest MSE, with weak prediction amplitude in Table 6.
  - Impacted by 109 days of missing data.
- **No Load Models**
  - Failed in all cases.
  - Do not capture consumption dynamics.
  - Residuals violate both independence and normality assumptions.

4.9 Particle Filter

4.9.1 Visual Comparison: Actual vs Predicted

Figures 32, 33, 11, 12, 34, and 35 present the actual versus predicted power consumption results using the Particle Filter for Households 3, 23, and 40 respectively. Each figure contains Load A and B predictions (top) and Q-Q plots of standardized residuals (bottom).

For household 3 in figure 32 and 33, load A and B predictions track actual consumption nearly perfectly. Residuals are very close to zero, shown in compressed Q-Q plots, though not normally distributed. Independence test is passed for kalman A, but rejected for kalman B.

For household 23 in figure 11 and 12, particle Filter yields excellent predictive alignment for both loads. Q-Q plots show low variability, indicating accurate predictions but also deviation from normality. Load B residuals pass the independence test, while Load A does not.

For household 40 in figure 34 and 35, the predictions follow the general structure, Load B underperforms. Residuals remain relatively small, yet Q-Q plots show skewed distributions. Independence holds, but normality fails significantly.

No Load Models

For no load in figure 36, 13, and 37, similar to kalman, predictions collapse to zero values. Model performance is drastically poor without load profiles.

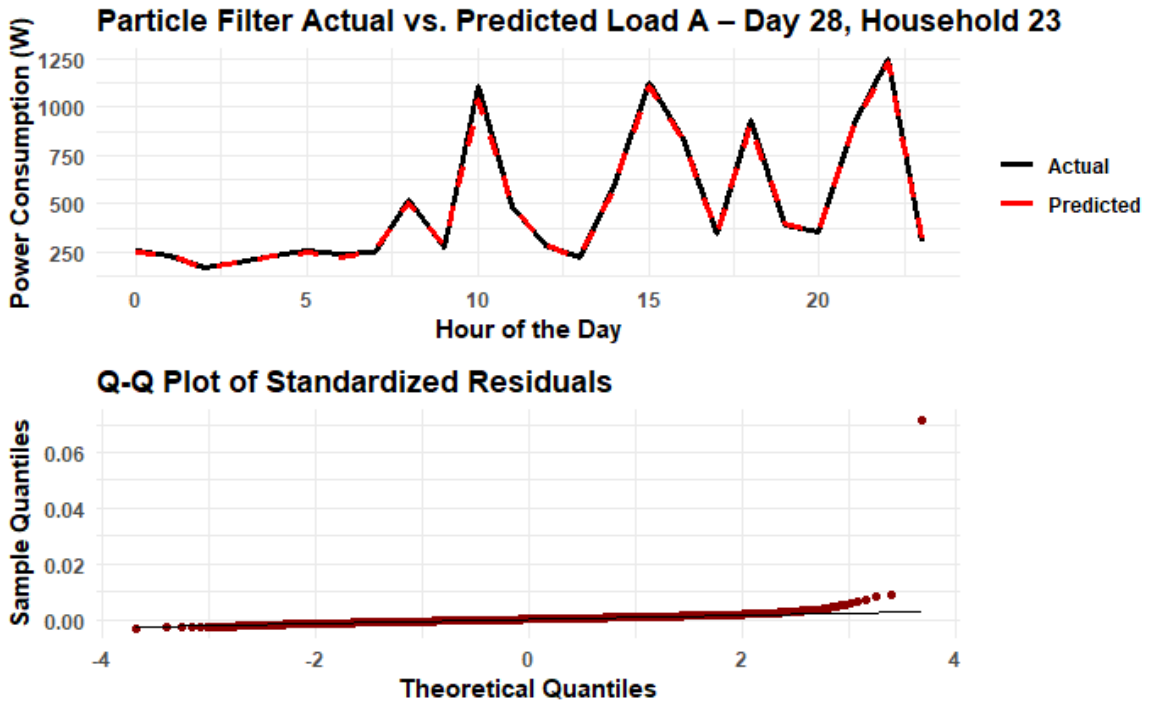

Figure 11: Particle Filter Load A Prediction with Residual - Household 23, Day 28

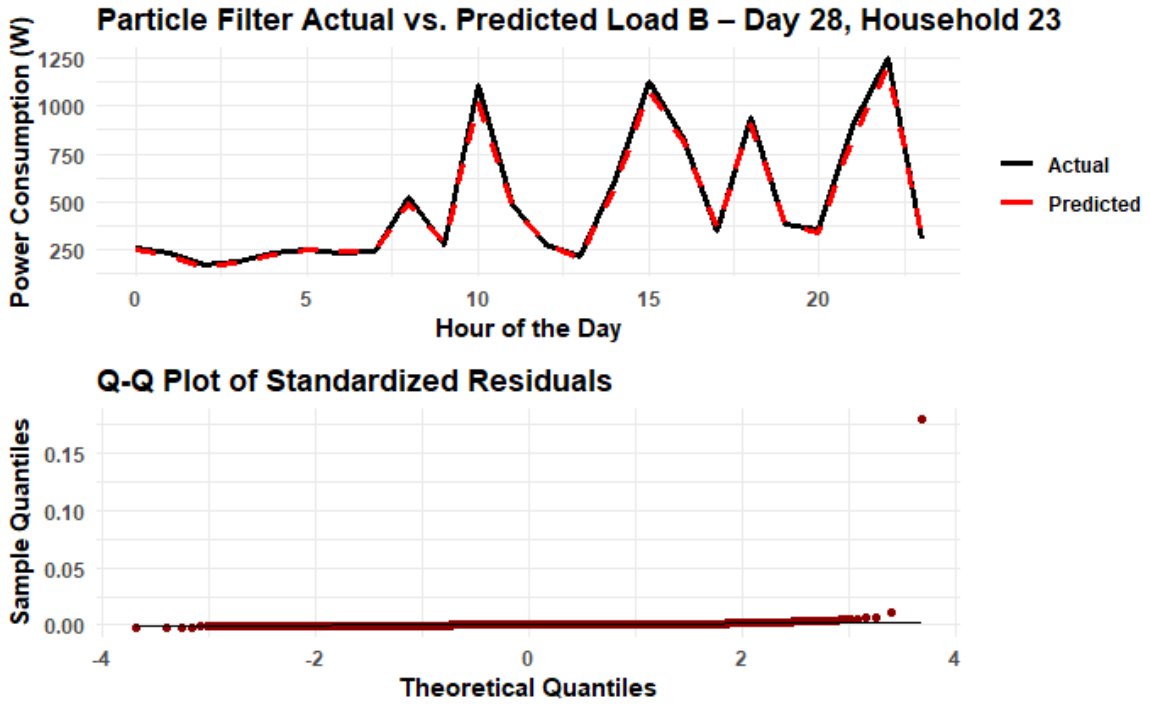

Figure 12: Particle Filter Load B Prediction with Residual - Household 23, Day 28

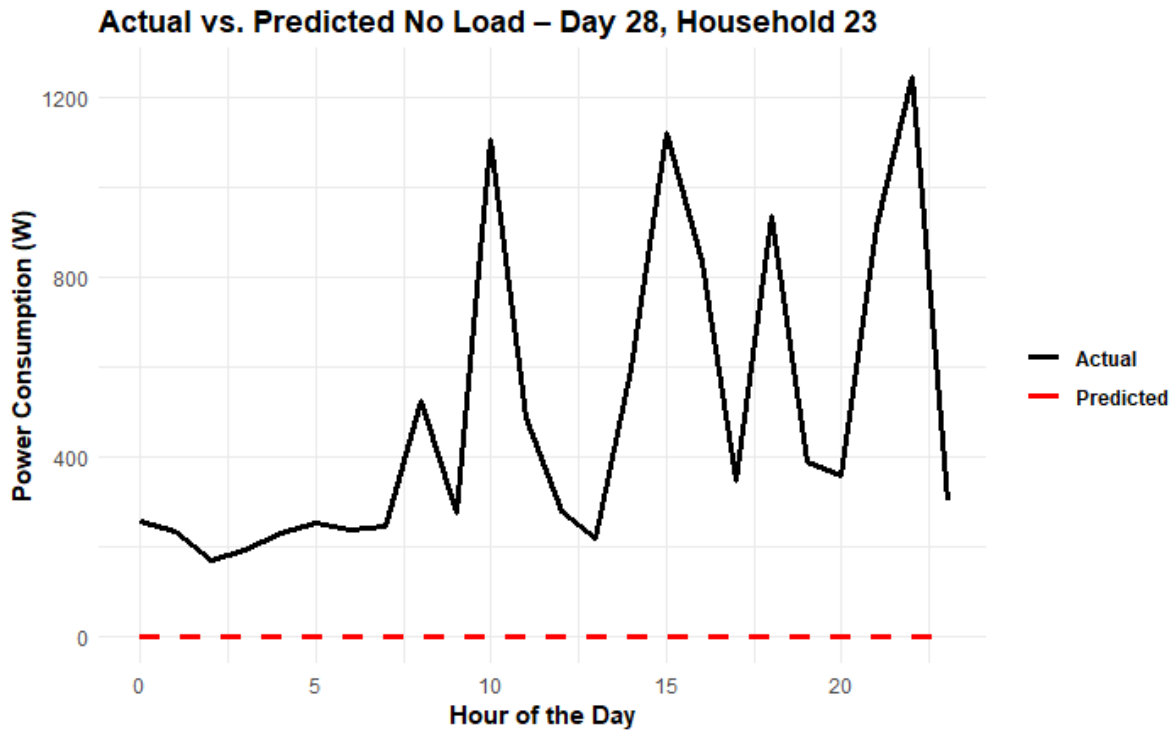

Figure 13: Particle Filter Prediction Without Load Profile for HH 3

4.9.2 Key Findings – Interpretation and Case Observations

Household 3

- Excellent visual match between actual and predicted values (Figure 32 and 33).
- Both Load A and B predictions track hour-by-hour fluctuations almost exactly.
- Very small residuals, with Q-Q plots showing high concentration near zero.
- Independence is accepted for Particle A but rejected for Particle B.
- Normality is rejected for both, despite low numerical variance.

Household 23

- Visual prediction quality is exceptionally high (Figure 11 and 12).
- Predicted curves overlap actual curves with minimal deviation, even around sharp peaks.
- Particle B residuals pass independence; Particle A fails.
- Both models fail normality, but residual magnitudes are very low.
- Lowest MSE across all households and filters, with Particle A = 124.60 in Table 8.

Table 7: Residual Diagnostics: Ljung-Box and Shapiro-Wilk for Particle Filter

Household	Model	Ljung-Box p	Independence	Shapiro-Wilk p	Normality
3	Particle A	0.9888	Accepted	4.43e-92	Rejected
	Particle B	2.2e-16	Rejected	2.00e-90	Rejected
	No Load	NA	Invalid	NA	Invalid
23	Particle A	0.0182	Rejected	9.16e-80	Rejected
	Particle B	0.8546	Accepted	1.65e-90	Rejected
	No Load	NA	Invalid	0	Rejected
40	Particle A	0.9724	Accepted	1.06e-82	Rejected
	Particle B	0.9242	Accepted	3.37e-36	Rejected
	No Load	NA	Invalid	0	Rejected

Household 40

- Predictions capture general trends but deviate significantly in amplitude, especially for Load B (Figure 34 and 35).
- Some peaks are overestimated or missed, though timing is relatively aligned.
- Residuals are more dispersed but still numerically small.
- Both models pass independence, yet normality is again rejected.
- Performance is affected by missing 109 days of data, limiting learning.

No Load Models

- Predictions are non-functional: predicted values collapse toward zero in all households (Figure 36, 13, and 37).
- Q-Q plots could not be evaluated meaningfully due to infinite or invalid residuals.
- Residual-based diagnostics fail to compute or return zero variance.
- MSEs are extremely large, confirming the importance of load profiles.

Table 8: MSE Comparison for Particle Filter Across Households and Models

Household	Particle A	Particle B	No Load
3	1,193.91	**1,139.62**	119,103.5
23	**124.60**	245.64	181,451.8
40	**1,237.90**	2,676.61	367,335.6

5 Summary

This project examined the forecasting of household-level electricity consumption using load profiles of 33 households equipped with heat pumps but without photovoltaic systems. The central goals were to model daily power consumption using smoothed multivariate curves, evaluate the influence of behavioral and seasonal factors, cluster households based on estimated effects, and predict short-term electricity demand with and without personalized load profiles.

Key findings of the analysis are summarized below:

- MANOVA showed that household, weekday, and season significantly affect load curves; interaction models had better AIC.
- Clustering on weekday curves produced strong group separation (Dunn index up to 2.48); weekend clusters were weaker.
- Forecasts for households 3, 23, and 40 improved when using personalized or cluster-based load profiles.
- Among forecasting models, Particle Filter B achieved the lowest MSE across all tested households: 1,139.62 (Household 3), 124.60 (Household 23), and 1,237.90 (Household 40). For comparison, No Load profile forecasts yielded substantially higher MSE values.
- Particle filters had better residual independence (Ljung-Box test) than kalman filters, though normality was often violated.
- Removing customized profiles reduced forecast accuracy, underscoring their benefit.

These results show the value of combining smoothing, clustering, and state-space filters for short-term electricity forecasting, and highlight the usefulness of MANOVA in identifying behavioral consumption patterns.

Despite strong results, caution is warranted. Residual diagnostics suggest model assumptions, particularly normality, may not hold consistently. Particle filters performed well but require careful tuning due to sensitivity to initialization.

Future research could explore the following directions:

- Incorporating temperature more explicitly into the forecasting models as a dynamic covariate.
- Refining cluster analysis by constructing separate weekday and weekend clusters to better capture behavioral heterogeneity.
- Investigating advanced or hybrid forecasting techniques, such as deep learning models, that can model non-linear and temporal dependencies.

Overall, this project reinforces the value of combining multivariate statistical methods and filtering-based forecasting in managing residential energy consumption. The synergy between data-driven clustering and personalized modeling offers a promising approach to enhance demand-side energy planning.

References

Brock, Gordon et al. (2008). "clValid: An R Package for Cluster Validation". In: *Journal of Statistical Software* 25.4, pp. 1–22.

Brockwell, Peter J. and Richard A. Davis (2016). *Introduction to Time Series and Forecasting*. 3rd ed. New York: Springer.

Cappé, Olivier, Simon Godsill, and Eric Moulines (2007). "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo". In: *Proceedings of the IEEE* 95.5, pp. 899–924.

Grolemund, Garrett and Hadley Wickham (2011). "Dates and Times Made Easy with lubridate". In: *Journal of Statistical Software* 40.3, pp. 1–25.

Heiberger, Richard M. and Burt Holland (2015). *Statistical Analysis and Data Display: An Intermediate Course with Examples in R*. 2nd ed. New York: Springer.

Kalman, Rudolf E. (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1, pp. 35–45.

Kassambara, Alboukadel (2017). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.5. URL: https://CRAN.R-project.org/package=factoextra.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rencher, Alvin C. (2002). *Methods of Multivariate Analysis*. 2nd ed. New York: Wiley-Interscience.

Rousseeuw, Peter J. (1987). "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis". In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65.

Shapiro, Samuel S. and Martin B. Wilk (1965). "An Analysis of Variance Test for Normality (Complete Samples)". In: *Biometrika* 52.3/4, pp. 591–611.

Walesiak, Marek and Adam Dudek (2020). *clusterSim: Searching for Optimal Clustering Procedure for a Data Set*. R package version 0.48-5. URL: https://cran.r-project.org/package=clusterSim.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

– (2022). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.10. URL: https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley et al. (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686.

A Additional figures

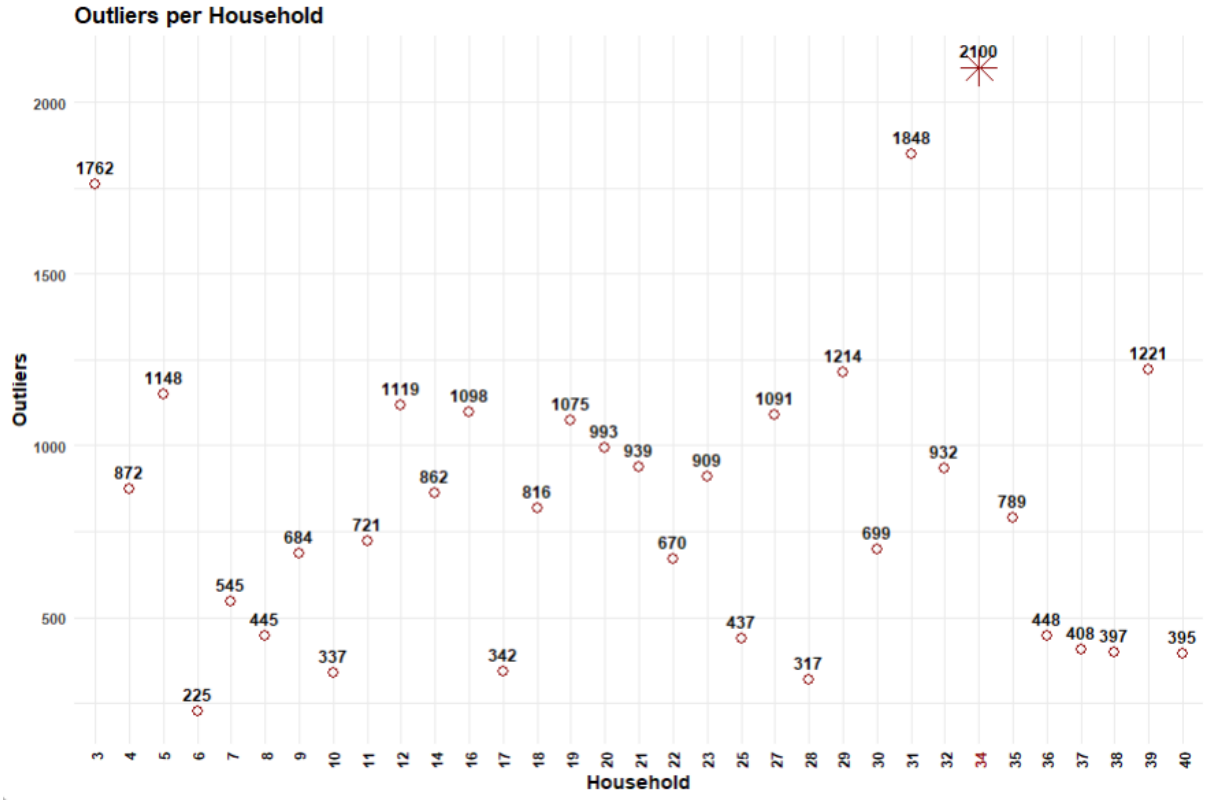

Figure 14: Number of outliers detected per household

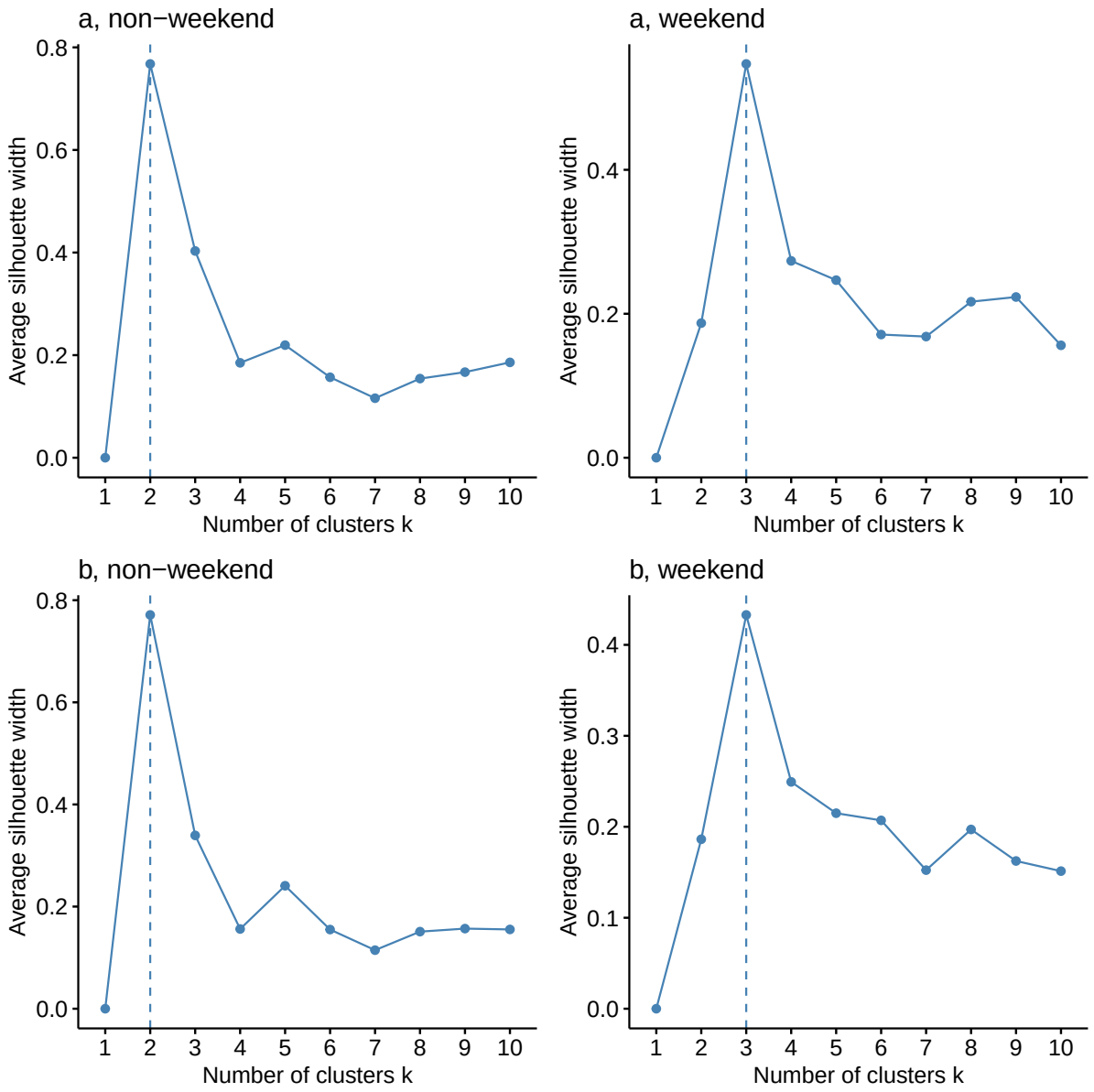

Figure 15: Silhouette plots used to identify the optimal number of clusters

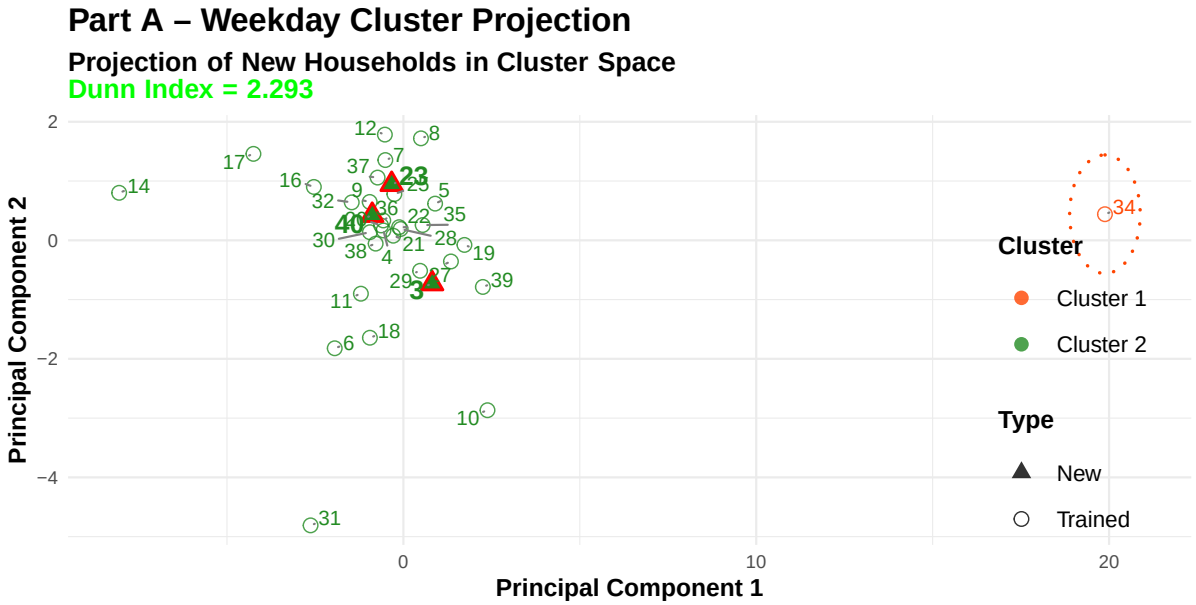

Figure 16: Part A – Weekday Cluster Projection

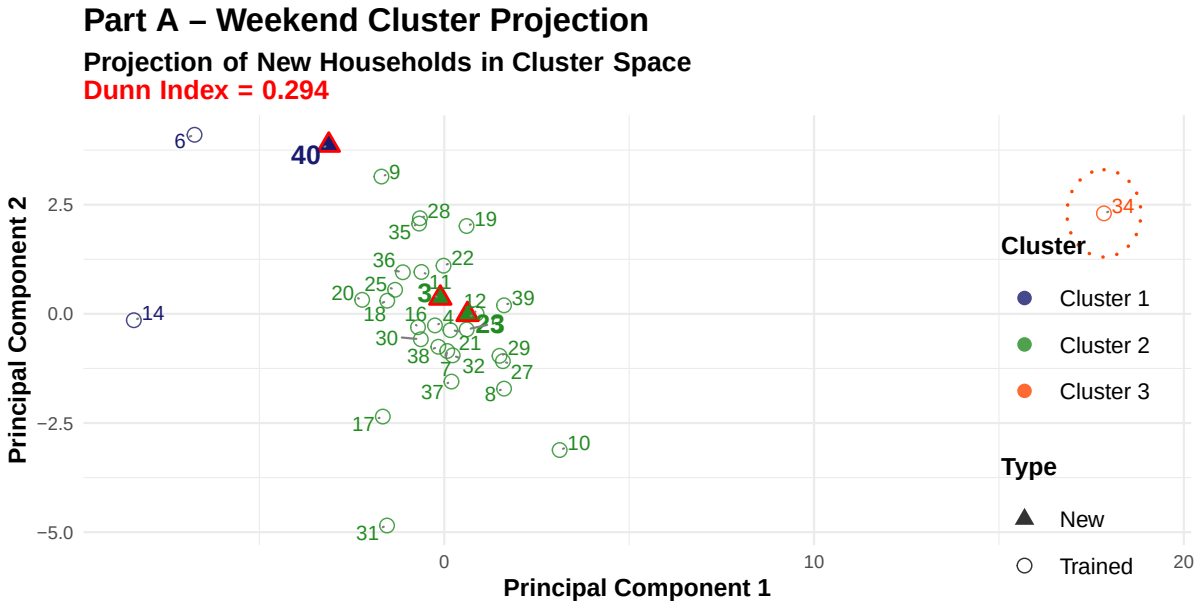

Figure 17: Part A – Weekend Cluster Projection

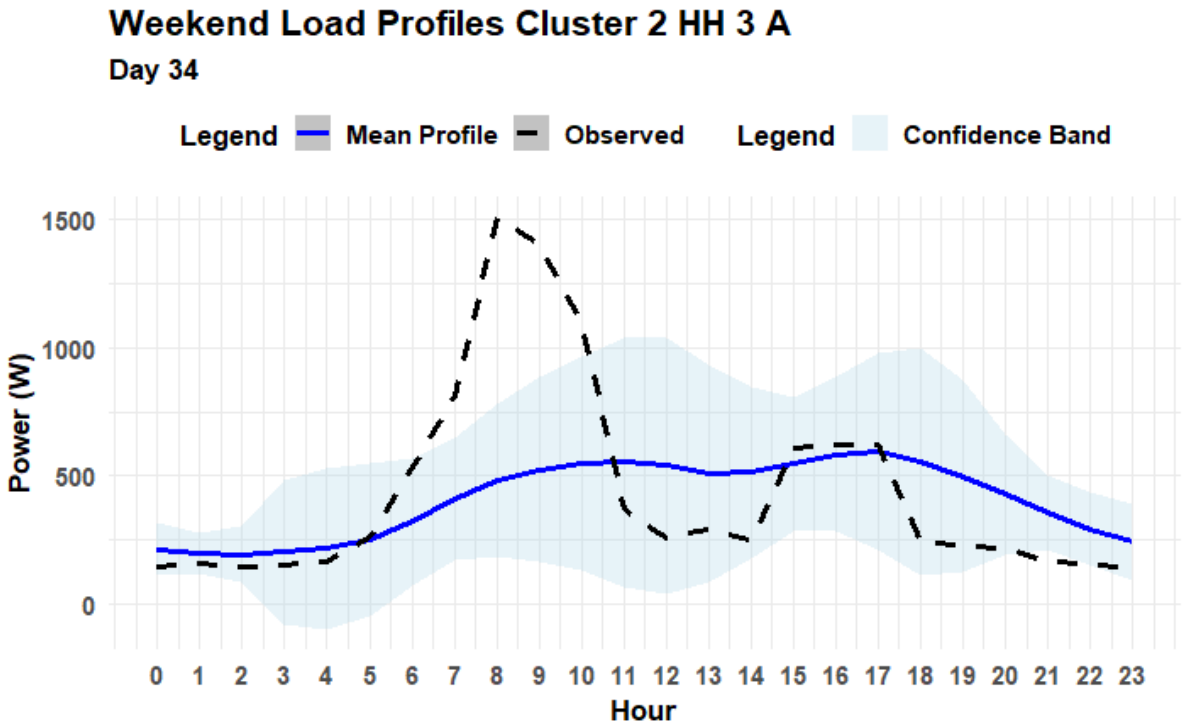

Figure 18: Weekend - Household 3 A Load Profile

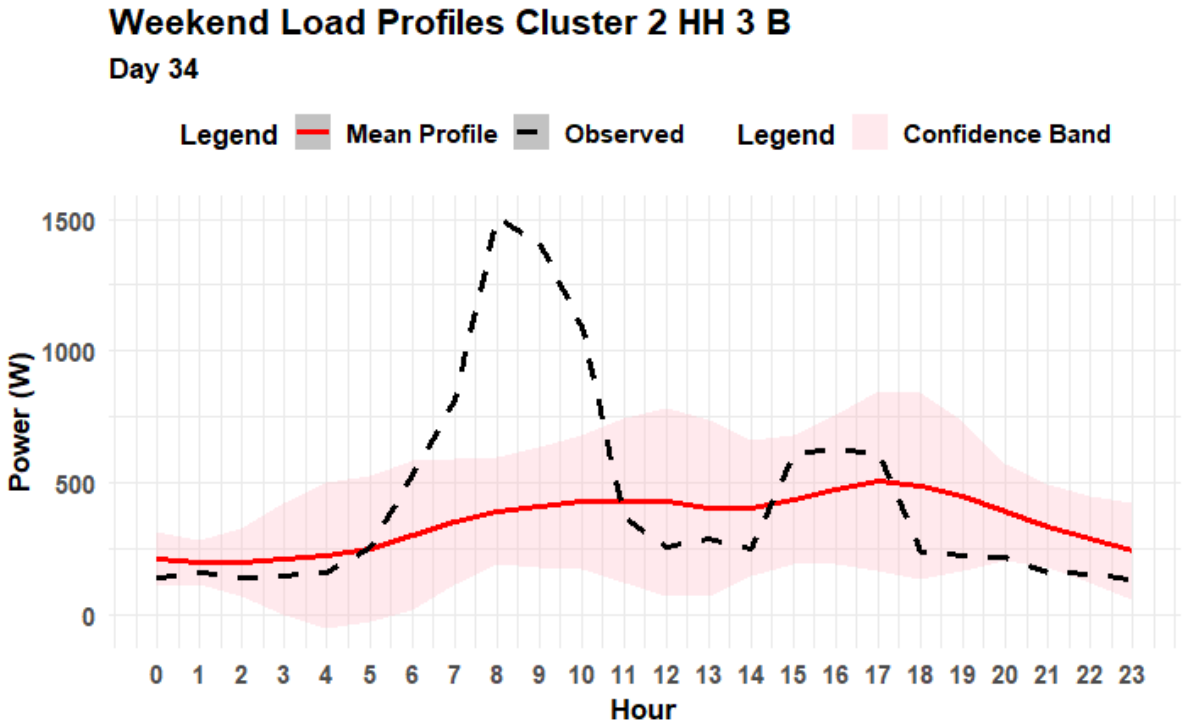

Figure 19: Weekend - Household 3 B Load Profile

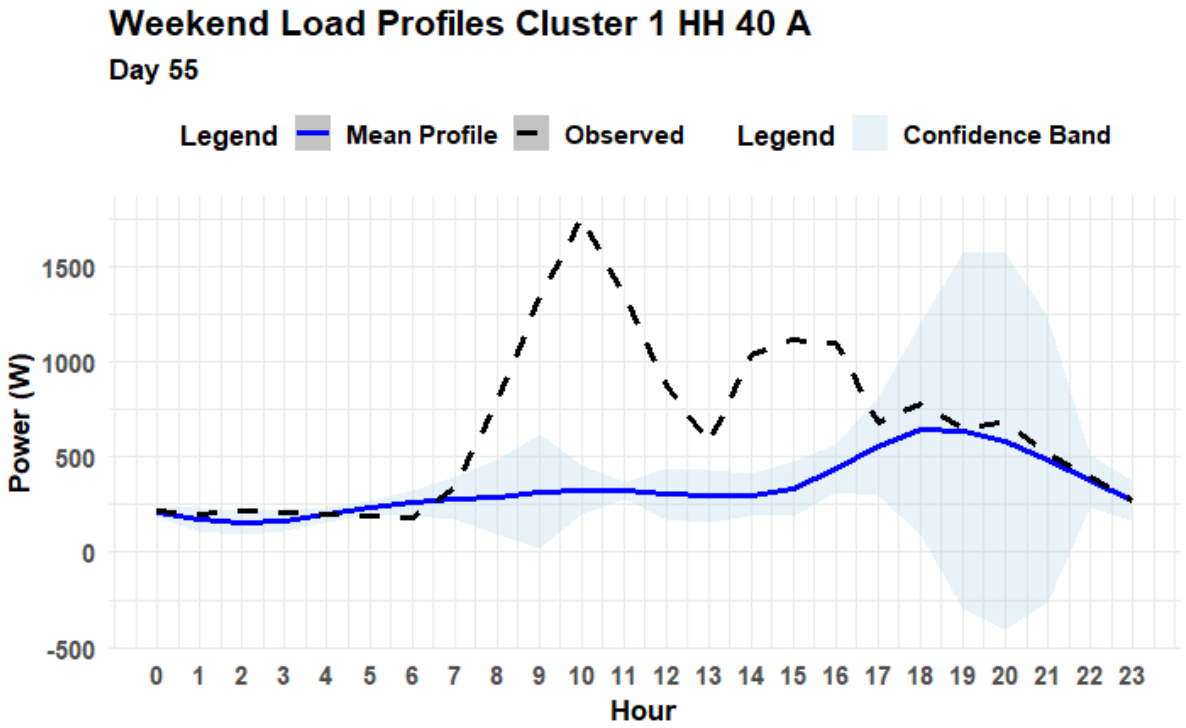

Figure 20: Weekend - Household 40 A Load Profile

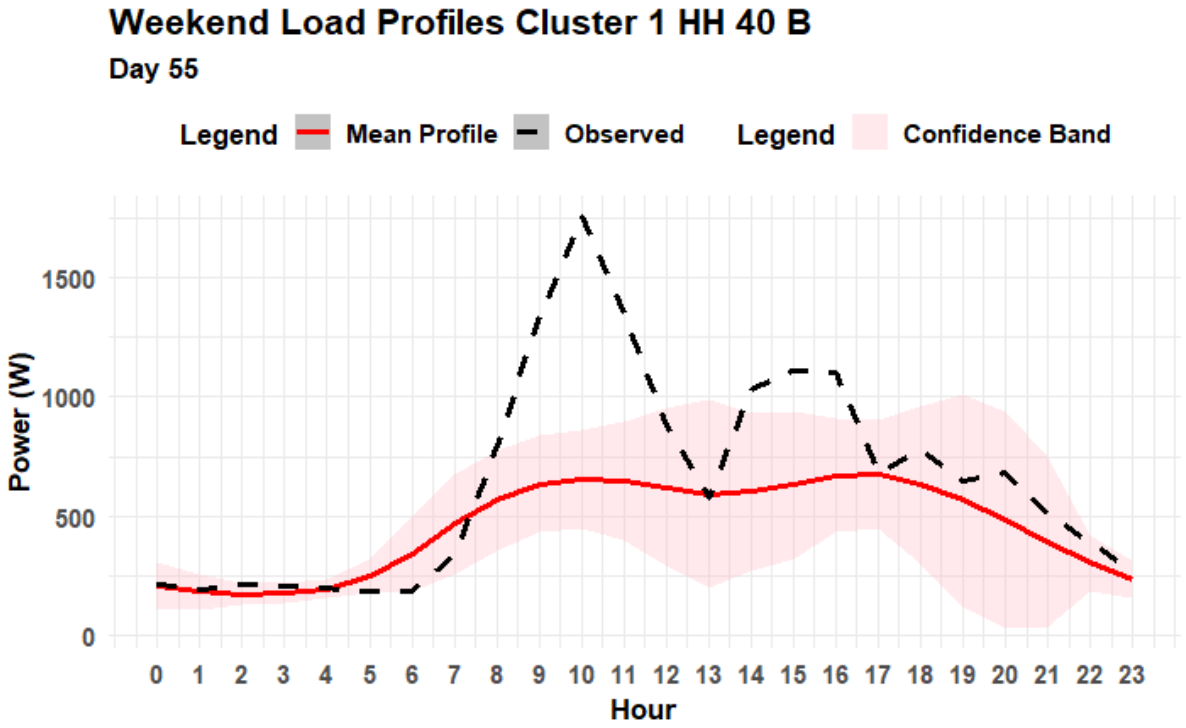

Figure 21: Weekend - Household 40 B Load Profile

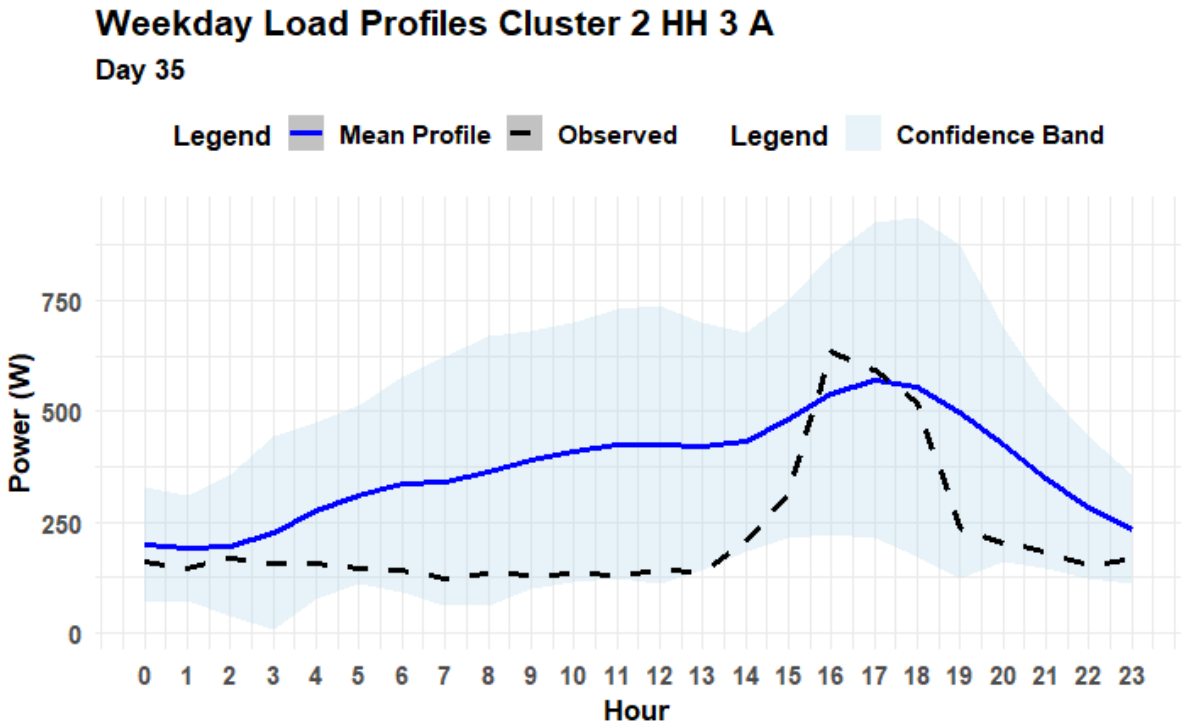

Figure 22: Weekday - Household 3 A Load Profile

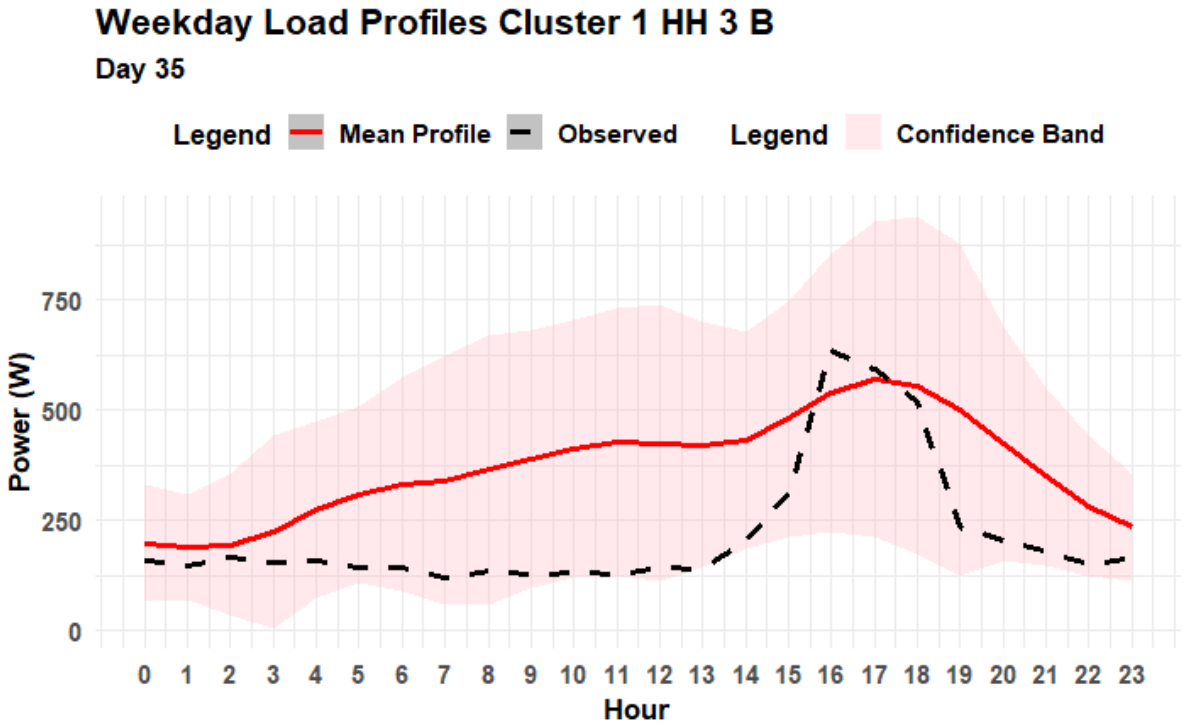

Figure 23: Weekday - Household 3 B Load Profile

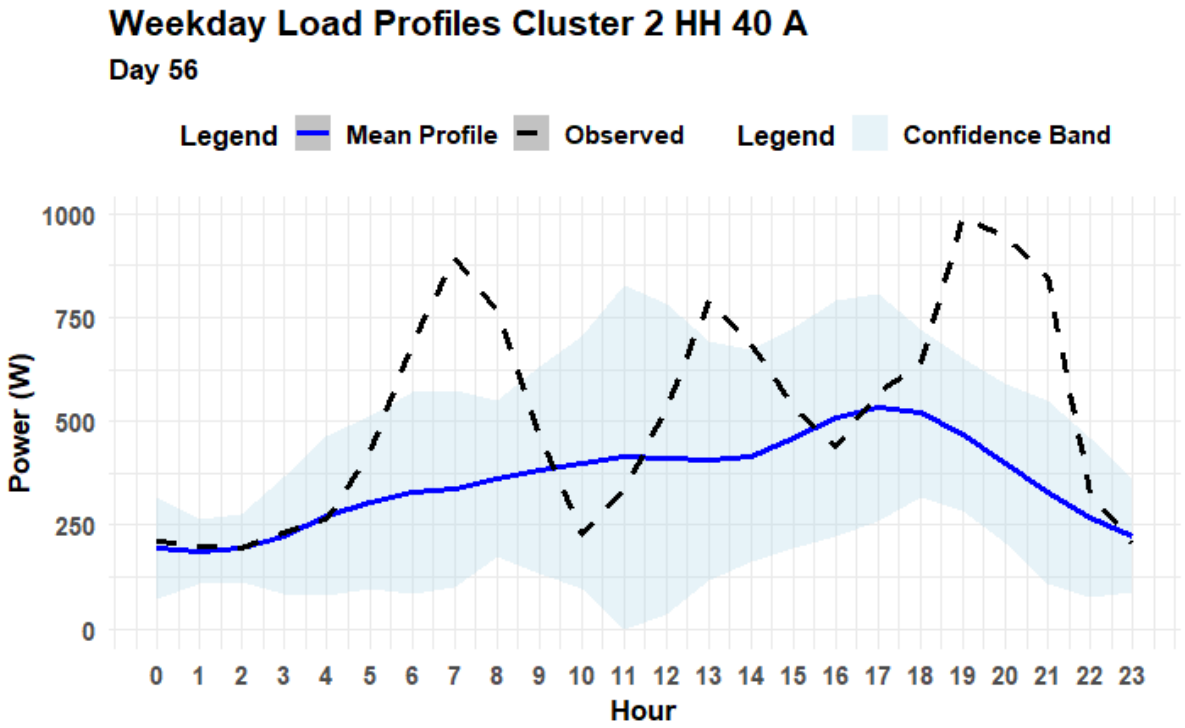

Figure 24: Weekday - Household 40 A Load Profile

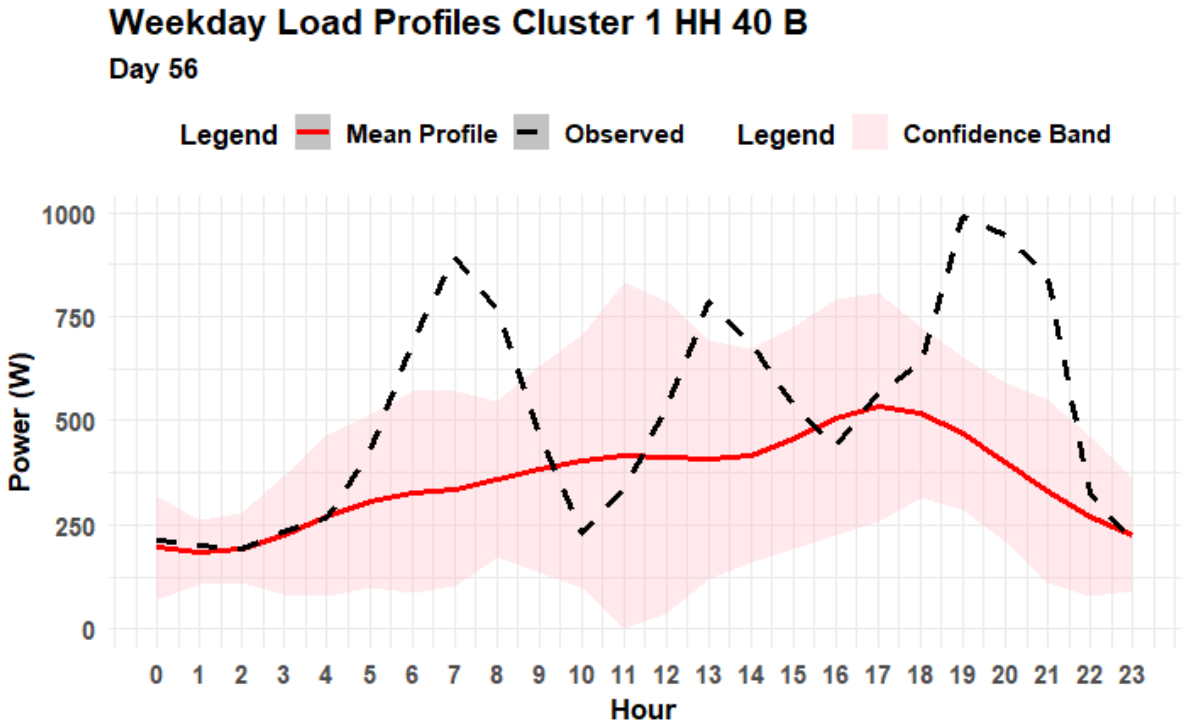

Figure 25: Weekday - Household 40 B Load Profile

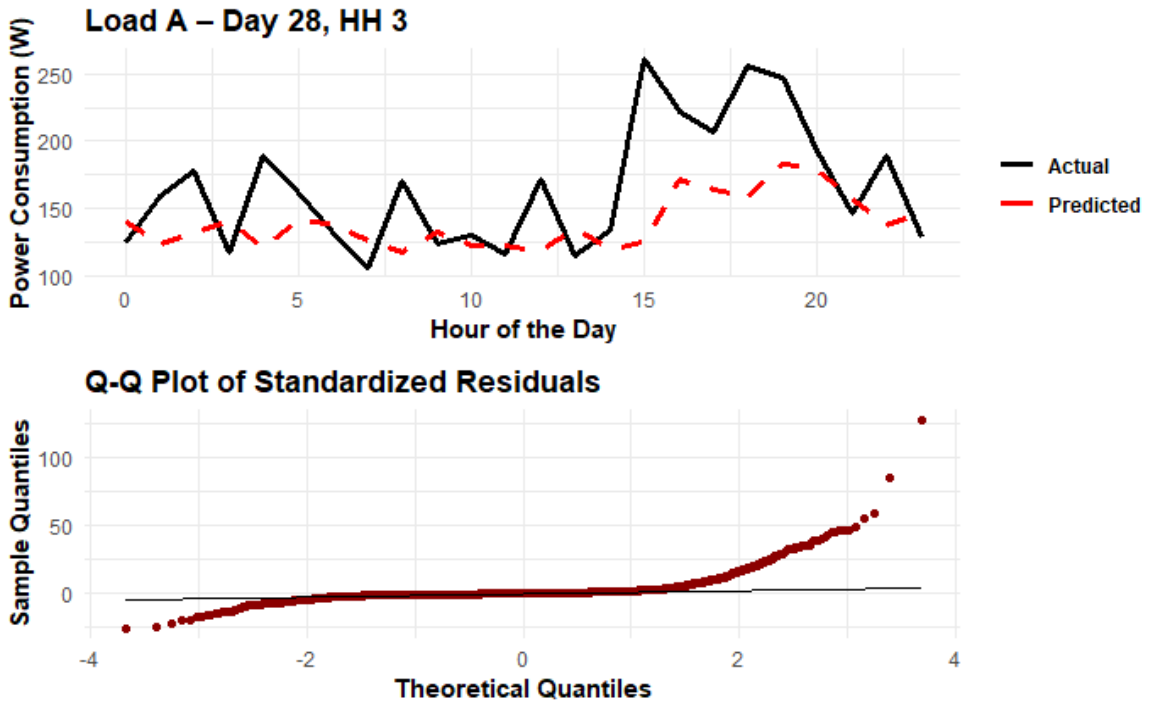

Figure 26: Kalman Filter Load A Predictions with Residuals – HH 3, Day 28

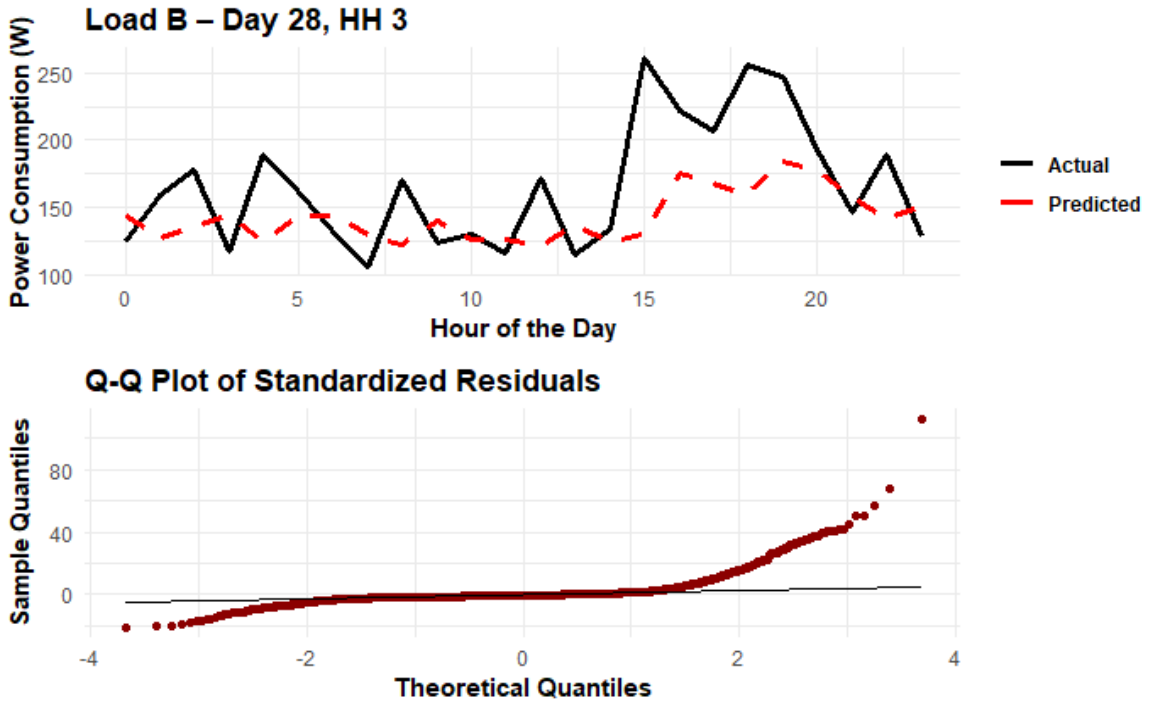

Figure 27: Kalman Filter Load B Predictions with Residuals – HH 3, Day 28

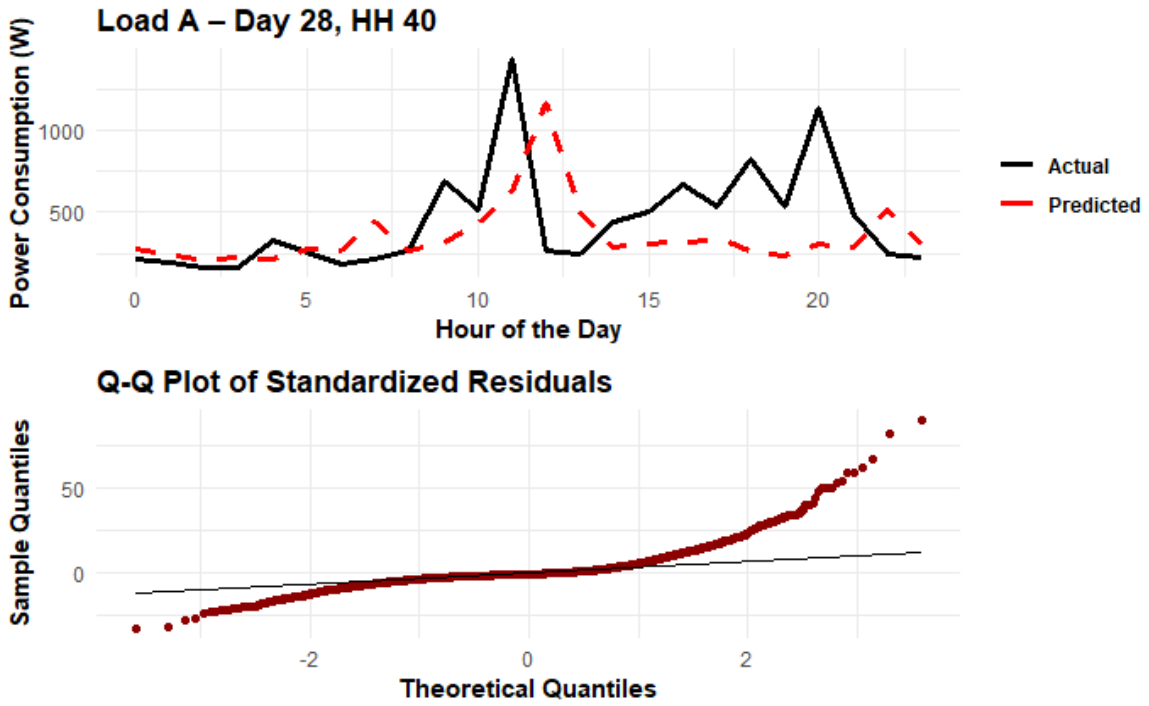

Figure 28: Kalman Filter Load A Predictions with Residuals - HH 40, Day 28

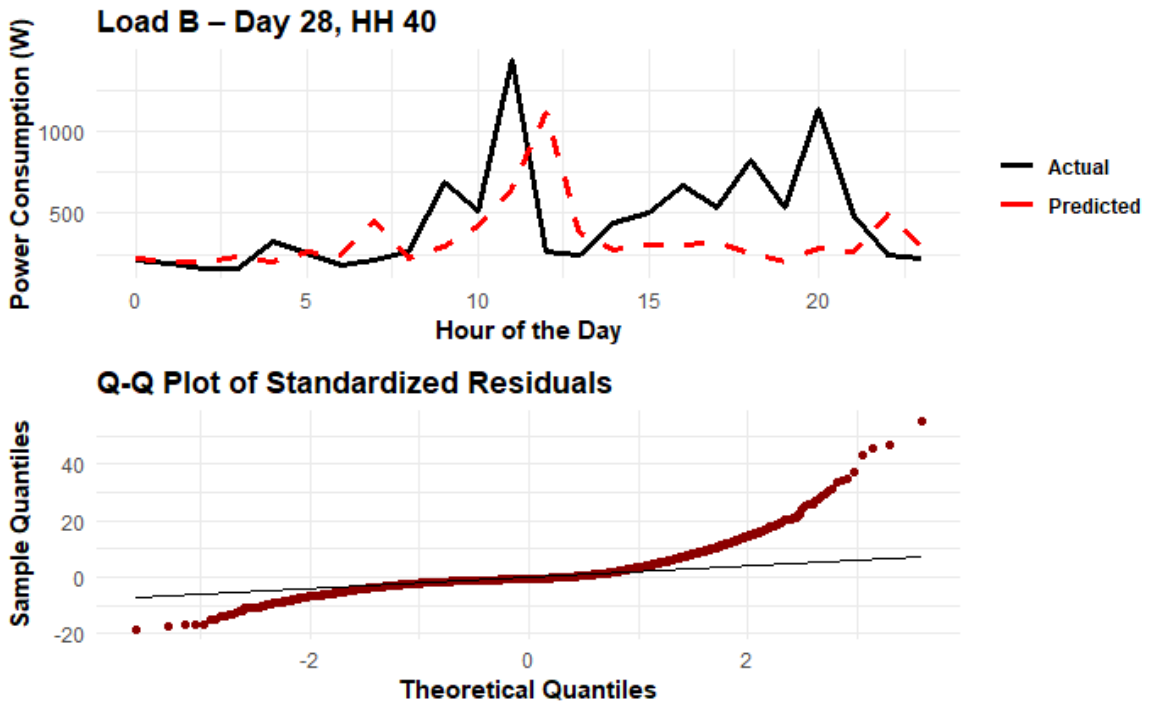

Figure 29: Kalman Filter Load B Predictions with Residuals - HH 40, Day 28

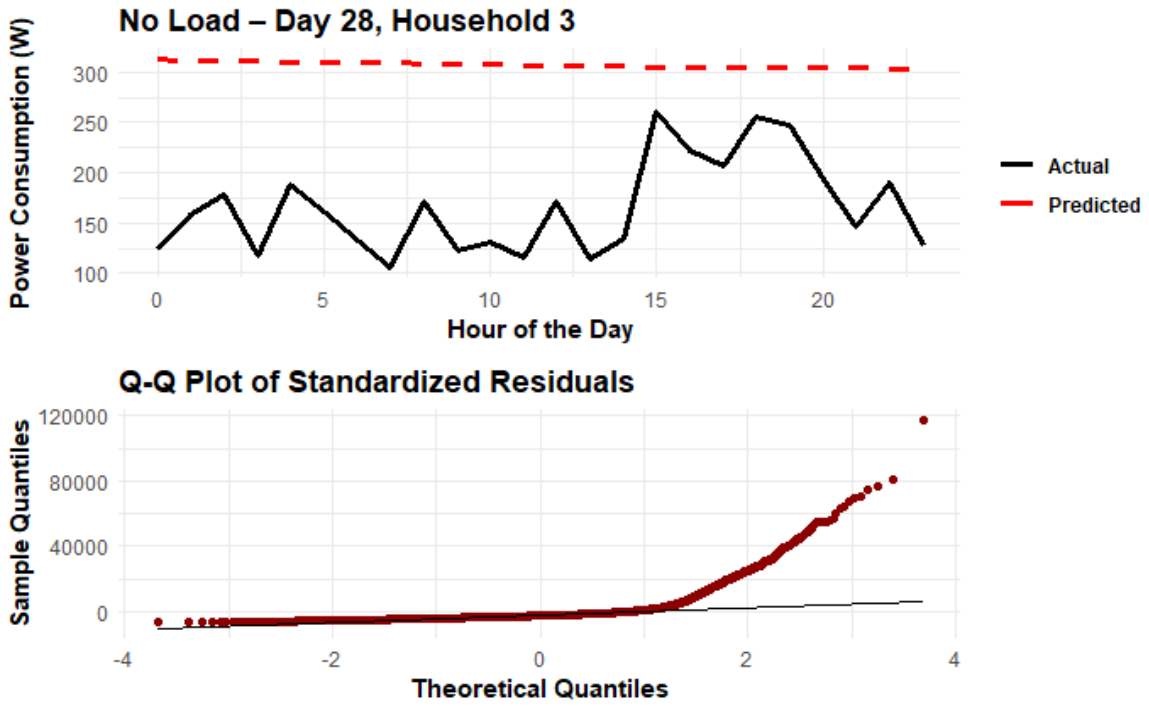

Figure 30: Prediction and Q-Q Residual Plot for No Load Model (HH 3)

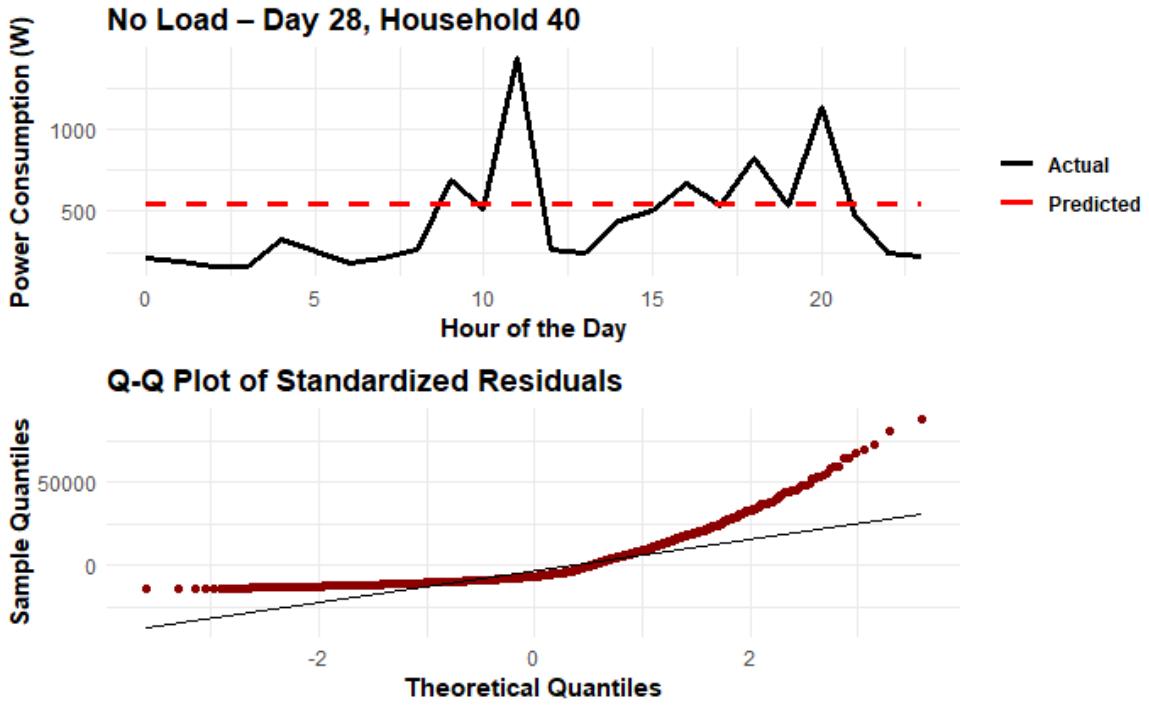

Figure 31: Prediction and Q-Q Residual Plot for No Load Model (HH 40)

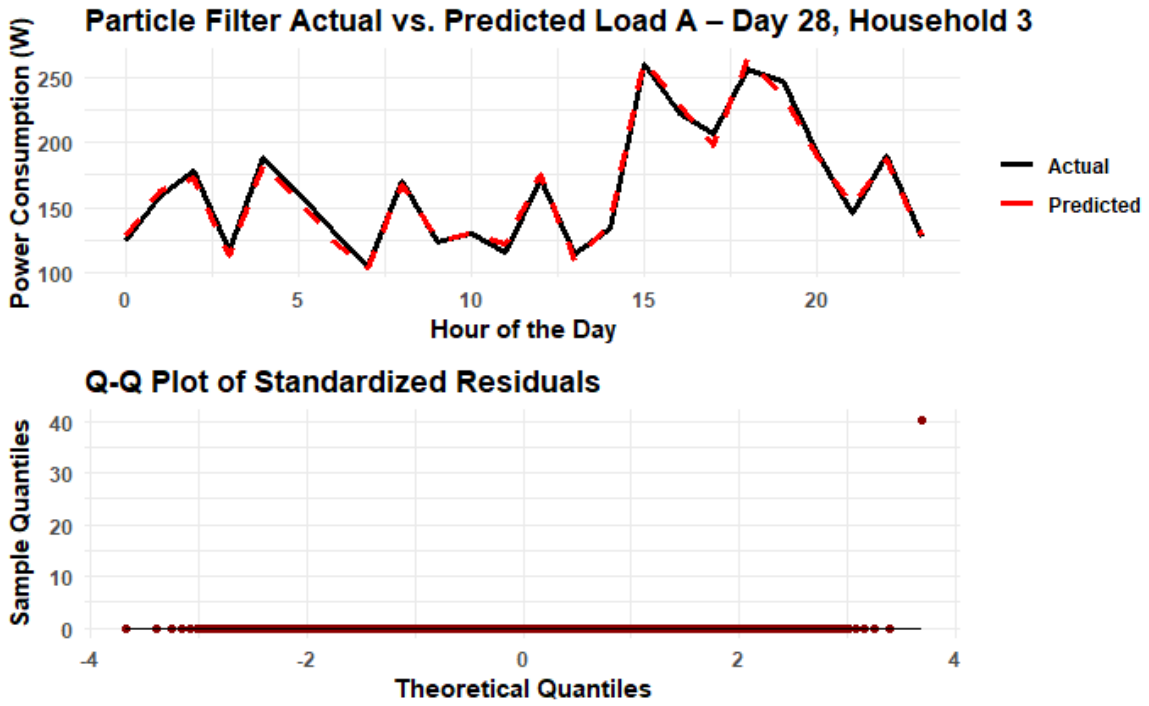

Figure 32: Particle Filter Load A Prediction with Residual - Household 3, Day 28

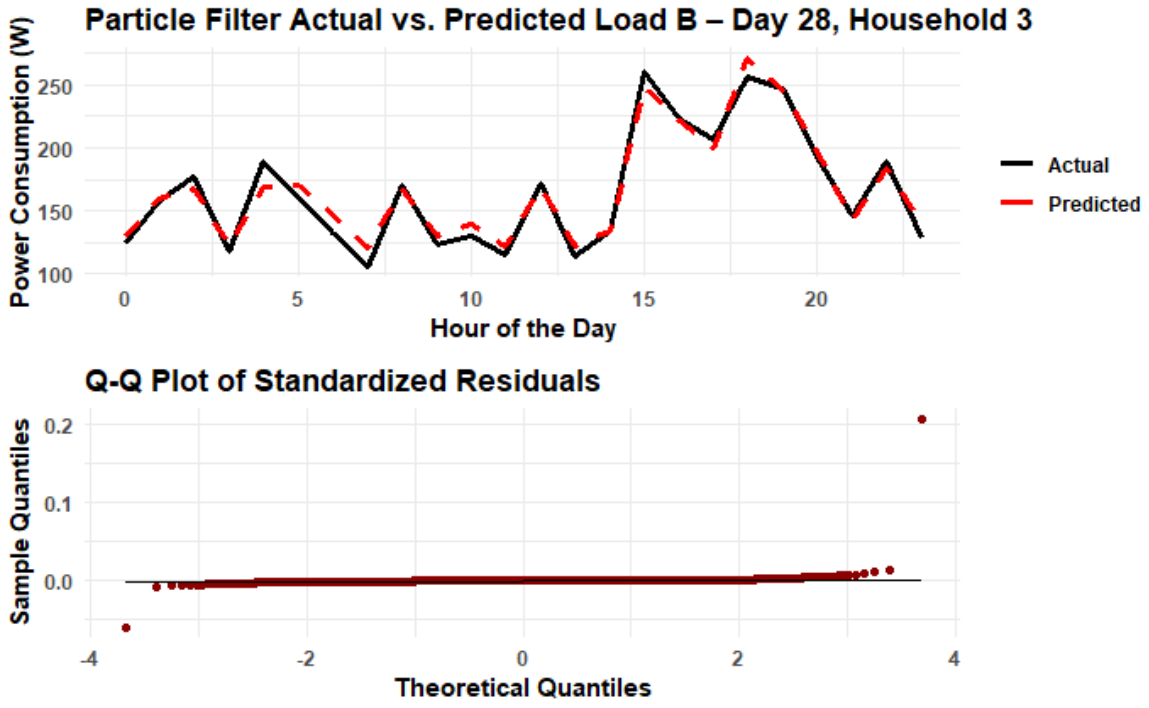

Figure 33: Particle Filter Load B Prediction with Residual - Household 3, Day 28

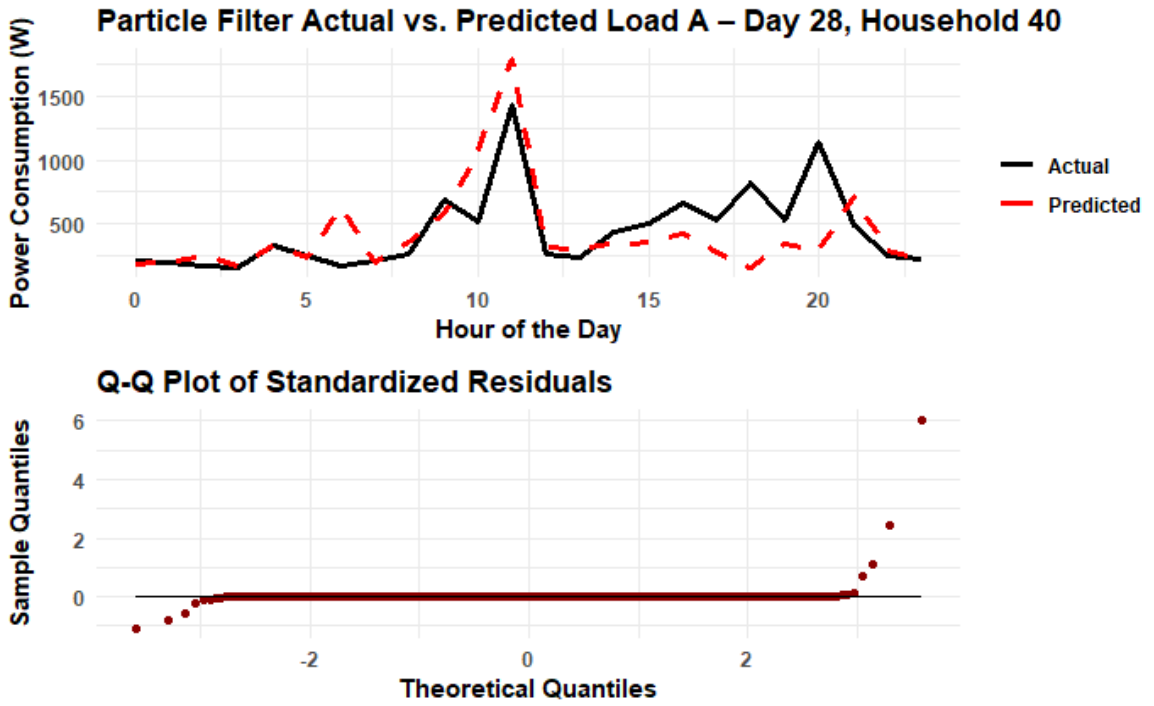

Figure 34: Particle Filter Load A Prediction with Residual - Household 40, Day 28

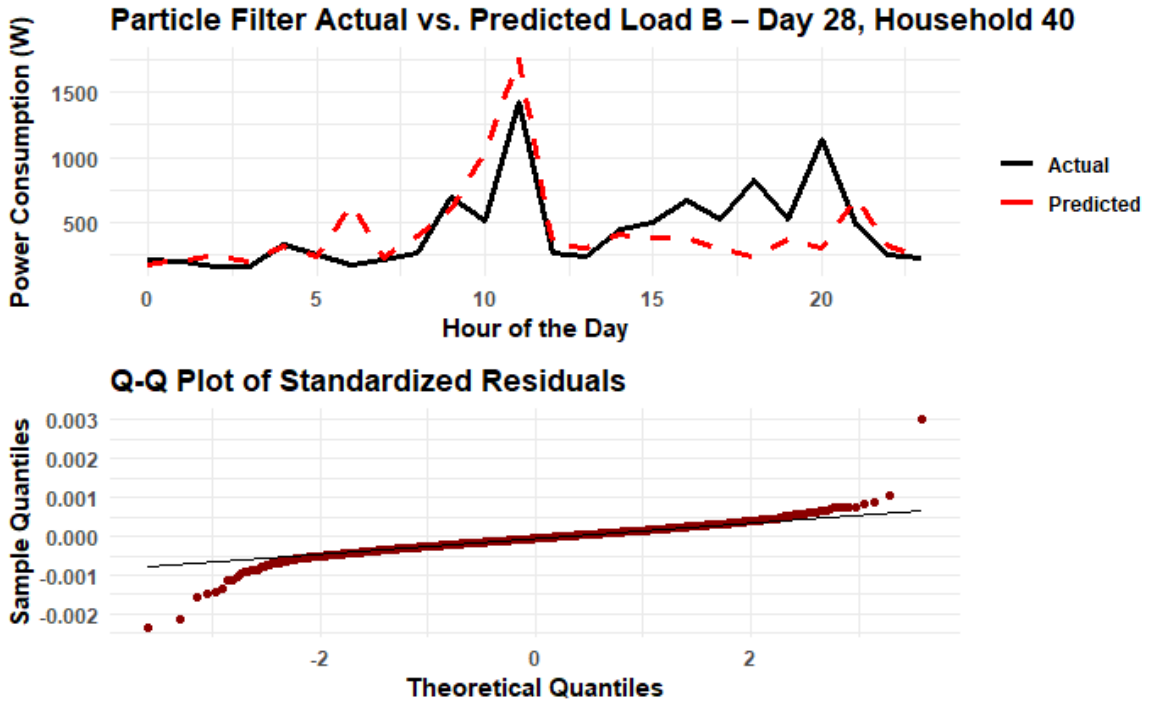

Figure 35: Particle Filter Load B Prediction with Residual - Household 40, Day 28

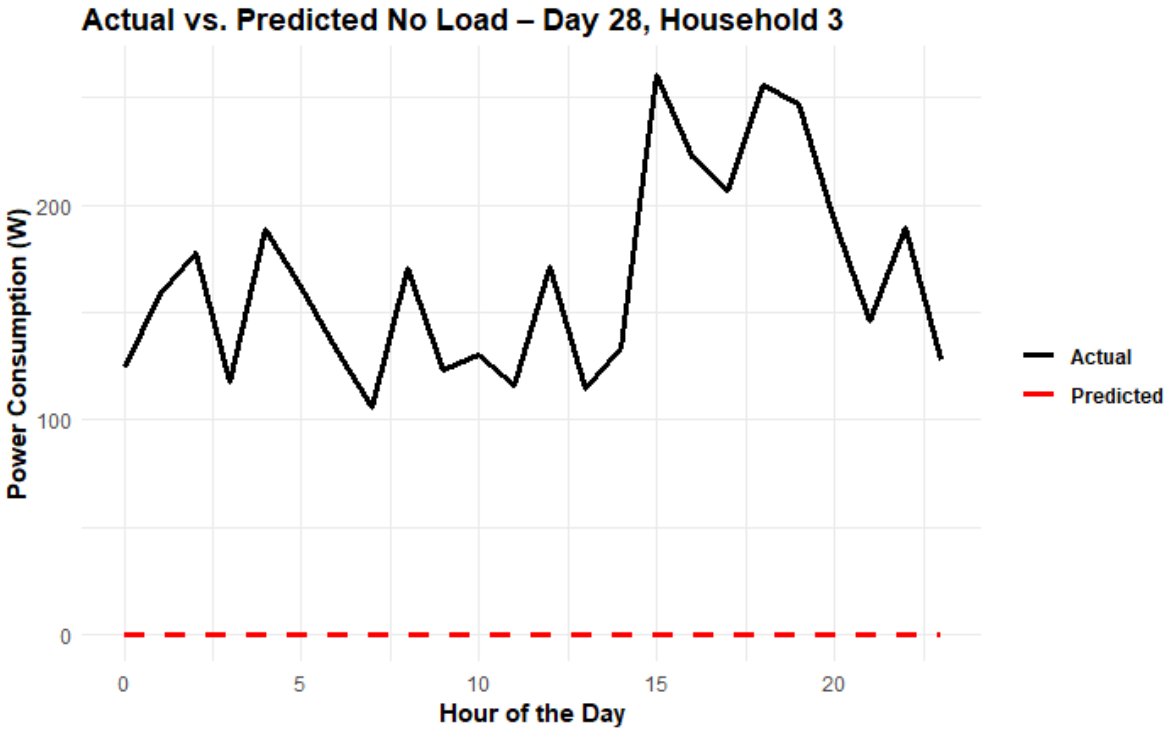

Figure 36: Particle Filter Prediction Without Load Profile for HH 3

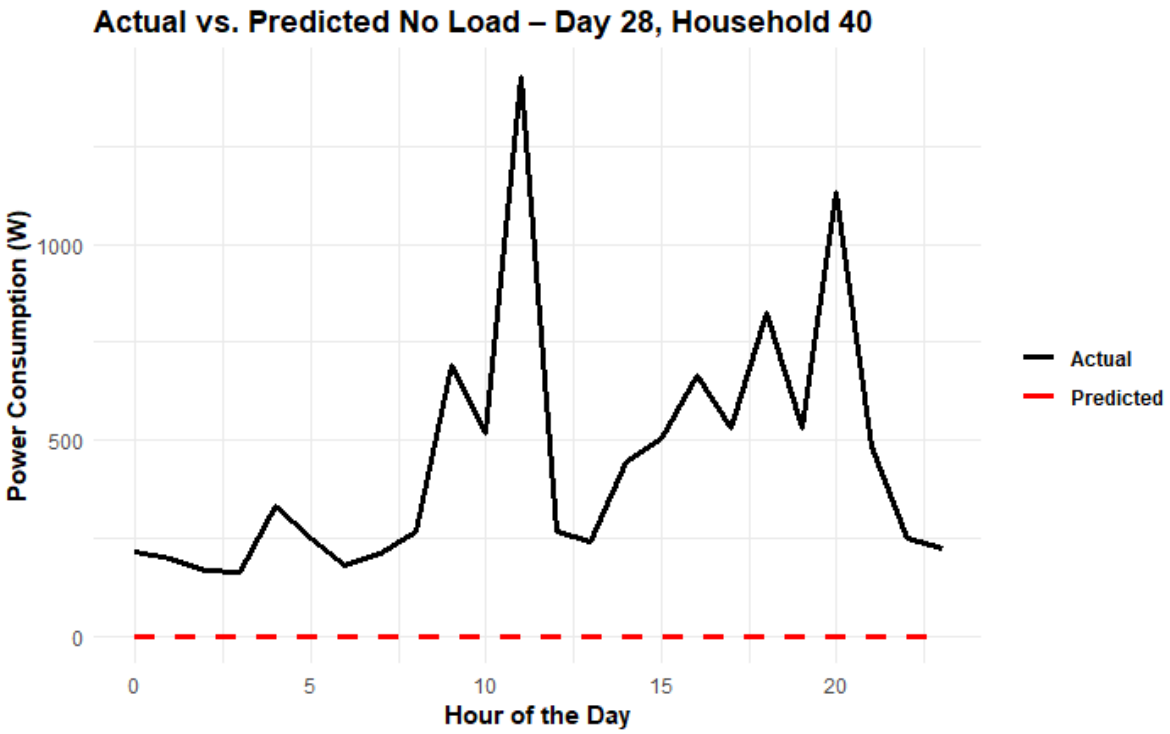

Figure 37: Particle Filter Prediction Without Load Profile for HH 40