



# University of Padova

---

DEPARTMENT OF MATHEMATICS TULLIO-LEVI CIVITA

MASTER THESIS IN DATA SCIENCE

## Linear and machine learning cross-temporal forecast reconciliation: An empirical investigation

SUPERVISOR

PROFESSOR LUISA BISAGLIA  
UNIVERSITY OF PADOVA

Co-SUPERVISOR

DOCTOR DANIELE GIROLIMETTO  
UNIVERSITY OF PADOVA

MASTER CANDIDATE

ROYA GHAMARI  
2071969

ACADEMIC YEAR

2024-2025



I DEDICATE THIS WORK TO MY PARENTS AND MY BROTHER, WHO HAVE ALWAYS SUPPORTED ME. I ALSO WANT TO DEDICATE THIS THESIS TO MY COUSIN, WHOSE LIFE WAS CUT SHORT AT A YOUNG AGE. THE PAIN OF LOSING HER WILL ALWAYS REMAIN WITH US, AND SHE WILL ALWAYS BE REMEMBERED WITH LOVE.



# Abstract

Forecast reconciliation is a post forecasting process that ensures coherence when dealing with linear constraints time series, aligning predictions at different levels of aggregation. Exploring the cross-temporal framework, this thesis compares different linear and machine learning-based solutions with two empirical applications: Citi Bike rental demand and Energy Load. While both datasets share the same temporal structure, they differ in cross-sectional hierarchy and industry context, allowing for a broader evaluation of reconciliation methods.

Key concepts in this study include evaluating different reconciliation strategies by exploring various covariance matrix structures, refining residual handling in linear and machine learning-based models, and ensuring consistency in non-negativity constraints across all reconciliation methods. The results indicate that, in contrast to prior findings, linear reconciliation methods consistently outperformed ML-based approaches. Among the linear models, heuristic and optimal linear reconciliation approaches demonstrated the highest forecast accuracy. On the other hand, Random Forest remained the strongest ML-based reconciliation method.

Sensitivity analyses revealed that including a wide range of temporal aggregation levels generally improves accuracy, reinforcing the value of comprehensive reconciliation structures. The choice between compact and complete feature matrices had a notable impact on ML reconciliation performance, with its effects varying across hierarchical levels. These findings highlight the importance of methodological choices in reconciliation and their influence on forecast accuracy.

Future work could explore reconciliation using neural network-based ML algorithms, automated feature selection techniques, and online learning approaches for dynamic model updates and improved computational efficiency.



# Contents

ABSTRACT	v
LIST OF FIGURES	ix
LIST OF TABLES	xi
LISTING OF ACRONYMS	xiii
<b>1 INTRODUCTION</b>	<b>I</b>
1.1 Time Series, Forecasting, and Hierarchical Structure . . . . .	I
1.2 Forecast Reconciliation Approaches . . . . .	2
1.3 Context of the Data . . . . .	3
1.4 Outlines of the Thesis . . . . .	4
<b>2 OBJECTIVES</b>	<b>7</b>
<b>3 DATA</b>	<b>9</b>
3.1 Bicycle sharing platform data . . . . .	9
3.2 Energy load data of Italy . . . . .	15
<b>4 METHODOLOGY</b>	<b>23</b>
4.1 Forecast set-up . . . . .	23
4.2 Traditional Time Series Models for Base Forecasting . . . . .	24
4.3 Machine Learning Methods for Forecast Reconciliation . . . . .	25
4.4 Linear Reconciliation Benchmarks for Cross-Temporal Hierarchies . . . . .	27
4.4.1 Reconciliation forecast methods . . . . .	27
4.4.2 Covariance matrices . . . . .	28
4.5 Evaluation Metrics . . . . .	29
4.5.1 Weighted Absolute Percentage Error . . . . .	29
4.5.2 Mean Absolute Scaled Error . . . . .	30
<b>5 RESULTS AND DISCUSSION</b>	<b>33</b>
5.1 Results for bicycle sharing data . . . . .	33
5.1.1 Overall forecast performance . . . . .	33
5.1.2 Sensitivity to Feature Matrices . . . . .	34
5.1.3 Sensitivity to Temporal Aggregation Orders . . . . .	41

5.2	Results for energy load data . . . . .	42
5.2.1	Overall forecast performance . . . . .	42
5.2.2	Sensitivity to Feature Matrices . . . . .	46
5.2.3	Sensitivity to Temporal Aggregation Orders . . . . .	56
6	CONCLUSION	63
6.1	Addressing the Study Objectives . . . . .	63
6.2	Summary of Results and Dataset Comparisons . . . . .	64
6.3	Limitations and Future Works . . . . .	66
	REFERENCES	67
	ACKNOWLEDGMENTS	71
	APPENDIX A ADDITIONAL FIGURES FOR DESCRIBING CITI BIKE DATASET	73
	APPENDIX B ADDITIONAL FIGURES FOR DESCRIBING ENERGY LOAD DATASET	77
	APPENDIX C MASE ACCURACY INDEX FOR CITI BIKE FORECASTS	81
	APPENDIX D MASE ACCURACY INDEX FOR ENERGY LOAD FORECASTS	89

# Listing of figures

3.1	Summary statistics for Citi Bike data. . . . .	10
3.2	Monthly bike demand distribution across different areas. . . . .	11
3.3	Decomposition analysis of bike demand load for NYC. . . . .	12
3.4	Bike demand time series over the year 2023: regional trends across NYC. . . . .	13
3.5	Daily time series plots of Bike demand: median and variability across different regions of NYC. . . . .	14
3.6	Bike demand for NYC and areas. . . . .	15
3.7	Summary statistics for Energy Load data. . . . .	17
3.8	Monthly Energy Load distribution across different regions. . . . .	18
3.9	Decomposition analysis of Energy load for Italy. . . . .	19
3.10	Energy Load time series over the year 2023: regional trends across Italy. . . . .	20
3.11	Daily time series plots of Energy Load: median and variability across different regions of Italy. . . . .	21
3.12	Energy Load for Italy and regions. . . . .	22
A.1	Boxplot of Citi Bike data by area. . . . .	73
A.2	Hourly bike demand variability across different regions of NYC. . . . .	74
A.3	Correlation matrix of Citi Bike. . . . .	74
A.4	Daily Bike demand patterns across different regions of NYC (January 1–7, 2023). . . . .	75
B.1	Boxplot of Energy Load data by region. . . . .	77
B.2	Hourly Energy Load variability across different regions of Italy. . . . .	78
B.3	Correlation matrix of Energy Load. . . . .	78
B.4	Daily Energy Load patterns across different regions of Italy (January 1–7, 2023). . . . .	79



# Listing of tables

5.1	Citi Bike H <sub>3</sub> Cells Forecast Results for Naive Base Method. . . . .	34
5.2	Citi Bike H <sub>3</sub> Cells Forecast Results for ETS Base Method. . . . .	35
5.3	Citi Bike H <sub>3</sub> Cells Forecast Results for SARIMA Base Method. . . . .	36
5.4	Citi Bike H <sub>3</sub> Cells Forecast Results for Forecast Combination Base Method. .	37
5.5	Citi Bike Market Forecast Results for Naive Base Method. . . . .	37
5.6	Citi Bike Market Forecast Results for ETS Base Method. . . . .	38
5.7	Citi Bike Market Forecast Results for SARIMA Base Method. . . . .	39
5.8	Citi Bike Market Forecast Results for Forecast Combination Base Method. .	40
5.9	Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike Naive Base Method. . . . .	42
5.10	Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike ETS Base Method. . . . .	43
5.11	Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike SARIMA Base Method. . . . .	44
5.12	Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike Forecast Combination Base Method. . . . .	45
5.13	Energy Load Zones Forecast Results for Naive Base Method. . . . .	46
5.14	Energy Load Zones Forecast Results for ETS Base Method. . . . .	47
5.15	Energy Load Zones Forecast Results for SARIMA Base Method. . . . .	48
5.16	Energy Load Zones Forecast Results for Forecast Combination Base Method. .	49
5.17	Energy Load Areas Forecast Results for Naive Base Method. . . . .	49
5.18	Energy Load Areas Forecast Results for ETS Base Method. . . . .	50
5.19	Energy Load Areas Forecast Results for SARIMA Base Method. . . . .	51
5.20	Energy Load Areas Forecast Results for Forecast Combination Base Method. .	52
5.21	Energy Load Italy Forecast Results for Naive Base Method. . . . .	52
5.22	Energy Load Italy Forecast Results for ETS Base Method. . . . .	53
5.23	Energy Load Italy Forecast Results for SARIMA Base Method. . . . .	54
5.24	Energy Load Italy Forecast Results for Forecast Combination Base Method. .	55
5.25	Sensitivity Analysis of Features Matrix for Zones: Results for Energy Load data.	56
5.26	Sensitivity Analysis of Features Matrix for Areas: Results for Energy Load data.	57
5.27	Sensitivity Analysis of Features Matrix for Italy: Results for Energy Load data.	57
5.28	Sensitivity Analysis of Temporal Frequencies: Results for Energy Load Naive Base Method. . . . .	58

5.29	Sensitivity Analysis of Temporal Frequencies: Results for Energy Load ETS Base Method. . . . .	59
5.30	Sensitivity Analysis of Temporal Frequencies: Results for Energy Load SARIMA Base Method. . . . .	60
5.31	Sensitivity Analysis of Temporal Frequencies: Results for Energy Load Forecast Combination Base Method. . . . .	61
C.1	Citi Bike H <sub>3</sub> Cells Forecast Results of MASE index for Naive Base Method. . . . .	81
C.2	Citi Bike H <sub>3</sub> Cells Forecast Results of MASE index for ETS Base Method. . . . .	82
C.3	Citi Bike H <sub>3</sub> Cells Forecast Results of MASE index for SARIMA Base Method. . . . .	83
C.4	Citi Bike H <sub>3</sub> Cells Forecast Results of MASE index for Forecast Combination Base Method. . . . .	84
C.5	Citi Bike Market Forecast Results of MASE accuracy index for Naive Base Method. . . . .	84
C.6	Citi Bike Market Forecast Results of MASE index for ETS Base Method. . . . .	85
C.7	Citi Bike Market Forecast Results of MASE index for SARIMA Base Method. . . . .	86
C.8	Citi Bike Market Forecast Results of MASE index for Forecast Combination Base Method. . . . .	87
D.1	Energy Load Zones Forecast Results of MASE index for Naive Base Method. . . . .	89
D.2	Energy Load Zones Forecast Results of MASE index for ETS Base Method. . . . .	90
D.3	Energy Load Zones Forecast Results of MASE index for SARIMA Base Method. . . . .	91
D.4	Energy Load Zones Forecast Results of MASE index for Forecast Combination Base Method. . . . .	92
D.5	Energy Load Areas Forecast Results for Naive Base Method. . . . .	92
D.6	Energy Load Areas Forecast Results of MASE index for ETS Base Method. . . . .	93
D.7	Energy Load Areas Forecast Results of MASE index for SARIMA Base Method. . . . .	94
D.8	Energy Load Areas Forecast Results of MASE index for Forecast Combination Base Method. . . . .	95
D.9	Energy Load Italy Forecast Results of MASE index for Naive Base Method. . . . .	95
D.10	Energy Load Italy Forecast Results of MASE index for ETS Base Method. . . . .	96
D.11	Energy Load Italy Forecast Results of MASE index for SARIMA Base Method. . . . .	97
D.12	Energy Load Italy Forecast Results of MASE index for Forecast Combination Base Method. . . . .	98

# **Listing of acronyms**

ARIMA .....	Autoregressive Integrated Moving Average
BIC .....	Bayesian Information Criterion
ETS .....	Exponential Smoothing
HTS .....	Hierarchical Time Series
LightGBM .....	Light Gradient Boosting Machine
MASE .....	Mean Absolute Scaled Error
ML .....	Machine Learning
MW .....	Megawatts
NYC .....	New York City
SARIMA .....	Seasonal Autoregressive Integrated Moving Average
SSE .....	Sum of Squared Errors
WAPE .....	Weighted Absolute Percentage Error
XGBoost .....	Extreme Gradient Boosting



# 1

## Introduction

This chapter is divided into four sections. Section 1.1 introduces fundamental concepts of time series data, forecasting, and hierarchical structures, emphasizing the role of cross-temporal aggregation. Section 1.2 explores forecast reconciliation approaches, detailing both linear and machine learning-based methods and their relevance to hierarchical forecasting. Section 1.3 provides an overview of the datasets used in this study, emphasizing their distinct characteristics, significance, and forecasting challenges. Finally, Section 1.4 presents an outline of the thesis structure to guide the reader through the subsequent chapters.

### 1.1 TIME SERIES, FORECASTING, AND HIERARCHICAL STRUCTURE

Time series forecasting plays a crucial role in various domains, including transportation, energy, and finance, where accurate predictions are essential for decision-making. A time series consists of observations collected sequentially over time, often exhibiting trends, seasonality, and irregular variations [1]. These components introduce challenges, as traditional statistical models often struggle to adapt to dynamic patterns in the data. The increasing availability of high-frequency and multi-level time series data has further emphasized the importance of dependency between different variables or the same variable of different frequency.

Forecast reconciliation is a post forecasting process design to deal with time series organized

in a hierarchy based on either cross-sectional aggregation (e.g., different geographic regions contributing to total demand) or temporal aggregation (e.g., hourly data aggregated into daily and weekly levels). These hierarchies introduce dependencies between levels, requiring forecasts at different granularities to be coherent, meaning that aggregated forecasts should match the sum of their lower-level components [2]. However, independent forecasting at each level often results in incoherent predictions, motivating the need for reconciliation techniques. A recent extension to these frameworks is cross-temporal forecasting, which simultaneously ensures consistency across both dimensions, cross-sectional and temporal, by enforcing coherence in multi-level forecasts across time and aggregation structures [3].

For a comprehensive overview of forecast reconciliation across cross-sectional, temporal, and cross-temporal hierarchies, readers are encouraged to consult [4]. The connection between forecast reconciliation and forecast combination methods is explored in detail by [5, 6]. Recent advancements in hierarchical forecasting are discussed in the work of [7]. Additionally, several recent studies have contributed to advancements in forecast reconciliation [8, 9, 10].

## 1.2 FORECAST RECONCILIATION APPROACHES

Forecast reconciliation techniques address inconsistencies in hierarchical forecasting by adjusting forecasts so that they align with aggregation constraints. Cross-sectional reconciliation corrects inconsistencies across linear constraints (e.g., spatial constraints), while temporal reconciliation ensures coherence across different time scales [11]. The extension to cross-temporal reconciliation integrates both, enabling improved accuracy and obtaining coherence.

Several methodologies have been proposed to achieve reconciliation. One of the foundational works in this area is by [11], who introduced an optimal linear reconciliation framework that minimizes forecast error variance. [12] provided further methodological insights into reconciliation techniques, while [13] unified various reconciliation approaches, including the game-theoretic perspective proposed by [14]. More recently, [15] extended these methods to probabilistic reconciliation, offering a generalized framework for linearly constrained multiple time series.

While these traditional techniques have been widely applied, recent advancements have introduced machine learning based reconciliation methods, which offer increased flexibility in capturing complex hierarchical dependencies. The work of [16] represents a key development in this area, introducing a cross-sectional ML-based reconciliation framework that trains tree-based models (e.g., Random Forest, XGBoost) to revise bottom-level forecasts based on hierar-

chical information. Unlike linear reconciliation, which seeks to minimize reconciliation error, the ML-based approach directly optimizes forecast accuracy while ensuring coherence, leading to significant performance gains in applications such as tourism and retail.

Further extensions of ML-based reconciliation methods have been explored. [17] employed deep learning and LightGBM to refine the bottom-up method of [16], correcting biases in lower-level forecasts to improve upper-level accuracy. [18] introduced a dynamic top-down, bottom-up reconciliation approach using tree-based methods and lasso regression, which proved effective for fast-moving consumer goods and large-scale hierarchical datasets like the M5 competition. Additionally, a growing body of research has investigated neural network-based reconciliation models, with [19] and [20] demonstrating the potential of deep learning techniques for end-to-end forecasting and reconciliation across various datasets.

Given the significance of these challenges, this thesis explores the application of forecast reconciliation in both cross-sectional and temporal settings, comparing traditional approaches with machine learning-based methods.

### 1.3 CONTEXT OF THE DATA

This thesis investigates forecast reconciliation techniques using datasets from two distinct industries: transportation and energy. By employing different datasets, we aim to show the applicability of cross-temporal reconciliation methods across various hierarchical structures and real-world challenges.

The bicycle-sharing dataset derives from the Citi Bike system in New York City and includes half-hourly trip counts across six geographical regions, providing a simpler but still hierarchical cross-sectional structure. Accurate forecasting in this domain is crucial for managing bike availability, redistribution strategies, and infrastructure planning. Key challenges include capturing high-frequency seasonality (e.g., hourly, daily, and weekly patterns), sudden demand spikes due to weather or events, and dependencies between regions.

The energy load dataset captures half-hourly electricity consumption across multiple areas, representing a broader and more complex cross-sectional hierarchy. Forecasting energy demand is vital for ensuring the stability and efficiency of energy grids, enabling operators to plan for demand fluctuations and optimize resource allocation. Challenges include long-term trends, non-stationarities caused by events or policy changes, and the need for coherent aggregation across multiple cross-sectional levels (e.g., zone, area, and national level). The deeper cross-sectional hierarchy of this dataset makes it particularly suitable for testing the scalability of reconciliation

methods.

Inspired by the framework outlined by [3], this thesis applies reconciliation techniques to datasets with distinct features compared to traditional low-frequency datasets often used in forecasting literature, such as those in the M5 competition. Unlike datasets with monthly or yearly frequencies, both the bicycle-sharing and energy load datasets feature high-frequency, streaming data. This high granularity enables detailed decision-making but also introduces challenges like non-stationarities and structural shifts caused by expansions, seasonal patterns, or external events.

By evaluating reconciliation techniques on these datasets, this thesis provides insights into their performance across different industries and hierarchical structures, addressing challenges such as high-frequency forecasting, temporal aggregation, and cross-sectional dependencies.

## I.4 OUTLINES OF THE THESIS

This thesis is structured to provide a comprehensive overview of the research conducted, with a clear progression from foundational concepts to experimental results and conclusions. Chapter 2 outlines the objectives of the study, exploring alternative solution to the machine learning reconciliation approaches proposed by [3]. Chapter 3 provides a detailed exploration of the datasets used and the preprocessing steps applied. Chapter 4 presents the methodology employed in this study. It begins by outlining the forecast setup, detailing the structure and design of the base forecasting and reconciliation process. This chapter delves into traditional time series models, machine learning methods for forecast reconciliation, and the linear reconciliation benchmarks applied to cross-temporal hierarchies. Special emphasis is placed on the reconciliation forecast methods and covariance matrices, as these are critical components in achieving coherent forecasts. The evaluation metrics, including Weighted Absolute Percentage Error (WAPE) and Mean Absolute Scaled Error (MASE), are also described to provide a comprehensive understanding of how forecast accuracy and reconciliation performance are assessed.

Chapter 5 discusses the results and findings of the study, which are categorized by dataset. The first part focuses on the bicycle-sharing data, analyzing the overall forecast performance, sensitivity to feature matrices, and sensitivity to temporal aggregation orders. The second part examines the energy load dataset, applying the same criteria to evaluate the methods' effectiveness. The chapter highlights the comparative performance of different forecasting and reconciliation approaches and explores how varying feature matrices and temporal aggregation orders

influence forecast accuracy across the two datasets. This chapter provides critical insights into the effectiveness and robustness of the proposed methodologies. Finally, Chapter 6 concludes the thesis by summarizing the key outcomes, discussing their implications for practical applications, and identifying potential future research directions to further enhance forecasting and reconciliation methodologies.



# 2

## Objectives

This chapter outlines the primary objectives of this thesis, emphasizing the specific challenges addressed and the innovations introduced. While this study builds on the work of [3], it extends their framework considering a new dataset from a different industry, refine reconciliation methodologies, and address practical limitations. By doing so, this work bridges gaps between linear and ML forecast reconciliation. Accurate and coherent forecasting plays a pivotal role in decision-making processes across industries, from urban transportation to energy management. While the original work laid the groundwork for cross-temporal reconciliation, this thesis aims to adapt and expand these methodologies to datasets with unique characteristics, such as high-frequency time series and complex hierarchical structures. Additionally, following modifications were introduced to perform a fair comparison between linear and ML approaches.

1. *FoReco 1.0:* This study utilized the most recent version of the FoReco package, replacing the older version used in the original work. This update provided access to enhanced functionalities and significantly reduced the computational time required for linear reconciliation. The improved efficiency allowed for a smoother integration of advanced reconciliation methods and ensured more reliable outcomes.

2. *Non-Negativity and Rounding Adjustments to Ensure Consistency:* A limitation identified in the original work was the inconsistent treatment of non-negativity and rounding between machine learning-based and linear reconciliation models. While non-negativity constraints and rounding were applied during the reconciliation process for machine learning models, these adjustments were not extended to linear models. To ensure fairness and consistency across

methodologies, in the reconciliation process with linear models, we applied rounding and non-negativity constraints to the base forecasts. Once the forecasts were reconciled, we applied rounding to the high-frequency bottom-level forecasts and subsequently followed it with cross-temporal bottom-up.

*3. Exploring Covariance Matrix Estimation for Linear Reconciliation Methods:* To explore how different covariance configurations could influence reconciliation performance, we extended the solution proposed in [3] by evaluating additional covariance matrix combinations for heuristic, iterative and optimal approaches. These extensions allowed us to assess the sensitivity of reconciliation outcomes to different covariance structures and determine whether alternative configurations could enhance forecast accuracy and coherence.

*4. Validation errors for Linear and Machine Learning Models:* In the original work, the way errors were handled between linear and machine learning (ML) reconciliation models was different. Specifically, the authors used in-sample residuals derived from the training periods in each rolling window, for linear reconciliation methods. In contrast, for ML-based reconciliation, they split the data into training and validation set. Notably, the ML models demonstrated superior forecasting accuracy under this setup.

To ensure a fair comparison and investigate whether this adjustment could improve the performance of linear reconciliation methods, we modified the framework to use validation errors for estimating the covariance matrix in the linear approach according to [3].

*5. Energy load data with 3 levels of hierarchy:* In addition to the Citi Bike dataset, we extended our analysis to a new domain, energy load forecasting, a crucial area for managing energy demand and resource allocation. This dataset introduces additional complexity in its cross-sectional hierarchy, featuring more aggregation levels beyond those present in the Citi Bike dataset. By applying the entire reconciliation framework to this new domain, we aimed to evaluate the adaptability of both linear and machine learning-based reconciliation methods to a structurally different dataset with a more complex aggregation structure. Ultimately, this extension enhances the generalizability of our findings, providing insights into how reconciliation methods perform across distinct forecasting applications.

# 3

## Data

This chapter is divided into two sections. Section 3.1 focuses on the Citi Bike data, outlining its structure and key features, while Section 3.2 examines the Energy Load Data of Italy. Both sections include visual representations to highlight important patterns and characteristics of the datasets.

### 3.1 BICYCLE SHARING PLATFORM DATA

The Citi Bike dataset provides demand data for bicycle rentals across six H<sub>3</sub> cells in New York City, derived from publicly available records for 2023.<sup>1</sup> The data was processed to create geospatial 30-minute time series, capturing bicycle rental demand from 12:00 a.m. to 11:30 p.m. each day. These H<sub>3</sub> cells, part of Uber's Hexagonal Hierarchical Spatial Index, represent hexagonal geospatial units, ensuring a structured cross-sectional hierarchy.

The temporal hierarchy includes ten aggregation levels: 30 minutes, 1, 1.5, 2, 3, 4, 6, 8, 12, and 24 hours. The cross-sectional hierarchy consists of two levels: the upper level represents New York City as a whole, while the bottom level comprises six individual H<sub>3</sub> cells. This diverse structure facilitates an evaluation of forecast reconciliation methods across both granular and aggregated temporal scales.

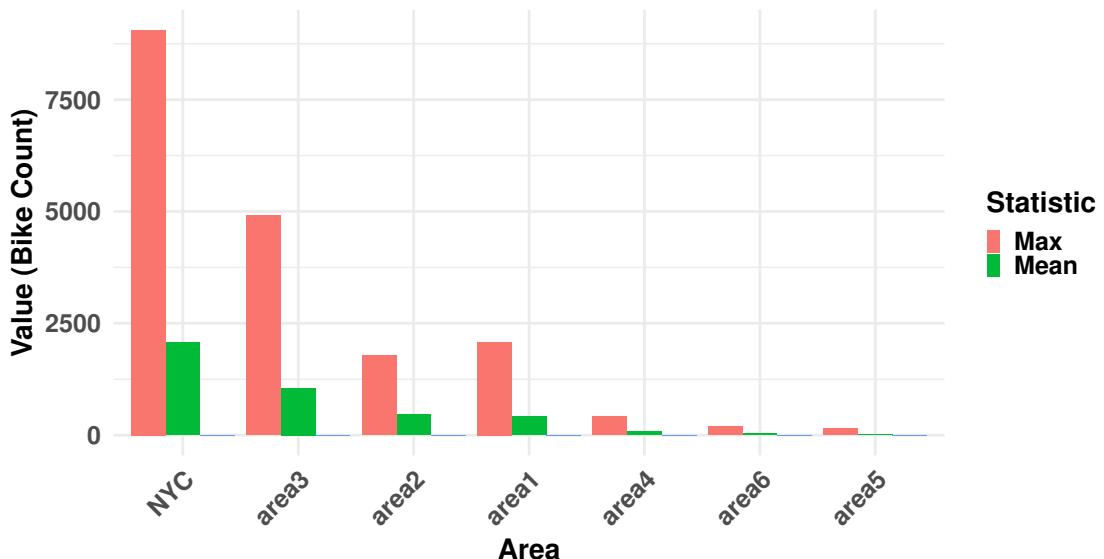
The dataset is well-suited for assessing cross-temporal forecast reconciliation, as it incorpo-

---

<sup>1</sup>The dataset is available at <https://citibikenyc.com/system-data>

rates high-frequency data crucial for decision-making in urban mobility systems. The use of this dataset highlights the challenges of modeling real-world demand with inherent variability and the need for accurate and coherent forecasts across all hierarchical levels.

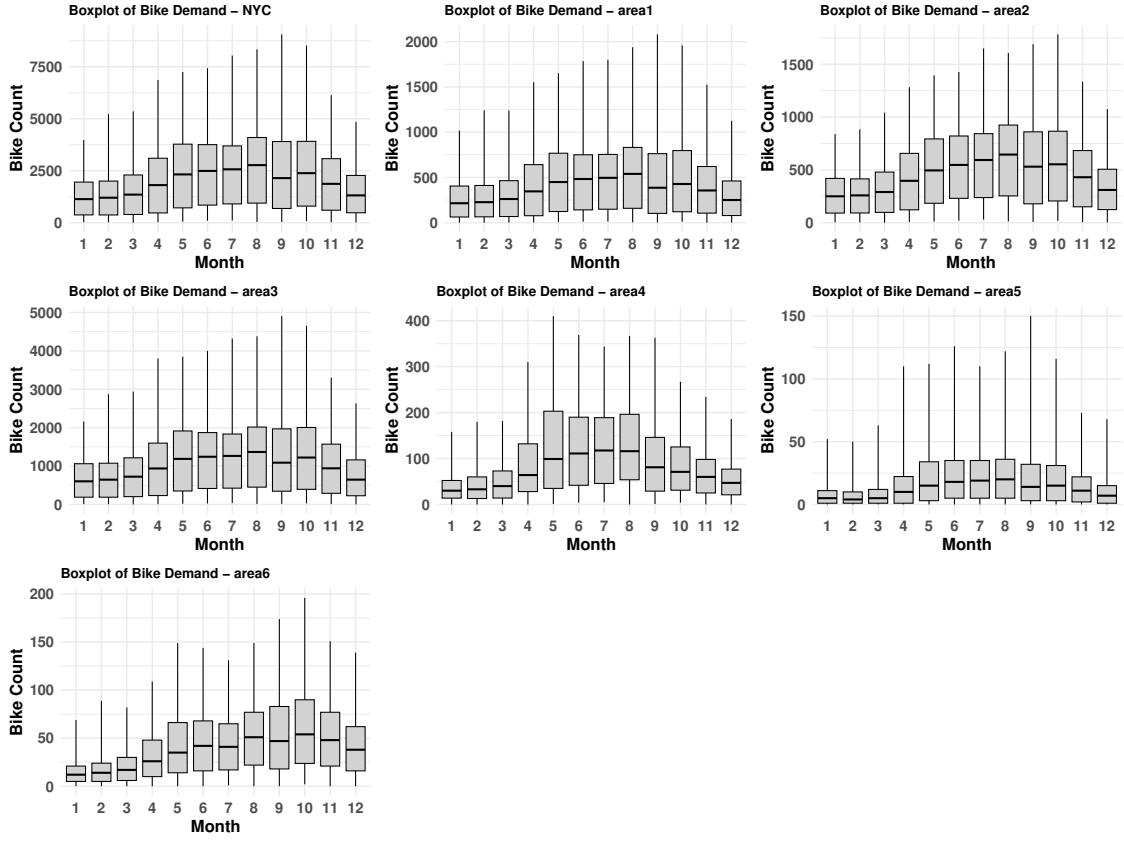
Figure 3.1 presents the summary statistics for Citi Bike usage across six areas in New York City (NYC). The statistics include the minimum, mean, and maximum bike counts for each area, highlighting variations in demand. The minimum values for all areas are zero and thus are not visible in the Figure 3.1. Area 3 and Area 2 exhibit the highest average demand, with maximum values exceeding 4,000 bikes, while Areas 5 and 6 show significantly lower usage levels. These differences underscore the heterogeneous nature of bike demand across regions, which is crucial for effective reconciliation and forecasting. Figure A.1 in Appendix A displays



**Figure 3.1:** Summary statistics for Citi Bike data.

box plots of bike counts for all areas and New York city in the Citi Bike dataset. The plot highlights the distribution of bike counts across regions. NYC, representing the aggregated demand, exhibits the highest range, while individual areas show distinct patterns, reflecting their localized demand dynamics.

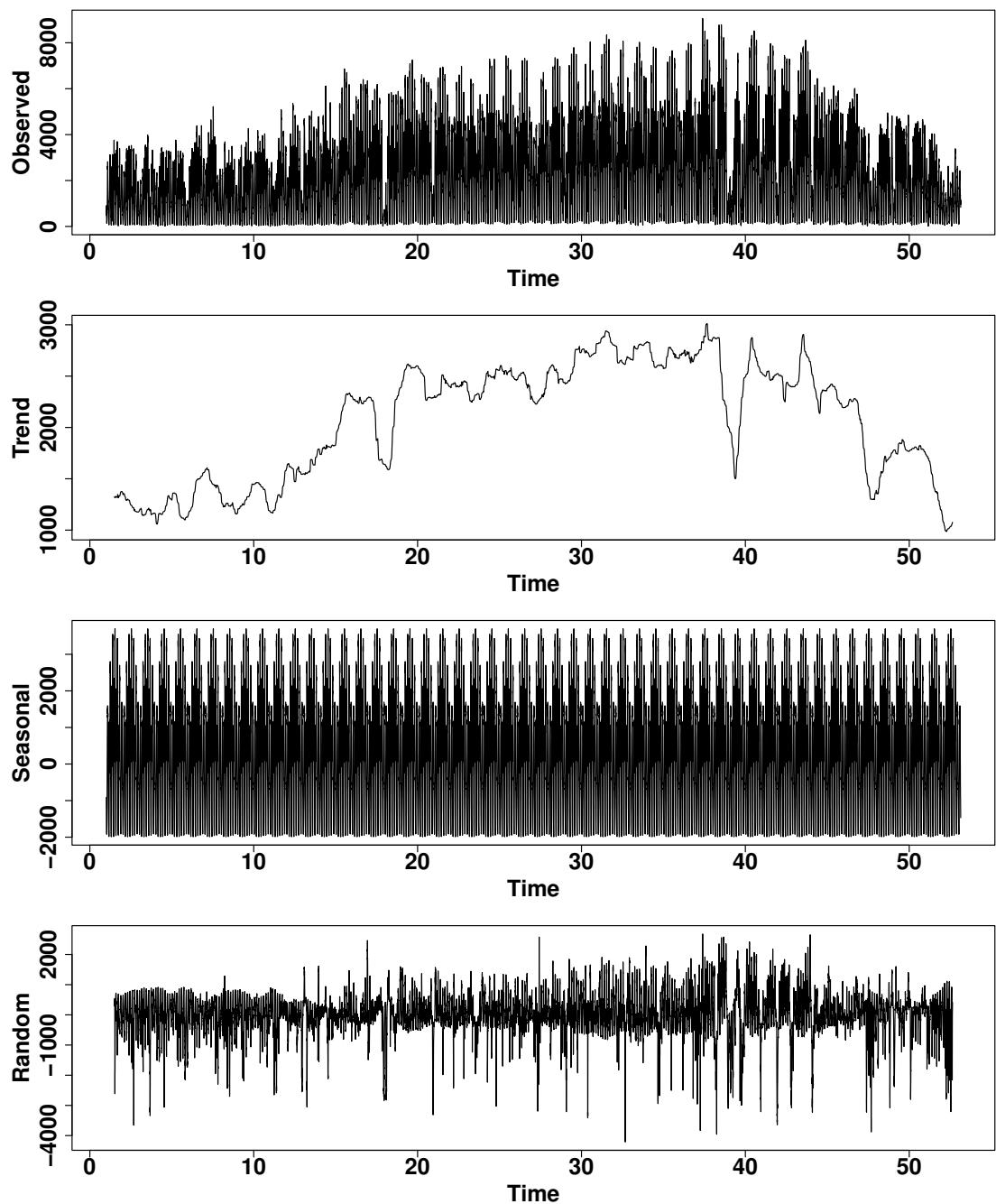
Figure 3.2 shows monthly boxplots of bike counts for each area in the Citi Bike dataset. The plots reveal the distribution of bike demand across months, highlighting seasonal trends and variations. Figure A.2 shows the hourly and daily seasonality of bike demand across different areas in New York City, including NYC as an aggregate (see Appendix A). Each blue line represents the bike demand for a single day, plotted over 24 hours. The transparency of the lines



**Figure 3.2:** Monthly bike demand distribution across different areas.

allows overlapping patterns to be visualized, highlighting variations in bike usage throughout the day and across different days. These plots reveal daily demand trends and seasonal fluctuations for each area.

Figure 3.3 presents the decomposition of Bike demand into trend, seasonal, and residual components. The trend captures long-term variations, showing an initial increase in bike usage followed by a decline toward the end of the observed period. The seasonal component reveals pronounced daily and weekly patterns, with peaks corresponding to typical commuting hours and weekends. The residual component highlights random fluctuations, which may result from weather or special events. This decomposition provides a comprehensive view of the factors driving bike demand in New York City. Figure A.3 in Appendix A, displays the correlation matrix for daily bike demand across different areas in New York City, including the aggregate demand for NYC. Strong positive correlations are observed between most areas, indicating similar daily demand patterns, particularly among central and high-demand regions.

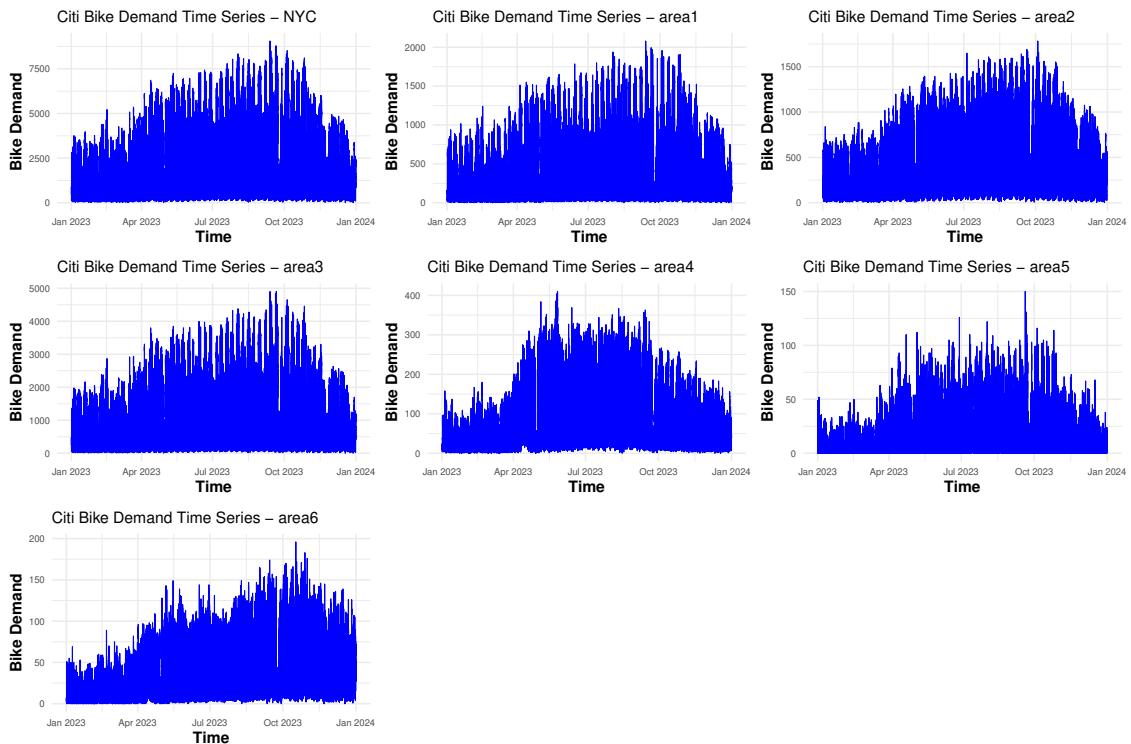


**Figure 3.3:** Decomposition analysis of bike demand load for NYC.

However, correlations tend to be weaker among lower-demand areas, reflecting more localized patterns in bike usage. This matrix highlights the interdependencies between regions, essential

for reconciliation and forecasting.

Figure 3.4 represents bike demand time series over the year 2023 for NYC and six specific areas. NYC, representing aggregate demand, exhibits clear seasonal patterns with peaks during warmer months, reflecting increased activity. The individual areas display distinct variations, with area 3 showing the highest demand among the regions, while areas 5 and 6 have consistently lower demand. These trends highlight the spatial and temporal diversity in bike usage across the city, critical for understanding demand dynamics. Figure A.4 illustrates the daily

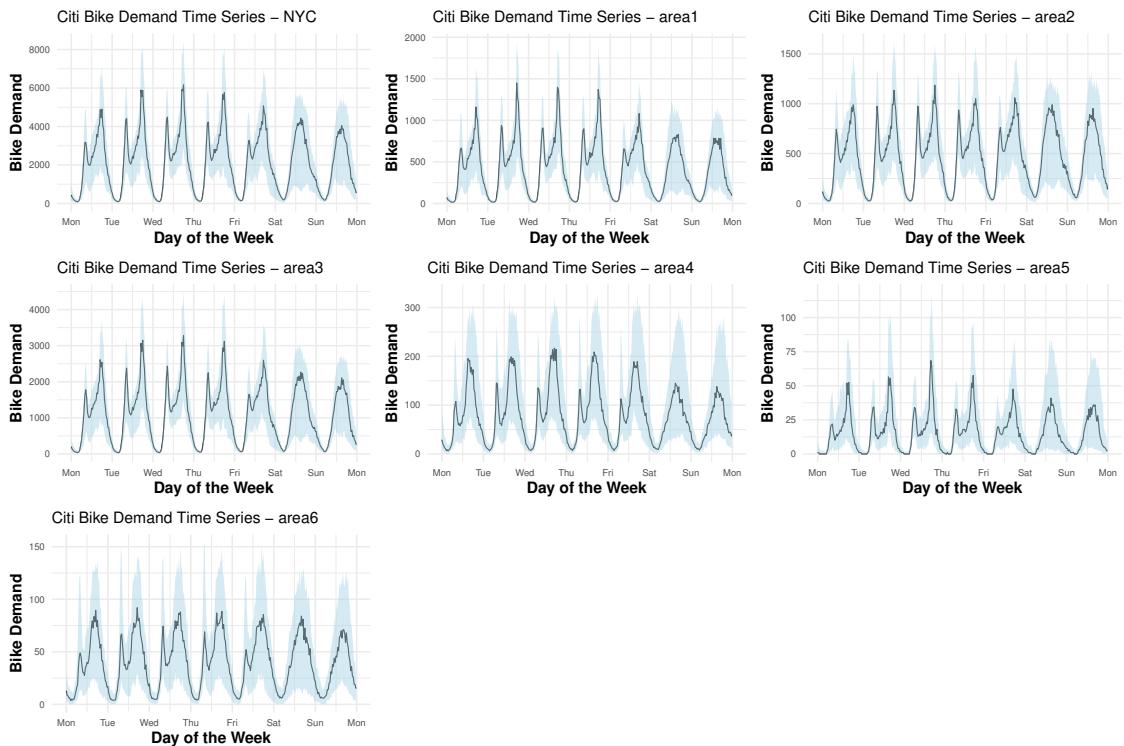


**Figure 3.4:** Bike demand time series over the year 2023: regional trends across NYC.

bike demand patterns across NYC and its six regions from January 1 to January 7 year 2023, as an example for a week demand (see Appendix A). NYC, as the aggregate, shows a peak in demand on weekends, reflecting increased recreational usage. Similarly, area 3 exhibits a high demand pattern, consistent with its overall activity levels, while areas 5 and 6 show significantly lower and more stable demand. The plots capture variations in demand throughout the week, highlighting the influence of both weekday commuting and weekend leisure activities.

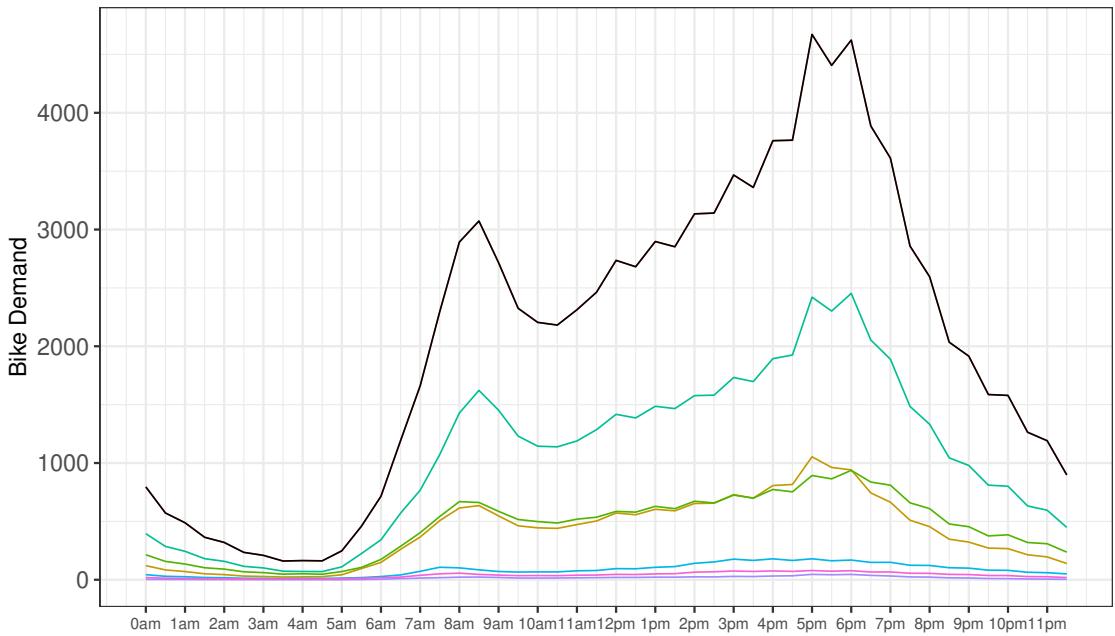
Figure 3.5 depicts the daily time series of bike demand across NYC and its six regions. The solid black line represents the median demand at each 30-minute interval over all weeks, high-

lighting the central tendency of bike usage. The shaded blue area illustrates the variability, spanning the 5th to 95th percentiles, providing a clear indication of the range in which most data points fall. It uses a solid line for the median and shaded areas for variability, providing a clear picture of the central tendency and spread across a week. NYC and areas display clear and regular daily patterns, with consistent peaks during commuting hours across the week. Moreover, lower-demand regions like area 4, area 5, and area 6 exhibit relatively higher variability, as reflected in the wider percentile bands. This suggests that demand in these areas is more erratic or influenced. Figure 3.6 illustrates the average 30-minute bike demand across individual H<sub>3</sub>



**Figure 3.5:** Daily time series plots of Bike demand: median and variability across different regions of NYC.

cells (colored lines) and the entire New York City area (black line). The temporal patterns are analyzed at various levels of aggregation, ranging from 30-minute intervals to 1-day summaries. This decomposition provides insights into both local demand patterns and city-wide trends, highlighting the spatial and temporal diversity in bike usage.



**Figure 3.6:** Bike demand for NYC and areas.

### 3.2 ENERGY LOAD DATA OF ITALY

The energy load dataset provides 30-minute electricity consumption data for various regions of Italy throughout the year 2023. This publicly available dataset can be accessed from Terna's Download Center.<sup>2</sup>

The cross-sectional hierarchy comprises seven zones: Calabria, Center-North, Center-South, North, Sardinia, Sicily, and South. These zones can be aggregated into areas and Italy as a whole, such that:

- Italy: The sum of all seven zones.
- Islands: The sum of Sardinia and Sicily.
- All-North: The sum of Center-North and North.
- All-South: The sum of Center-South, South, and Calabria.

The temporal hierarchy includes ten aggregation levels: 30 minutes, 1, 1.5, 2, 3, 4, 6, 8, 12, and 24 hours. The high-frequency data, combined with its cross-sectional structure, reflects

---

<sup>2</sup>The dataset is available at <https://dati.terna.it/en/download-center#/Load/Total-load>

the complexity of managing energy consumption. The dataset also highlights the importance of accurate forecasting for ensuring grid reliability, managing energy demand, and planning resource allocation.

Similarly to the Citi Bike dataset, we generated a series of plots for the energy load data to analyze its key characteristics and provide insights into its hierarchical structure. Figure 3.7 presents the summary statistics which are maximum, mean, and minimum values for the energy load data.

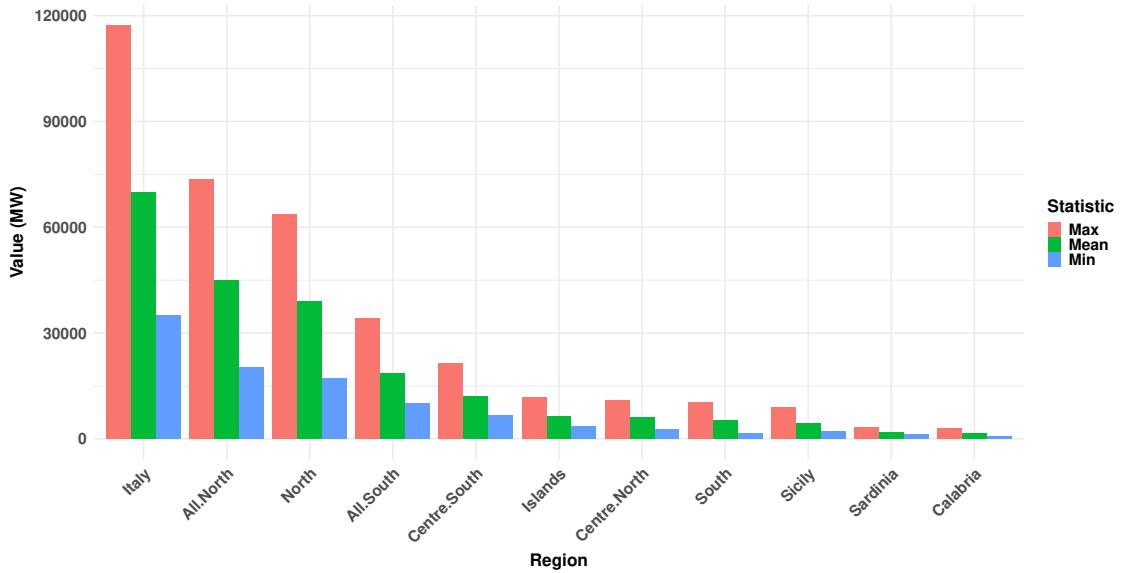
The results reveal significant variations in energy load across different regions. Italy, as a whole, exhibits the highest maximum, mean, and minimum energy load values, as expected. Among the areas, the All-North zone shows consistently higher energy loads compared to the All-South zone, reflecting the higher energy consumption patterns in Northern Italy.

The Islands (Sardinia and Sicily combined) exhibit considerably lower energy loads, indicating reduced energy consumption in these areas. Among individual zones, North and Center-North demonstrate the highest values, while Calabria and Sardinia have the lowest, with minimal differences between maximum, mean, and minimum values.

These results emphasize the geographical disparities in energy consumption patterns within Italy, influenced by factors such as population density, industrial activity, and regional energy needs. The hierarchical aggregation structure, from individual zones to larger areas and the entire country, provides a valuable perspective for assessing the performance of reconciliation methods across different levels of aggregation.

Figure B.1 in Appendix B presents boxplots of energy load data across various zones and areas in Italy. Each boxplot summarizes the distribution of 30-minute energy load data for the year 2023. The plot highlights significant differences in energy load distribution across areas. All-North and North exhibit the next highest medians, reflecting higher energy consumption due to their larger population and industrial activity. Conversely, smaller regions like Calabria, Sardinia, and Sicily show much lower medians and reduced variability, indicative of their smaller populations and less industrialized nature. Notably, the plot emphasizes the disparity in energy load between northern and southern areas. The All-South area, though smaller in scale compared to the northern areas, demonstrates relatively higher variability compared to individual southern zones, suggesting that aggregation can introduce more variability in energy patterns.

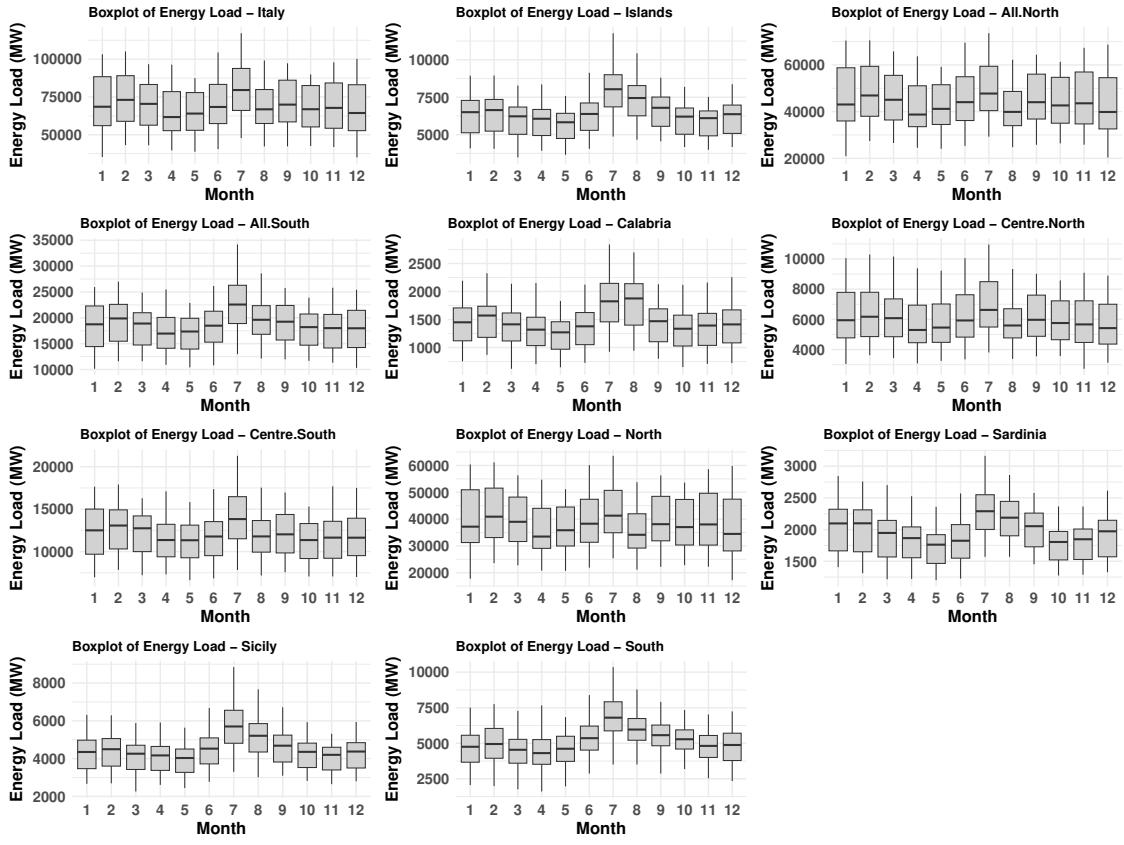
Figure 3.8 presents monthly boxplots of energy load data for different zones and areas in Italy, highlighting seasonal variations throughout 2023. The data reveal significant seasonal trends, with higher energy consumption observed during the summer months (July and August), especially for aggregated areas like Italy and All-North. These peaks likely correspond



**Figure 3.7:** Summary statistics for Energy Load data.

to increased energy demand for cooling. Conversely, the winter months exhibit relatively stable but lower energy consumption across all zones, with slight increases in specific areas such as Center-South and South. Areas with smaller populations, such as Calabria, Sardinia, and Sicily, show more stable energy loads with narrower distributions, reflecting less pronounced seasonal fluctuations. In contrast, larger and industrialized zones like North and Center-North exhibit broader distributions, indicative of greater variability and higher energy demand during peak months.

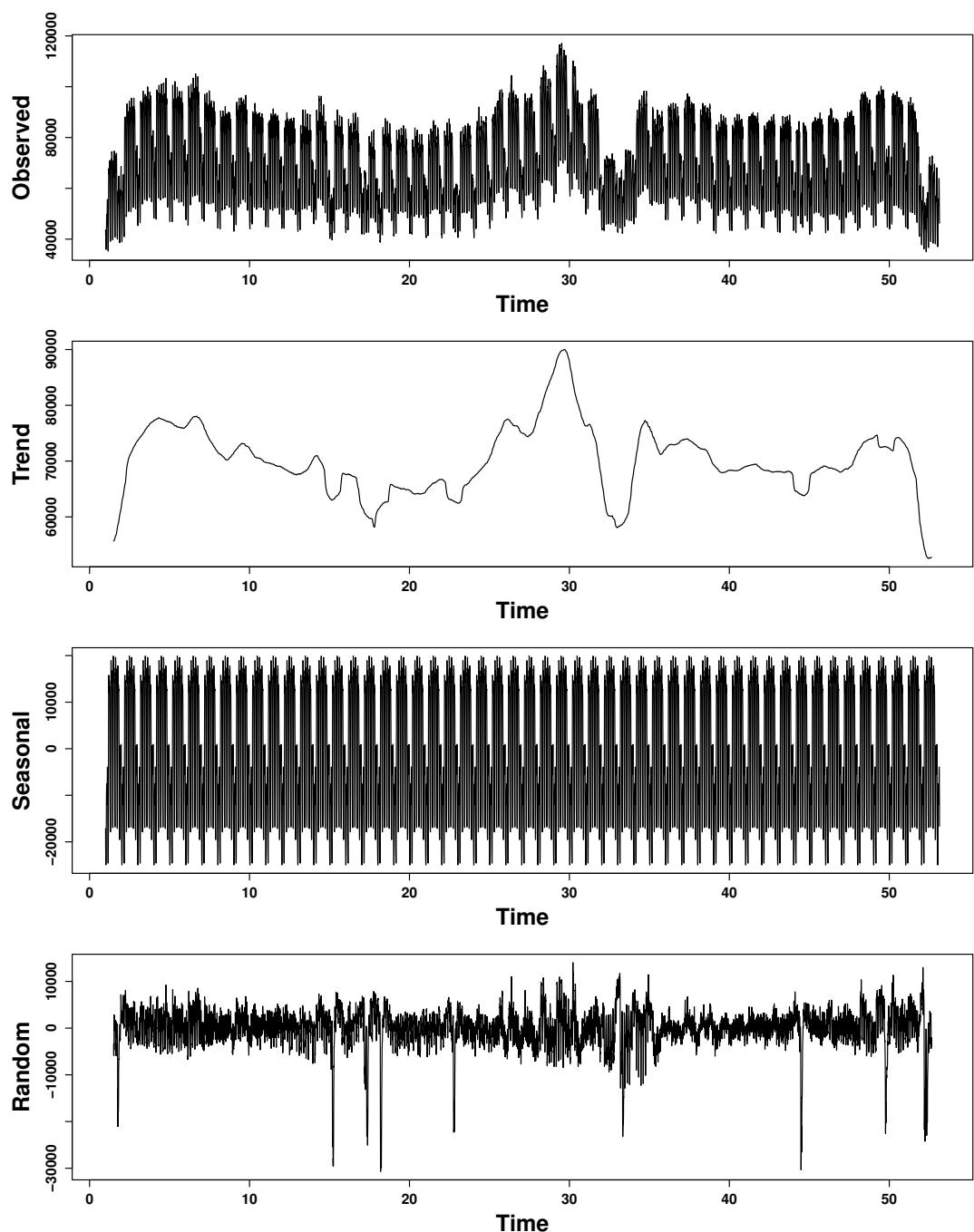
Figure B.2 shows the daily seasonality of energy load across various zones and areas in Italy, including Italy as an aggregate (see Appendix B). Each blue line represents the energy load for a single day, plotted over 24 hours, at 30-minute intervals. The transparency of the lines highlights overlapping patterns, making it easier to visualize trends and variations across different days. The plots reveal distinct energy consumption patterns across all zones and areas. There is a noticeable increase in energy load during the morning hours (approximately 6:00 to 9:00 AM), a plateau or slight dip around midday, and a secondary peak in the evening (approximately 6:00 to 9:00 PM). This pattern is consistent across all variables, reflecting typical daily activities such as work and residential energy usage. Smaller zones, such as Sardinia, Calabria, and Sicily, exhibit much lower energy loads, which is consistent with their smaller populations and lower industrial activity. On the other hand, areas like Center-North and North show higher energy consumption, aligning with their larger economic activity and urbanization.



**Figure 3.8:** Monthly Energy Load distribution across different regions.

Figure 3.9 presents a decomposition of the energy load data for Italy into observed, trend, seasonal, and random components. The trend component demonstrates a gradual increase and subsequent decrease in energy demand, which could be indicative of seasonal shifts in consumption patterns. The seasonal component highlights consistent cyclical patterns, likely corresponding to daily and weekly usage cycles. The random component reveals occasional fluctuations and anomalies, which might be linked to unplanned events or irregularities in energy consumption. This decomposition illustrates the intricate interplay between predictable cycles and irregular variations in energy load, emphasizing the complexity of modeling and forecasting such data.

Figure B.3 in Appendix B displays the correlation matrix of energy loads across various zones, illustrating the strength of linear relationships between regions. The matrix reveals strong positive correlations, as indicated by the dominant red hues, particularly among adjacent or aggregated regions, such as Italy with its subcomponents (North, Center-North, South, and Islands).



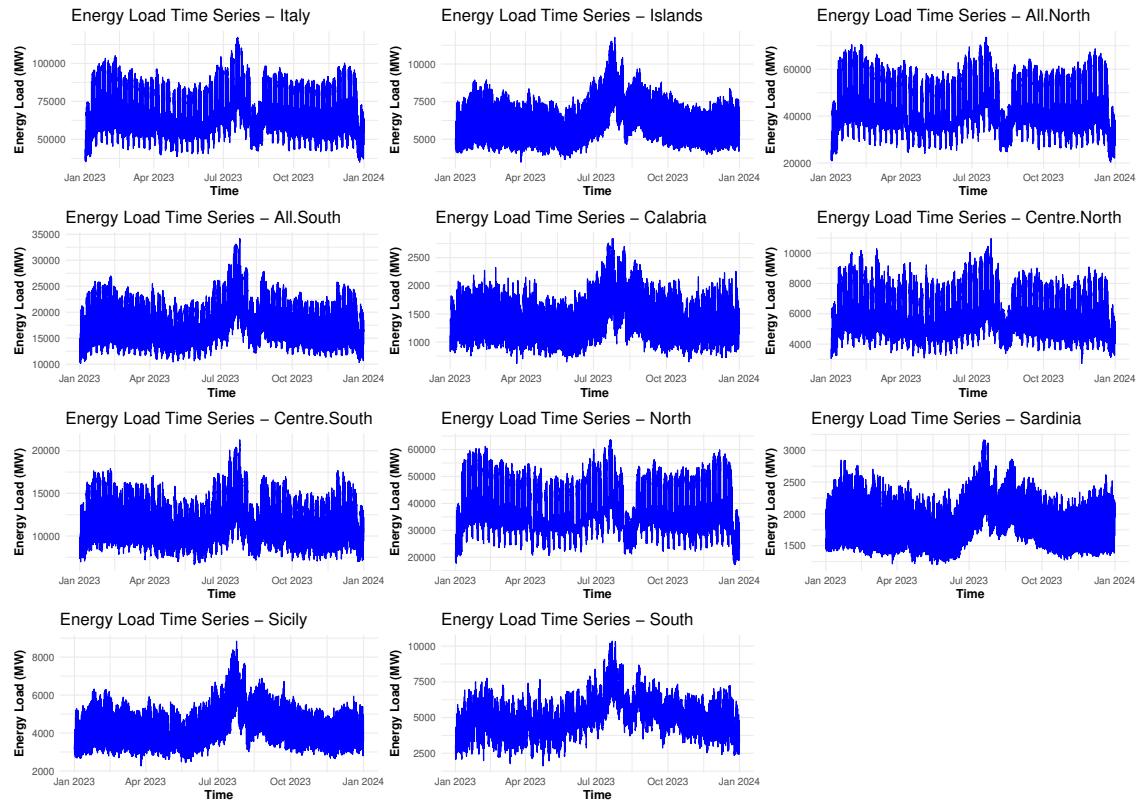
**Figure 3.9:** Decomposition analysis of Energy load for Italy.

High correlations are evident between All-North and North, as well as between All-South, Center-South, and South, reflecting the hierarchical structure of the zones. This consistency

highlights shared consumption patterns or dependencies among regions, which are crucial for cross-sectional reconciliation in hierarchical forecasting. Lower correlations or slight variations in color gradients, such as between Calabria and other regions, suggest localized behaviors or deviations in energy usage.

Figure 3.10 presents the energy load time series for 2023 across Italy and its hierarchical zones, illustrating variations in energy consumption throughout the year. Each subplot corresponds to a specific zone, from the aggregate level (Italy) to smaller zones like Sardinia and Calabria.

The time series reveal consistent energy load levels for most of the year, with notable drops in August and September. This decline could be related to reduced industrial or commercial activities, possibly due to vacation periods or seasonal changes in energy demand patterns. The variability is more significant at higher aggregation levels, such as Italy and All-North, compared to smaller zones like Calabria and Sardinia, where fluctuations are less pronounced.

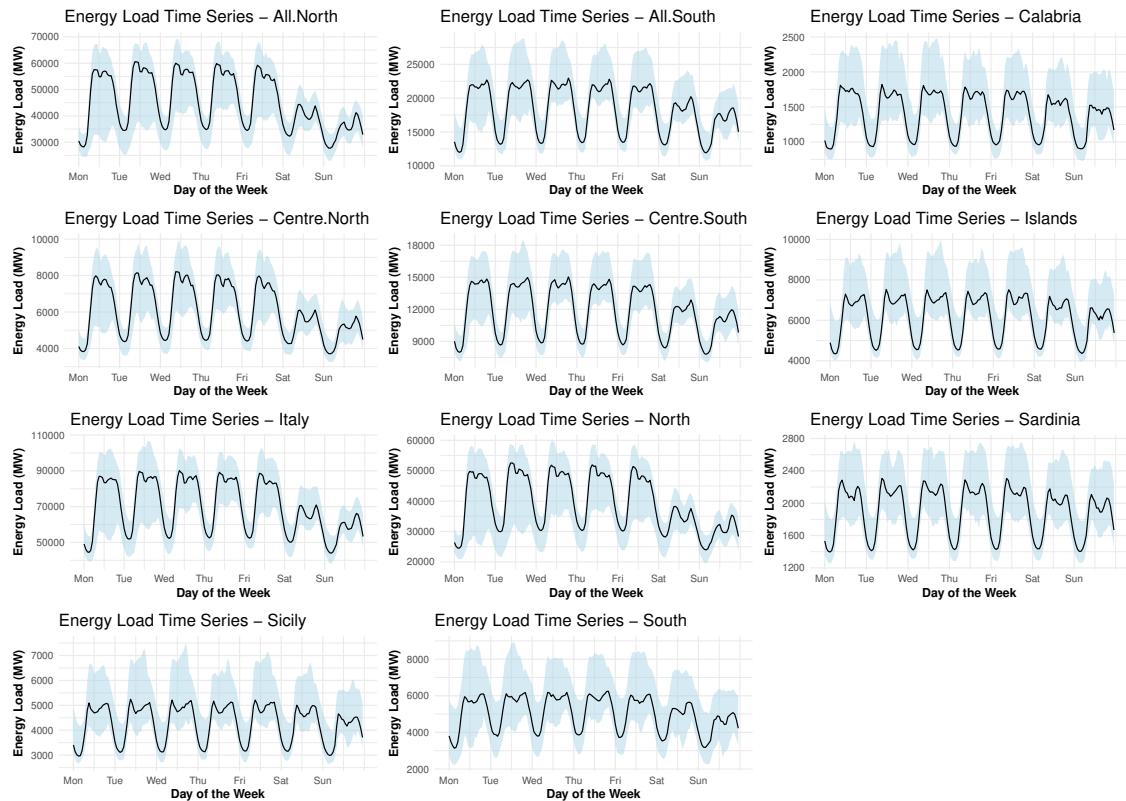


**Figure 3.10:** Energy Load time series over the year 2023: regional trends across Italy.

Figure B.4 in Appendix B illustrates the daily and weekly energy load patterns for Italy and zones for the week of January 1–7, 2023. The x-axis represents the days of the week, while the

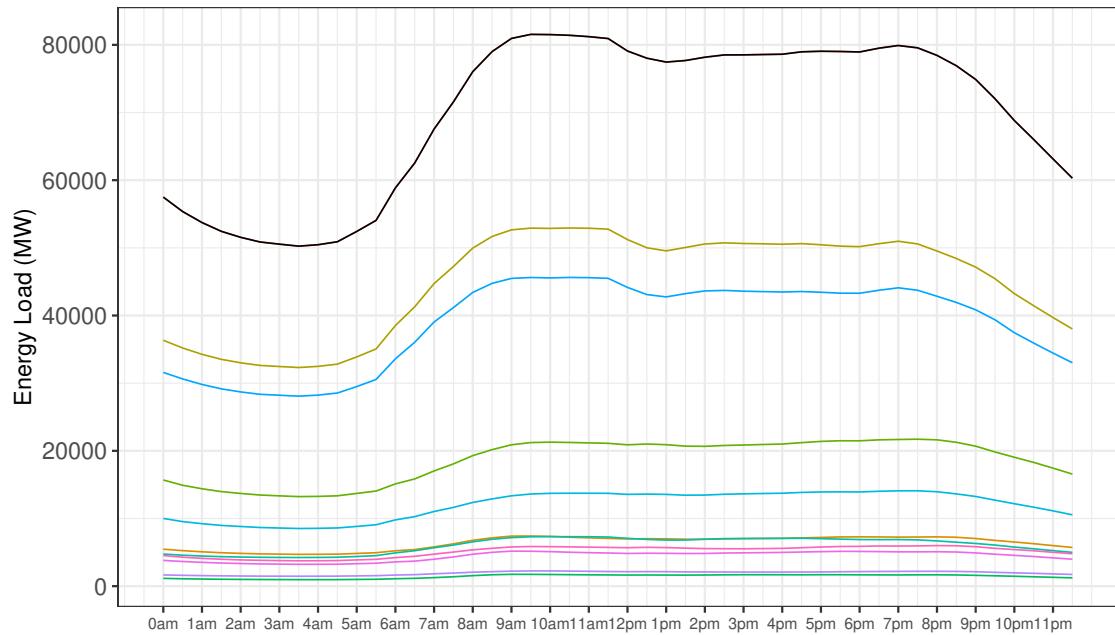
y-axis denotes energy load in megawatts (MW). Across all zones, there is a consistent cyclical pattern indicating higher energy load during weekdays and a dip during the weekend, particularly on Sundays. The Italy plot highlights the national trend, where weekday demand peaks reflect work and industry activity, while weekend troughs likely result from reduced industrial energy consumption.

Figure 3.11 illustrates the daily time series of energy load for Italy and its various zones, with data spanning Sunday to Monday. The solid black line represents the median energy load at each 30-minute interval throughout the week, highlighting the central tendency of energy consumption patterns. The shaded blue region reflects the variability, encompassing the 5th to 95th percentiles, thus capturing the range in which the majority of data points fall. This visualization reveals consistent daily seasonality across all regions. While the overall shape remains similar throughout the week, a drop in intensity is observed during weekends, particularly in high-demand regions.



**Figure 3.11:** Daily time series plots of Energy Load: median and variability across different regions of Italy.

Figure 3.12 illustrates the average 30-minute energy load for Italian zones and aggregate levels. Each colored line represents a specific zone, while the black line represents the overall energy load across Italy. The plot provides insights into the temporal patterns of energy consumption across different levels of aggregation, ranging from individual zones to the national level. This figure highlights a distinct peak in energy load during the morning and early evening hours, with a dip during midday and late night hours. The aggregate energy load for Italy shows a smooth and consistent pattern, whereas individual zones exhibit varying magnitudes of energy usage. This differentiation underscores the regional disparities in energy consumption, with zones like All-North and Italy showing significantly higher loads compared to smaller zones like Calabria and Sardinia.



**Figure 3.12:** Energy Load for Italy and regions.

# 4

## Methodology

This chapter is divided into five sections, each detailing different aspects of the methodology. Section 4.1 describes the forecast setup, including the rolling-window approach and dataset-specific configurations. Section 4.2 focuses on the traditional time series models used for base forecasting. Section 4.3 introduces the machine learning methods employed for forecast reconciliation. Section 4.4 outlines the linear reconciliation benchmarks for cross-temporal hierarchies, including their methods and covariance matrix estimators. Finally, Section 4.5 explains the evaluation metrics used to assess the accuracy and effectiveness of the forecast reconciliation methods.

### 4.1 FORECAST SET-UP

A rolling-window framework is employed to evaluate the forecast performance of the reconciliation method in comparison to its benchmarks. The setup for the two datasets is outlined to provide a clear understanding of the approach.

*Citi Bike Data:* The Citi Bike data forecast evaluation employs a rolling-window approach to compare the performance of the reconciliation methods against their benchmarks. For this setup, the test sample spans June 18, 2023, to December 31, 2023. The rolling-window configuration includes an estimation set of 140 days (Q), a forecast horizon of 1 day (H), and a 28-day validation period that updates continuously within each rolling iteration (R), resulting in a total combined sample size (N) of 168 days. In each iteration of the rolling-window process, 140

days of data are used to estimate the base models, which generate forecasts for both the validation and test periods. The subsequent 28 days serve as the validation set to calculate residuals. These residuals are then leveraged for the reconciliation process:

- Machine Learning Methods use the observed values from the validation set as target values and the base forecasts from the validation set as input features to train the machine learning models.
- Linear Methods utilize validation residuals to directly adjust the test forecasts, applying patterns from the validation period without traditional training.

In other words, the base models are estimated using the most recent 140 days of data, and their performance is evaluated on the test horizon of 1 day for each iteration. The test forecast horizon spans 1 day ( $H=1$ ) per iteration, and the rolling window advances by one day for each subsequent iteration. This approach produces 197 outer rolling windows, ensuring robust evaluation of the reconciliation methods across various temporal settings. This iterative design allows both ML and linear methods to comprehensively leverage patterns from validation residuals, ensuring accurate and consistent forecasts across cross-temporal hierarchies.

*Energy Load Data:* The forecast evaluation for the energy load dataset follows the same rolling-window framework as the Citi Bike dataset. The entire year of 2023 is used as the total dataset, ensuring a direct comparison with the Citi Bike forecasting setup. The same configurations are applied: the training period spans 140 days to estimate the base models, while the validation period covers 28 days, providing forecasts and errors for both machine learning and linear reconciliation methods. The forecast horizon is set to 1 day per iteration, and the rolling window advances by one day in each iteration, resulting in a total of 197 rolling windows for evaluation.

This setup ensures consistency in comparing forecast reconciliation performance across different datasets while considering their distinct cross-sectional structures.

## 4.2 TRADITIONAL TIME SERIES MODELS FOR BASE FORECASTING

We consider four base forecast models: a widely used industry forecasting model, two versatile time series models designed for univariate forecasting, and a forecast combination approach that ensembles the predictions from the first three models. These base forecasts serve as the foundation for subsequent reconciliation and performance evaluation.

*Naive:* The first base forecast model is a straightforward yet widely adopted approach, referred to as ‘Naive’. This model assumes that the forecast for a specific time slot (e.g., 30-min, hourly, or daily) for the next week is equal to the observed value in the same slot and day of the previous week. This approach performs particularly well in datasets with strong seasonal patterns. Its computational efficiency stems from the absence of parameter estimation, making it a preferred choice in many industrial applications [21].

*Exponential Smoothing (ETS):* The ETS model is a popular state-space forecasting approach, utilizing exponential smoothing methods to model trends, seasonality, and errors. Implemented using the `ets` function from the `forecast` package in R, this method automatically selects the optimal error, trend, and seasonality components. For 30-min, 1-hour, 1.5-hour, 2-hour and 3-hour, intra-day seasonality is estimated and for the 4-hour, 6-hour, 8-hour, 12-hour series, weekly seasonality is captured using Fourier series and then ETS is applied to the residual errors for accurate forecasting [22].

*Seasonal Autoregressive Integrated Moving Average (SARIMA):* The SARIMA model extends the ARIMA framework by incorporating seasonality components. For daily data, SARIMA explores the optimal model parameters in the space of seasonal ARIMA models, allowing weekly seasonality ( $m=7$ ) with up to three autoregressive ( $p$ ) and moving average ( $q$ ) terms. Seasonal parameters are restricted to one each ( $P=1, Q=1$ ), and at most one non-seasonal difference ( $d=1$ ) is applied. For higher-frequency data, such as hourly and 30-min series, SARIMA incorporates Fourier series to capture long seasonal patterns. Optimal Fourier terms are selected using the Bayesian Information Criterion (BIC) before modeling residuals with ARIMA [23].

*Forecast Combination:* The Forecast Combination model aggregates the predictions of the Naive, ETS, and SARIMA models by computing their arithmetic mean. This ensemble approach has been shown to enhance forecast accuracy, especially in hierarchical forecasting settings [24]. The combination uses equal weights, chosen for its simplicity and proven effectiveness in prior forecasting studies [25].

## 4.3 MACHINE LEARNING METHODS FOR FORECAST RECONCILIATION

We employ three widely recognized machine learning (ML) models: Random Forest, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). Random Forest and XGBoost have been previously explored in hierarchical time series contexts, such as

in [16], while LightGBM has recently gained attention for its remarkable performance in the M5 forecasting competition [26], which focused on hierarchical retail sales data.

In this study, we utilize the standard implementations of these ML models with their default hyperparameter settings. This approach is twofold: it mitigates excessive computational demands when working with streaming platform data and reflects the practical advantage of off-the-shelf tree-based methods, which often achieve robust performance across diverse applications (e.g., [27]).

For the main results, the models are trained using the sum of squared errors (SSE) as the default loss function. However, the proposed framework is flexible, allowing practitioners to adopt alternative loss functions that may better align with specific characteristics of their data or objectives.

*Random Forest:* Random forests, introduced by [28], are a robust ensemble learning method that combines multiple decision trees to enhance forecasting accuracy. Each tree in a random forest is built using a subset of the data and features, which helps mitigate overfitting and improves generalization. The method uses a bootstrap sampling technique to select data points and a random selection of predictors for each split, thereby ensuring diversity among trees.

In this study, we use the standard implementation provided by the randomForest package [29] in R, employing default hyperparameters: 500 trees (ntree), a third of the total predictors used at each split (mtry), and a minimum node size of 5 (nodesize). These settings balance computational efficiency and accuracy, making random forests particularly suitable for high-dimensional datasets like ours.

*XGBoost:* Extreme Gradient Boosting is an optimized implementation of the gradient boosting framework. Gradient boosting works by sequentially constructing small regression trees, known as weak learners, to minimize the residuals from previous trees. These trees are then combined into a single ensemble to make predictions. XGBoost enhances this process by improving computational efficiency and flexibility.

In this study, we used the xgboost package in R with default hyperparameters to maintain consistency and avoid excessive computational demands. This choice aligns with its established performance across diverse forecasting applications, including hierarchical and time series contexts [30]. The primary loss function used for training XGBoost models is the sum of squared errors, which allows for robust learning without hyperparameter tuning. We fixed the hyperparameters to the default values as follows: 100 boosting iterations (nrounds), 6 as max tree depth (max depth), 0.3 as learning rate (eta), 1 as subsample ratio (subsample), 1 as subsample ratio of columns (colsample bytree), 1 as minimum sum of instance weight (hessian) (min child

weight) and  $\alpha$  as minimum loss reduction (gamma).

*LightGBM*: Light Gradient Boosting Machine, developed by Microsoft, is another gradient boosting algorithm. Like XGBoost, it uses tree-based models to learn residuals and improve predictions. However, LightGBM incorporates specialized features, such as gradient-based one-side sampling and exclusive feature bundling, enabling it to handle large datasets with improved computational efficiency.

We employed the `lightgbm` package in R with its standard configurations: 100 boosting iterations (`nrounds`), 31 as maximum number of leaves (`num leaves`), 0.1 as learning rate (`eta`), 1 as subsample ratio (`subsample`), 1 as subsample ratio of columns (`colsample bytree`), 0.001 as minimum sum of instance weight (`hessian`) (`min child weight`) and 0 as  $\ell_1$ -regularization (`lambda l1`) and no limit to the max depth (`max depth`) for the tree model. This choice ensures ease of implementation while leveraging LightGBM’s proven effectiveness in complex forecasting tasks [31]. As with XGBoost, the loss function used for training is the sum of squared errors, focusing on generating accurate base forecasts without additional tuning.

## 4.4 LINEAR RECONCILIATION BENCHMARKS FOR CROSS-TEMPORAL HIERARCHIES

This section explores various reconciliation techniques and their application in the proposed framework. Emphasis is placed on the integration of cross-temporal and cross-sectional aggregation to maintain consistency across hierarchical levels. The analysis compares our newly proposed machine learning-based forecast reconciliation approach with base forecasts and five state-of-the-art linear reconciliation benchmarks for cross-temporal hierarchies, implemented in the `FoReco` package [32] using R [33].

### 4.4.1 RECONCILIATION FORECAST METHODS

This subsection outlines the forecast reconciliation methods used in this study, including Bottom-Up, `tcs`, `cst`, `ite`, and `oct`. These approaches differ in their reconciliation order and computational strategies, offering distinct advantages depending on the hierarchy and data structure.

*Bottom Up*: Simple aggregation of bottom level forecasts.

*tcs*: Heuristic (first-temporal-then-cross-sectional) cross-temporal forecast reconciliation. It replicates the procedure by [34]: (i) for each time series the forecasts at any temporal aggregation order are reconciled using temporal hierarchies; (ii) time-by-time cross-sectional reconcili-

ation is performed; and (iii) the projection matrices obtained at step (ii) are then averaged and used to cross-sectionally reconcile the forecasts obtained at step (i).

*cst*: Heuristic(first-cross-sectional-then-temporal) cross-temporal forecast reconciliation. The order of application of the two reconciliation steps (temporal first, then cross-sectional), is inverted compared to tcs [32].

*ite*: Iterative cross-temporal forecast reconciliation starting with cross-sectional reconciliation. It performs the iterative procedure described in [32], which produces cross-temporally reconciled forecasts by alternating forecast reconciliation along one single dimension (either cross-sectional or temporal) at each iteration step.

*oct*: Optimal combination cross-temporal forecast reconciliation in a single reconciliation step. It performs optimal (in least squares sense) combination cross-temporal forecast reconciliation. Non-negative and immutable reconciled forecasts can be considered by [35] and [36] respectively.

#### 4.4.2 COVARIANCE MATRICES

To ensure effective reconciliation of forecasts, the choice of covariance matrix estimators plays a critical role. These estimators incorporate information from the residuals of the base forecasts to capture the variability and dependencies within and across the series. Depending on the method and hierarchical structure, different covariance matrix approximations are used to improve the coherence and accuracy of the reconciled forecasts. This subsection outlines key estimators used in hierarchical and cross-temporal reconciliation, based on methods detailed by [32].

*wlsv*: For the tcs, cst, ite, and oct benchmark methods, the temporal reconciliation covariance matrix estimator ‘wlsv’ is the series variance scaling matrix which is the diagonal matrix that contains the estimated variances of the validation errors across each level [37].

*wls*: Used in methods tcs, cst, and ite for the cross-sectional step, the covariance matrix is estimated as a diagonal matrix of the validation error variances. This approach uses the errors to scale the importance of different series [12], ensuring that series with higher variability are weighted less during reconciliation.

*acov*: Applied to tcs, cst, ite, and oct, this method leverages the auto-covariance structure of residuals within each temporal aggregation level. By capturing temporal dependencies, it improves the alignment of forecasts with the inherent autocorrelations in the data. However, acov does not model relationships between different levels of aggregation, focusing solely on

improving reconciliation within each frequency separately. As described in [38], this method is particularly useful when there is enough data to reliably estimate the autocovariance matrix for each temporal level.

*shr*: Used for methods tcs, cst, and ite, the cross-sectional reconciliation covariance matrix estimator is a shrinkage covariance matrix of the errors, where the off-diagonal elements are shrunk towards zero [11].

*sari*: Utilized in tcs, cst, and ite, the sari estimator applies first-order autoregressive modeling to capture temporal dependencies in the errors [38].

*bdsr*: The cross-temporal covariance matrix is estimated using a block-diagonal matrix where each block is the cross-sectional shrinkage covariance of each temporal level [11].

*str*: Primarily used in the oct method, the str covariance matrix is a cross-sectional reconciliation estimator. It relies on structural variances, assuming that the variance at each level of the hierarchy is approximately equal. This method distributes variance according to the hierarchical structure, ensuring that the reconciliation process preserves the relationships between aggregated and disaggregated levels [37].

*Ssbr*: The cross-temporal covariance matrix is estimated using a block diagonal matrix, where each block contains the temporal shrinkage covariance matrix for each temporal level [11].

## 4.5 EVALUATION METRICS

The Weighted Absolute Percentage Error (WAPE) and the Mean Absolute Scaled Error (MASE) are two of the key metrics used to evaluate forecast accuracy. These metrics provide insights into the performance of the models across different scales and time series structures, offering complementary perspectives for forecasting quality.

### 4.5.1 WEIGHTED ABSOLUTE PERCENTAGE ERROR

WAPE is a commonly used metric in forecasting, particularly for evaluating models in platforms dealing with heterogeneous data. WAPE is calculated as the normalized ratio of the total absolute error to the total actual values, enabling a scale-independent measure of forecast accuracy. This property makes it ideal for comparing performance across time series with varying scales or aggregation levels. For a specific temporal aggregation level  $k$  and a time series  $i$ , the WAPE

formula is given by:

$$\text{WAPE}_i^{[k]} = \frac{\sum_{j=1}^{T_{\text{test}}/k} |A_{i,j}^{[k]} - F_{i,j}^{[k]}|}{\sum_{j=1}^{T_{\text{test}}/k} A_{i,j}^{[k]}}, \quad (4.1)$$

where  $A_{i,j}^{[k]}$  and  $F_{i,j}^{[k]}$  denote the actual and forecasted values, respectively, and  $T_{\text{test}}$  is the number of high-frequency observations in the test set. By multiplying WAPE by 100, it can be expressed as a percentage. To summarize performance across multiple series, the individual WAPEs are averaged at each cross-temporal level. For example, the overall WAPE for bottom-level series (e.g., 30-minute time series for  $n_b$  areas) is computed as:

$$\text{WAPE}^{[1]} = \frac{1}{n_b} \sum_{b=1}^{n_b} \text{WAPE}_b^{[1]}, \quad (4.2)$$

where  $\text{WAPE}_b^{[1]}$  represents the WAPE for each bottom-level series. The metric emphasizes periods with higher demand, ensuring that forecasts are evaluated based on their impact on critical business decisions. WAPE is particularly effective for datasets with intermittent demand, as it avoids issues with zero values in the denominator. For further details on its application in demand forecasting, see [39].

#### 4.5.2 MEAN ABSOLUTE SCALED ERROR

As a robustness check, results are also reported for MASE, which is a robust and widely used metric for evaluating forecast accuracy. Unlike WAPE, which assumes stability in the mean over time, MASE accounts for trends, seasonality, and other repetitive patterns in the data. This makes it especially useful for datasets with recurring behaviors, such as hourly or daily demand patterns. MASE is calculated for each temporal aggregation level  $k$  and cross-sectional series as the formula:

$$\text{MASE}_i^{[k]} = \frac{\frac{1}{m_k H} \sum_{j=m_k N+1}^{m_k(N+H)} |A_{i,j}^{[k]} - F_{i,j}^{[k]}|}{\frac{1}{m_k(N-7)} \sum_{j=7m_k+1}^{m_k N} |A_{i,j}^{[k]} - A_{i,j-7m_k}^{[k]}|}, \quad (4.3)$$

where  $A_{i,j}^{[k]}$  and  $F_{i,j}^{[k]}$  represent the actual and forecasted values, respectively. The denominator is based on a “naive” forecast method, which assumes that the forecast for a given time slot matches the actual value observed in the same slot a week earlier.

This scaling ensures that MASE is independent of the magnitude of the data, facilitating comparisons across series with different scales. It provides meaningful insights into forecast

accuracy, even in the presence of seasonal patterns or non-stationary behavior. For a more detailed discussion of MASE and its properties, see [40].



# 5

## Results and discussion

This chapter is divided in two main parts. In Sections 5.1 and 5.2, we present the results for the Citi Bike and Energy load datasets.

### 5.1 RESULTS FOR BICYCLE SHARING DATA

Section 5.1.1 provides an analysis of the overall performance of the forecast reconciliation methods applied to the Citi Bike dataset. Sections 5.1.2 and 5.1.3 explore two sensitivity analyses, focusing on the choice of temporal aggregation orders for all reconciliations methods and the choice of features for the ML models.

#### 5.1.1 OVERALL FORECAST PERFORMANCE

Tables 5.1 to 5.4 summarize the WAPE forecast accuracy for the H<sub>3</sub> cells on the complete temporal hierarchy for the different combinations of base forecasts and reconciliation methods; the results for the New York City market are displayed in Tables 5.5 to 5.8. In contrast to the findings of [3] which the ML-based forecast reconciliation methods generally result in substantial improvements in forecast accuracy compared to the linear reconciliation methods, our results differ. Specifically, linear methods tcs and oct, using validation errors and covariance matrices such as Sshr, acov and acov-shr outperformed and achieved the best results under these settings. Our results also offer a consistent improvement over the different base forecasts. Be-

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	0.2883	0.2675	0.2583	0.2525	0.2434	0.2365	0.2286	0.2199	0.2132	0.1981
	Randomforest	0.2682	0.2530	0.2456	0.2400	0.2317	0.2241	0.2136	0.2014	0.1962	0.1775
	XGBoost	0.2951	0.2775	0.2682	0.2620	0.2513	0.2426	0.2305	0.2157	0.2099	0.1879
	LightGBM	0.2882	0.2723	0.2639	0.2583	0.2484	0.2408	0.2278	0.2154	0.2089	0.1869

**Table 5.1:** Citi Bike H3 Cells Forecast Results for Naive Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks.)

tween the machine learning methods, RandomForest generally delivered the most promising results as find in [3]. It is worth noting that, unlike linear reconciliation methods, ML-based reconciliation methods demonstrated significantly greater stability across different base forecasts. Table 5.4 highlights that we achieved the best WAPEs for H<sub>3</sub> cells when using the forecast combination as base forecasts and reconciling it with tcs (acov-wls). In this case, the forecast accuracy for the highest and lowest temporal frequencies were 0.2391 and 0.1492, respectively. At the market level, the results similarly favor forecast combination as the base method and oct (Sshr) as the reconciliation method. Table 5.8 shows that in this case, the forecast accuracy for the highest and lowest temporal frequencies were 0.1879 and 0.1459, respectively.

Changing the metric to the MASE forecast accuracy index instead of the WAPE, similar conclusions can be drawn from Tables C.1 to C.8 (see Appendix C). The best forecast reconciliation results for each base forecast (Naive, ETS, SARIMA, Forecast Combination) are highlighted in each table.

### 5.1.2 SENSITIVITY TO FEATURE MATRICES

The results for the sensitivity analysis on feature matrices are summarized in Table B.1 of [3]. Since no changes or modifications have been made to this part of the original authors' work, we have opted not to reproduce the table here. However, this subsection is included to maintain the logical flow of the discussion, as we will later apply the same sensitivity analysis to our new dataset, the Energy Load data, in Section 5.2.2.

This sensitivity analysis compares the performance of the compact features matrix and the complete features matrix for ML models. The compact features matrix contains all 10 temporal frequencies for the bottom-level series, while the upper-level series are included only at their highest available frequency. In contrast, the complete features matrix incorporates information across the entire cross-temporal structure, encompassing all hierarchical levels. The results are

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.3699	0.3382	0.3314	0.3246	0.3089	0.2677	0.2482	0.2383	0.2237	0.1552	
	Bottom Up	0.3699	0.3593	0.3542	0.3498	0.3408	0.3305	0.3183	0.3052	0.2958	0.2555	
tcs	<i>wlsv-shr</i>	0.3303	0.3180	0.3120	0.3070	0.2959	0.2815	0.2703	0.2525	0.2370	0.1895	
	<i>sari-shr</i>	0.3155	0.3027	0.2960	0.2904	0.2782	0.2620	0.2479	0.2334	0.2141	0.1572	
	<i>acov-shr</i>	0.2503	0.2371	0.2311	0.2264	0.2200	0.2109	0.2015	0.1932	0.1836	0.1542	
	<i>acov-wls</i>	0.2487	0.2355	0.2295	0.2248	0.2184	0.2095	0.2002	0.1920	0.1826	0.1537	
	<i>sari-wls</i>	0.3157	0.3030	0.2962	0.2907	0.2785	0.2623	0.2481	0.2337	0.2143	0.1573	
cst	<i>wlsv-shr</i>	0.3276	0.3150	0.3088	0.3036	0.2917	0.2770	0.2648	0.2474	0.2309	0.1830	
	<i>sari-shr</i>	0.3158	0.3029	0.2961	0.2905	0.2780	0.2618	0.2470	0.2325	0.2130	0.1559	
	<i>acov-shr</i>	0.2805	0.2677	0.2611	0.2564	0.2472	0.2378	0.2265	0.2166	0.2019	0.1577	
	<i>acov-wls</i>	0.2752	0.2626	0.2560	0.2515	0.2428	0.2339	0.2231	0.2134	0.2000	0.1568	
	<i>sari-wls</i>	0.3166	0.3039	0.2972	0.2917	0.2794	0.2632	0.2490	0.2341	0.2145	0.1576	
ETS	ite	<i>wlsv-shr</i>	0.3272	0.3146	0.3084	0.3032	0.2912	0.2764	0.2641	0.2469	0.2303	0.1826
		<i>sari-shr</i>	0.3148	0.3020	0.2952	0.2896	0.2772	0.2610	0.2463	0.2321	0.2129	0.1559
		<i>acov-wls</i>	0.2493	0.2361	0.2300	0.2254	0.2190	0.2100	0.2006	0.1926	0.1830	0.1539
		<i>sari-wls</i>	0.3157	0.3030	0.2963	0.2907	0.2785	0.2623	0.2481	0.2337	0.2143	0.1573
	oct	<i>wlsv</i>	0.3310	0.3186	0.3127	0.3077	0.2967	0.2823	0.2713	0.2534	0.2380	0.1906
		<i>bdshr</i>	0.3228	0.3099	0.3037	0.2978	0.2856	0.2702	0.2566	0.2415	0.2248	0.1746
		<i>acov</i>	0.2471	0.2335	0.2276	0.2228	0.2164	0.2078	0.1983	0.1902	0.1814	0.1541
		<i>str</i>	0.3239	0.3110	0.3039	0.2977	0.2870	0.2697	0.2590	0.2421	0.2231	0.1712
		<i>Sshr</i>	0.2455	0.2316	0.2252	0.2205	0.2141	0.2064	0.1972	0.1904	0.1817	0.1594
	Randomforest	0.2606	0.2446	0.2367	0.2306	0.2213	0.2115	0.1989	0.1908	0.1827	0.1654	
	XGBoost	0.2822	0.2642	0.2541	0.2475	0.2369	0.2271	0.2125	0.2034	0.1941	0.1738	
	LightGBM	0.2742	0.2580	0.2489	0.2430	0.2328	0.2240	0.2096	0.1994	0.1924	0.1718	

**Table 5.2:** Citi Bike H3 Cells Forecast Results for ETS Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.3467	0.3110	0.3022	0.2982	0.2706	0.2321	0.2110	0.2004	0.1700	0.1547	
	Bottom Up	0.3467	0.3353	0.3295	0.3248	0.3137	0.3019	0.2868	0.2728	0.2583	0.2169	
tcs	<i>wlsv-shr</i>	0.3072	0.2936	0.2866	0.2804	0.2668	0.2502	0.2353	0.2210	0.2040	0.1639	
	<i>sari-shr</i>	0.2917	0.2777	0.2702	0.2639	0.2506	0.2326	0.2175	0.2046	0.1874	0.1514	
	<i>acov-shr</i>	0.2461	0.2314	0.2246	0.2189	0.2122	0.2016	0.1913	0.1825	0.1720	0.1505	
	<i>acov-wls</i>	0.2466	0.2319	0.2251	0.2195	0.2129	0.2022	0.1922	0.1830	0.1727	0.1511	
	<i>sari-wls</i>	0.2916	0.2776	0.2701	0.2638	0.2504	0.2325	0.2174	0.2045	0.1873	0.1515	
cst	<i>wlsv-shr</i>	0.3061	0.2926	0.2854	0.2790	0.2653	0.2486	0.2327	0.2193	0.2021	0.1618	
	<i>sari-shr</i>	0.2915	0.2774	0.2698	0.2634	0.2500	0.2320	0.2166	0.2039	0.1865	0.1504	
	<i>acov-shr</i>	0.2668	0.2529	0.2459	0.2403	0.2311	0.2182	0.2064	0.1969	0.1817	0.1512	
	<i>acov-wls</i>	0.2656	0.2519	0.2449	0.2394	0.2307	0.2187	0.2076	0.1979	0.1830	0.1521	
	<i>sari-wls</i>	0.2921	0.2781	0.2706	0.2644	0.2509	0.2329	0.2179	0.2046	0.1872	0.1515	
SARIMA	ite	<i>wlsv-shr</i>	0.3057	0.2921	0.2850	0.2786	0.2648	0.2481	0.2321	0.2187	0.2015	0.1615
		<i>sari-shr</i>	0.2910	0.2769	0.2693	0.2628	0.2495	0.2317	0.2163	0.2038	0.1865	0.1506
		<i>acov-wls</i>	0.2473	0.2326	0.2257	0.2202	0.2135	0.2027	0.1928	0.1835	0.1731	0.1514
		<i>sari-wls</i>	0.2916	0.2776	0.2701	0.2638	0.2504	0.2325	0.2174	0.2045	0.1873	0.1515
	oct	<i>wlsv</i>	0.3072	0.2937	0.2867	0.2805	0.2669	0.2504	0.2356	0.2212	0.2043	0.1645
		<i>bdshr</i>	0.3036	0.2898	0.2828	0.2759	0.2622	0.2450	0.2286	0.2162	0.1990	0.1591
		<i>acov</i>	0.2483	0.2333	0.2262	0.2206	0.2137	0.2034	0.1926	0.1838	0.1735	0.1523
		<i>str</i>	0.3083	0.2940	0.2853	0.2777	0.2632	0.2441	0.2290	0.2151	0.1973	0.1597
		<i>Sshr</i>	0.2523	0.2374	0.2298	0.2251	0.2176	0.2085	0.1990	0.1907	0.1809	0.1626
	Randomforest	0.2604	0.2445	0.2361	0.2303	0.2201	0.2085	0.1946	0.1865	0.1772	0.1589	
	XGBoost	0.2798	0.2621	0.2514	0.2449	0.2330	0.2230	0.2076	0.1982	0.1902	0.1681	
	LightGBM	0.2713	0.2548	0.2456	0.2395	0.2285	0.2186	0.2035	0.1945	0.1860	0.1642	

**Table 5.3:** Citi Bike H3 Cells Forecast Results for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.2942	0.2660	0.2578	0.2532	0.2364	0.2109	0.2017	0.1911	0.1758	0.1527
	Bottom Up	0.2942	0.2820	0.2759	0.2719	0.2641	0.2544	0.2443	0.2336	0.2253	0.1951
Forecast Combination	tcs <i>wlsv-shr</i>	0.2743	0.2608	0.2540	0.2491	0.2399	0.2282	0.2183	0.2068	0.1952	0.1639
	<i>sari-shr</i>	0.2666	0.2528	0.2458	0.2405	0.2308	0.2186	0.2082	0.1968	0.1839	0.1515
	<i>acov-shr</i>	0.2393	0.2241	0.2172	0.2123	0.2055	0.1976	0.1889	0.1808	0.1725	0.1491
	<i>acov-wls</i>	0.2391	0.2240	0.2170	0.2122	0.2053	0.1974	0.1887	0.1807	0.1724	0.1492
	<i>sari-wls</i>	0.2666	0.2528	0.2458	0.2405	0.2309	0.2186	0.2082	0.1968	0.1838	0.1515
	cst <i>wlsv-shr</i>	0.2736	0.2599	0.2532	0.2482	0.2388	0.2269	0.2167	0.2054	0.1933	0.1623
	<i>sari-shr</i>	0.2660	0.2522	0.2451	0.2399	0.2302	0.2179	0.2073	0.1961	0.1829	0.1508
	<i>acov-shr</i>	0.2489	0.2346	0.2275	0.2228	0.2148	0.2063	0.1970	0.1888	0.1777	0.1493
	<i>acov-wls</i>	0.2482	0.2340	0.2270	0.2223	0.2144	0.2063	0.1974	0.1892	0.1785	0.1499
	<i>sari-wls</i>	0.2665	0.2528	0.2458	0.2405	0.2308	0.2185	0.2082	0.1967	0.1835	0.1516
ite	<i>wlsv-shr</i>	0.2734	0.2598	0.2530	0.2479	0.2387	0.2268	0.2164	0.2052	0.1931	0.1621
	<i>sari-shr</i>	0.2661	0.2523	0.2452	0.2399	0.2302	0.2181	0.2074	0.1963	0.1833	0.1509
	<i>acov-wls</i>	0.2394	0.2242	0.2173	0.2124	0.2056	0.1977	0.1889	0.1809	0.1725	0.1493
	<i>sari-wls</i>	0.2666	0.2528	0.2458	0.2405	0.2308	0.2186	0.2082	0.1968	0.1838	0.1515
oct	<i>wlsv</i>	0.2744	0.2609	0.2542	0.2492	0.2401	0.2284	0.2186	0.2069	0.1953	0.1642
	<i>bdshsr</i>	0.2720	0.2582	0.2514	0.2462	0.2368	0.2248	0.2144	0.2033	0.1914	0.1600
	<i>acov</i>	0.2396	0.2242	0.2171	0.2122	0.2053	0.1974	0.1885	0.1808	0.1724	0.1501
	<i>str</i>	0.2707	0.2568	0.2496	0.2439	0.2340	0.2210	0.2110	0.1993	0.1866	0.1547
	<i>Sshr</i>	0.2409	0.2257	0.2184	0.2134	0.2066	0.1990	0.1910	0.1840	0.1754	0.1583
	Randomforest	0.2521	0.2365	0.2288	0.2237	0.2155	0.2075	0.1963	0.1893	0.1816	0.1646
	XGBoost	0.2755	0.2579	0.2481	0.2418	0.2315	0.2237	0.2111	0.2029	0.1951	0.1759
	LightGBM	0.2692	0.2535	0.2443	0.2388	0.2295	0.2222	0.2094	0.2022	0.1948	0.1751

**Table 5.4:** Citi Bike H3 Cells Forecast Results for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	0.2270	0.2243	0.2226	0.2213	0.2179	0.2150	0.2097	0.2043	0.1999	0.1891
	Randomforest	0.2225	0.2190	0.2166	0.2143	0.2102	0.2057	0.1976	0.1875	0.1832	0.1695
	XGBoost	0.2338	0.2295	0.2263	0.2237	0.2183	0.2133	0.2052	0.1943	0.1885	0.1732
	LightGBM	0.2356	0.2315	0.2287	0.2266	0.2215	0.2176	0.2080	0.1971	0.1914	0.1751

**Table 5.5:** Citi Bike Market Forecast Results for Naive Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.3839	0.3044	0.2978	0.2928	0.2751	0.2386	0.2166	0.2115	0.1954	0.1469	
	Bottom Up	0.3474	0.3446	0.3430	0.3409	0.3328	0.3260	0.3150	0.3031	0.2993	0.2737	
tcs	<i>wlsv-shr</i>	0.2876	0.2835	0.2808	0.2780	0.2679	0.2558	0.2473	0.2272	0.2130	0.1824	
	<i>sari-shr</i>	0.2699	0.2653	0.2616	0.2591	0.2473	0.2328	0.2205	0.2067	0.1873	0.1457	
	<i>acov-shr</i>	0.1983	0.1962	0.1945	0.1927	0.1886	0.1837	0.1763	0.1702	0.1619	0.1445	
	<i>acov-wls</i>	0.1958	0.1939	0.1922	0.1904	0.1862	0.1816	0.1743	0.1682	0.1605	0.1442	
	<i>sari-wls</i>	0.2702	0.2656	0.2620	0.2594	0.2476	0.2333	0.2208	0.2071	0.1876	0.1458	
cst	<i>wlsv-shr</i>	0.2841	0.2795	0.2768	0.2737	0.2629	0.2509	0.2410	0.2214	0.2057	0.1764	
	<i>sari-shr</i>	0.2699	0.2652	0.2616	0.2591	0.2470	0.2329	0.2198	0.2060	0.1866	0.1460	
	<i>acov-shr</i>	0.2250	0.2213	0.2192	0.2172	0.2104	0.2043	0.1958	0.1874	0.1744	0.1461	
	<i>acov-wls</i>	0.2173	0.2140	0.2122	0.2102	0.2042	0.1985	0.1910	0.1823	0.1717	0.1454	
	<i>sari-wls</i>	0.2708	0.2663	0.2628	0.2603	0.2486	0.2344	0.2228	0.2081	0.1885	0.1482	
ETS	ite	<i>wlsv-shr</i>	0.2831	0.2785	0.2757	0.2725	0.2616	0.2494	0.2394	0.2203	0.2044	0.1744
		<i>sari-shr</i>	0.2692	0.2644	0.2607	0.2582	0.2462	0.2318	0.2185	0.2056	0.1863	0.1446
		<i>acov-wls</i>	0.1966	0.1946	0.1930	0.1912	0.1870	0.1824	0.1749	0.1689	0.1609	0.1443
		<i>sari-wls</i>	0.2702	0.2656	0.2620	0.2594	0.2477	0.2333	0.2208	0.2071	0.1876	0.1458
	oct	<i>wlsv</i>	0.2886	0.2845	0.2818	0.2791	0.2691	0.2571	0.2485	0.2285	0.2145	0.1836
		<i>bdshr</i>	0.2789	0.2741	0.2712	0.2673	0.2558	0.2430	0.2317	0.2155	0.1998	0.1668
		<i>acov</i>	0.1945	0.1926	0.1909	0.1890	0.1849	0.1805	0.1732	0.1670	0.1600	0.1455
		<i>str</i>	0.2762	0.2714	0.2680	0.2643	0.2534	0.2379	0.2290	0.2118	0.1930	0.1577
		<i>Sshr</i>	0.1958	0.1936	0.1921	0.1903	0.1868	0.1831	0.1772	0.1705	0.1656	0.1495
	Randomforest		0.2154	0.2104	0.2075	0.2044	0.1979	0.1912	0.1804	0.1724	0.1689	0.1562
	XGBoost		0.2231	0.2180	0.2146	0.2115	0.2048	0.1989	0.1876	0.1790	0.1759	0.1609
	LightGBM		0.2201	0.2153	0.2120	0.2097	0.2026	0.1975	0.1856	0.1779	0.1750	0.1622

**Table 5.6:** Citi Bike Market Forecast Results for ETS Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.2915	0.2796	0.2775	0.2779	0.2451	0.2100	0.1988	0.1909	0.1596	0.1573
	Bottom Up	0.2870	0.2830	0.2802	0.2774	0.2661	0.2541	0.2423	0.2224	0.2097	0.1817
tcs	<i>wlsv-shr</i>	0.2649	0.2604	0.2570	0.2533	0.2406	0.2266	0.2132	0.1993	0.1829	0.1522
	<i>sari-shr</i>	0.2519	0.2472	0.2435	0.2406	0.2283	0.2136	0.2001	0.1893	0.1719	0.1431
	<i>acov-shr</i>	0.1976	0.1956	0.1937	0.1917	0.1872	0.1806	0.1726	0.1662	0.1577	0.1420
	<i>acov-wls</i>	0.1976	0.1956	0.1938	0.1917	0.1873	0.1806	0.1729	0.1664	0.1580	0.1424
	<i>sari-wls</i>	0.2519	0.2471	0.2435	0.2405	0.2282	0.2134	0.2001	0.1892	0.1718	0.1431
cst	<i>wlsv-shr</i>	0.2637	0.2592	0.2557	0.2520	0.2390	0.2250	0.2101	0.1974	0.1807	0.1494
	<i>sari-shr</i>	0.2515	0.2467	0.2430	0.2399	0.2276	0.2134	0.1995	0.1892	0.1716	0.1429
	<i>acov-shr</i>	0.2192	0.2158	0.2133	0.2111	0.2040	0.1950	0.1853	0.1780	0.1645	0.1430
	<i>acov-wls</i>	0.2176	0.2143	0.2120	0.2098	0.2032	0.1947	0.1858	0.1785	0.1652	0.1440
	<i>sari-wls</i>	0.2521	0.2473	0.2437	0.2407	0.2284	0.2140	0.2006	0.1898	0.1723	0.1437
SARIMA											
ite	<i>wlsv-shr</i>	0.2638	0.2592	0.2557	0.2520	0.2391	0.2249	0.2101	0.1974	0.1807	0.1494
	<i>sari-shr</i>	0.2512	0.2465	0.2428	0.2398	0.2273	0.2129	0.1990	0.1886	0.1712	0.1425
	<i>acov-wls</i>	0.1994	0.1972	0.1954	0.1933	0.1889	0.1820	0.1746	0.1676	0.1590	0.1430
	<i>sari-wls</i>	0.2519	0.2471	0.2435	0.2405	0.2282	0.2134	0.2001	0.1892	0.1718	0.1431
oct	<i>wlsv</i>	0.2649	0.2604	0.2570	0.2533	0.2407	0.2266	0.2136	0.1994	0.1832	0.1527
	<i>bdshr</i>	0.2628	0.2581	0.2546	0.2504	0.2378	0.2235	0.2078	0.1964	0.1803	0.1490
	<i>acov</i>	0.1978	0.1957	0.1937	0.1916	0.1871	0.1807	0.1722	0.1661	0.1574	0.1422
	<i>str</i>	0.2570	0.2519	0.2484	0.2443	0.2317	0.2157	0.2031	0.1918	0.1745	0.1456
	<i>Sshr</i>	0.1957	0.1932	0.1914	0.1899	0.1856	0.1808	0.1741	0.1680	0.1609	0.1458
Randomforest											
	Randomforest	0.2143	0.2087	0.2054	0.2023	0.1948	0.1869	0.1752	0.1684	0.1619	0.1491
XGBoost											
	XGBoost	0.2155	0.2099	0.2059	0.2027	0.1959	0.1900	0.1780	0.1705	0.1663	0.1510
LightGBM											
	LightGBM	0.2131	0.2078	0.2038	0.2005	0.1934	0.1873	0.1761	0.1690	0.1640	0.1496

**Table 5.7:** Citi Bike Market Forecast Results for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.2604	0.2339	0.2304	0.2291	0.2113	0.1912	0.1844	0.1768	0.1613	0.1498
	Bottom Up	0.2522	0.2497	0.2474	0.2460	0.2400	0.2333	0.2239	0.2128	0.2076	0.1889
Forecast Combination	tcs <i>wlsv-shr</i>	0.2324	0.2293	0.2269	0.2248	0.2180	0.2087	0.2007	0.1889	0.1783	0.1578
	<i>sari-shr</i>	0.2246	0.2215	0.2189	0.2167	0.2089	0.1992	0.1896	0.1796	0.1670	0.1436
	<i>acov-shr</i>	0.1896	0.1875	0.1858	0.1843	0.1806	0.1760	0.1694	0.1637	0.1575	0.1421
	<i>acov-wls</i>	0.1894	0.1873	0.1855	0.1841	0.1804	0.1758	0.1691	0.1635	0.1575	0.1422
	<i>sari-wls</i>	0.2246	0.2215	0.2189	0.2167	0.2089	0.1992	0.1897	0.1796	0.1669	0.1436
	cst <i>wlsv-shr</i>	0.2313	0.2281	0.2256	0.2235	0.2166	0.2070	0.1983	0.1872	0.1758	0.1557
	<i>sari-shr</i>	0.2239	0.2208	0.2182	0.2160	0.2084	0.1988	0.1893	0.1797	0.1671	0.1441
	<i>acov-shr</i>	0.2014	0.1987	0.1966	0.1950	0.1900	0.1841	0.1771	0.1703	0.1605	0.1415
	<i>acov-wls</i>	0.1999	0.1972	0.1951	0.1937	0.1888	0.1835	0.1767	0.1700	0.1610	0.1420
	<i>sari-wls</i>	0.2244	0.2213	0.2187	0.2166	0.2089	0.1992	0.1902	0.1803	0.1678	0.1448
ite	<i>wlsv-shr</i>	0.2313	0.2281	0.2257	0.2236	0.2167	0.2072	0.1985	0.1873	0.1760	0.1559
	<i>sari-shr</i>	0.2241	0.2210	0.2184	0.2162	0.2084	0.1988	0.1889	0.1792	0.1665	0.1431
	<i>acov-wls</i>	0.1898	0.1877	0.1859	0.1844	0.1807	0.1761	0.1696	0.1638	0.1577	0.1422
	<i>sari-wls</i>	0.2246	0.2215	0.2189	0.2167	0.2089	0.1992	0.1897	0.1796	0.1669	0.1436
	oct <i>wlsv</i>	0.2326	0.2294	0.2270	0.2250	0.2182	0.2089	0.2010	0.1891	0.1785	0.1580
	<i>bdshsr</i>	0.2301	0.2268	0.2245	0.2221	0.2149	0.2054	0.1967	0.1860	0.1747	0.1544
	<i>acov</i>	0.1891	0.1871	0.1853	0.1839	0.1802	0.1762	0.1692	0.1637	0.1581	0.1430
	<i>str</i>	0.2264	0.2231	0.2204	0.2177	0.2102	0.1993	0.1912	0.1809	0.1687	0.1462
	<i>Sshr</i>	0.1879	0.1857	0.1840	0.1827	0.1794	0.1753	0.1702	0.1637	0.1581	0.1459
	Randomforest	0.2064	0.2023	0.1997	0.1979	0.1931	0.1873	0.1785	0.1742	0.1676	0.1580
	XGBoost	0.2135	0.2089	0.2062	0.2039	0.1990	0.1929	0.1842	0.1791	0.1728	0.1621
	LightGBM	0.2123	0.2081	0.2050	0.2038	0.1982	0.1934	0.1834	0.1789	0.1738	0.1619

**Table 5.8: Citi Bike Market Forecast Results for Forecast Combination Base Method.**

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

as follows:

- For the H<sub>3</sub> cells, using the complete features matrix improves forecast accuracy across all base and machine learning forecast methods.
- For the whole market, the results are either virtually unaffected or show a slight deterioration.

These findings highlight that the relative performance of ML-based reconciliation methods when comparing the complete versus compact feature matrices is application-specific.

### 5.1.3 SENSITIVITY TO TEMPORAL AGGREGATION ORDERS

This section evaluates whether there is a gain in forecast accuracy for certain temporal frequencies of interest, specifically 30min, 1h, and 24h, when using all ten temporal frequencies for reconciliation instead of only the three of interest. The analysis is summarized in Tables 5.9 to 5.12, which report the relative WAPE values for the temporal frequencies of interest under these two settings.

A relative WAPE value below one indicates better performance when using the full set of temporal granularities rather than the reduced set. Similar to the findings in [3], the results confirm that forecast performance improves consistently when employing the full set of ten temporal frequencies in the reconciliation procedure. This finding supports the inclusion of intermediate temporal frequencies beyond the main frequencies of interest.

However, unlike the results presented in [3], where ML reconciliation methods showed the largest improvements, the best performance in this analysis was observed for linear methods: tcs and cst when combined with covariance matrices such as acov-shr and acov-wls; ite with acov-wls; and oct with acov and Sshr. These methods demonstrated significant improvements in forecast accuracy across the evaluated temporal granularities.

It worth to note that base forecasts and the bottom-up reconciliation method remain unaffected under the choice of temporal frequencies but are included in the tables for completeness. Although the use of all temporal frequencies introduces additional computational costs, especially for ML-based reconciliation methods, the improvements justify their inclusion for achieving the most accurate forecasts.

Base Method	Forecast Combination	H <sub>3</sub> Cells			Market		
		30min	1h	24h	30min	1h	24h
Naive	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Randomforest	0.9708	0.9750	0.9874	0.9674	0.9739	0.9924
	XGBoost	0.9426	0.9503	0.9217	0.9453	0.9551	0.9538
	LightGBM	0.9409	0.9459	0.9294	0.9529	0.9614	0.9615

**Table 5.9:** Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike Naive Base Method. (Notes: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)

## 5.2 RESULTS FOR ENERGY LOAD DATA

Similarly to Section 5.1, this section examines the performance of forecast reconciliation methods, but in the context of the energy load dataset. Section 5.2.1 presents an overall evaluation of the reconciliation approaches applied to this dataset. Sections 5.2.2 and 5.2.3 focus on sensitivity analyses, investigating the impact of different temporal aggregation orders on reconciliation accuracy and evaluating the influence of feature selection in ML-based methods.

### 5.2.1 OVERALL FORECAST PERFORMANCE

Tables 5.13 to 5.16 present the WAPE forecast accuracy for the energy load dataset at the zone level, which represents the bottom-level series in the cross-sectional hierarchy. The results for the middle level are summarized in Tables 5.17 to 5.20. Finally, the aggregated forecasts for Italy, representing the highest level of the hierarchy, are reported in Tables 5.21 to 5.24.

Similarly to the Citi Bike dataset, the linear reconciliation models outperformed the machine learning reconciliation methods in the energy load dataset. In particular, models like tcs with acov-wls, and with acov-shr, and oct with Sshr demonstrated the best forecasting performances which are quite the same as the ones with best performances for the citi bike dataset. Additionally, among the machine learning methods, again Random Forest consistently delivered the most promising results, reaffirming its robustness in reconciliation tasks across different datasets.

Table 5.15 highlights that we achieved the best WAPEs for zones when reconciling the SARIMA base forecasts with tcs (acov-shr). In this case, the forecast accuracies for the highest and lowest

Base Method	Forecast Combination	H <sub>3</sub> Cells			Market		
		30min	1h	24h	30min	1h	24h
	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Bottom Up	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
tcs	<i>wlsv-shr</i>	0.9801	0.9785	0.9902	0.9502	0.9475	0.9377
	<i>sari-shr</i>	0.9639	0.9616	0.9583	0.9377	0.9345	0.9257
	<i>acov-shr</i>	0.7272	0.7183	0.8210	0.6448	0.6504	0.7909
	<i>acov-wls</i>	0.7196	0.7106	0.8152	0.6315	0.6383	0.7836
	<i>sari-wls</i>	0.9636	0.9613	0.9569	0.9369	0.9337	0.9227
	<i>wlsv-shr</i>	0.9706	0.9682	0.9562	0.9386	0.9340	0.9130
	<i>sari-shr</i>	0.9644	0.9618	0.9498	0.9373	0.9337	0.9258
	<i>acov-shr</i>	0.8261	0.8206	0.8451	0.7416	0.7423	0.8123
	<i>acov-wls</i>	0.8084	0.8027	0.8356	0.7126	0.7147	0.8024
	<i>sari-wls</i>	0.9661	0.9639	0.9576	0.9387	0.9358	0.9355
ETS							
ite	<i>wlsv-shr</i>	0.9705	0.9680	0.9569	0.9339	0.9291	0.8983
	<i>sari-shr</i>	0.9618	0.9591	0.9508	0.9346	0.9307	0.9196
	<i>acov-wls</i>	0.7212	0.7123	0.8162	0.6342	0.6409	0.7842
	<i>sari-wls</i>	0.9636	0.9613	0.9568	0.9368	0.9336	0.9227
oct	<i>wlsv</i>	0.9783	0.9766	0.9863	0.9463	0.9437	0.9306
	<i>bdshr</i>	0.9586	0.9551	0.9380	0.9242	0.9193	0.8841
	<i>acov</i>	0.6898	0.6837	0.8041	0.6163	0.6236	0.7693
	<i>str</i>	0.9121	0.9021	0.7460	0.9146	0.9093	0.7931
	<i>Sshr</i>	0.7192	0.7071	0.8806	0.6362	0.6421	0.8332
Randomforest		0.9218	0.9163	0.9544	0.9003	0.8967	0.9739
XGBoost		0.8984	0.8987	0.8865	0.8911	0.8947	0.9371
LightGBM		0.8930	0.8893	0.8809	0.8840	0.8865	0.9468

**Table 5.10:** Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike ETS Base Method.  
 (Notes: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)

Base Method	Forecast Combination	H <sub>3</sub> Cells			Market		
		30min	1h	24h	30min	1h	24h
	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Bottom Up	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
tcs	<i>wlsv-shr</i>	0.9413	0.9362	0.9342	0.9567	0.9551	0.9609
	<i>sari-shr</i>	0.9138	0.9063	0.9492	0.9192	0.9165	0.9421
	<i>acov-shr</i>	0.7486	0.7368	0.8906	0.7127	0.7190	0.9119
	<i>acov-wls</i>	0.7497	0.7381	0.8915	0.7120	0.7184	0.9092
	<i>sari-wls</i>	0.9138	0.9063	0.9493	0.9192	0.9164	0.9414
cst	<i>wlsv-shr</i>	0.9394	0.9343	0.9265	0.9562	0.9543	0.9537
	<i>sari-shr</i>	0.9138	0.9063	0.9462	0.9190	0.9161	0.9431
	<i>acov-shr</i>	0.8199	0.8115	0.8996	0.7942	0.7960	0.9224
	<i>acov-wls</i>	0.8149	0.8068	0.8981	0.7862	0.7889	0.9185
	<i>sari-wls</i>	0.9150	0.9078	0.9488	0.9195	0.9166	0.9433
SARIMA							
ite	<i>wlsv-shr</i>	0.9386	0.9334	0.9262	0.9558	0.9540	0.9535
	<i>sari-shr</i>	0.9127	0.9050	0.9475	0.9185	0.9158	0.9417
	<i>acov-wls</i>	0.7519	0.7404	0.8930	0.7183	0.7245	0.9127
	<i>sari-wls</i>	0.9138	0.9063	0.9494	0.9192	0.9164	0.9414
oct	<i>wlsv</i>	0.9412	0.9359	0.9331	0.9564	0.9548	0.9563
	<i>bdshr</i>	0.9292	0.9225	0.9089	0.9436	0.9415	0.9506
	<i>acov</i>	0.7273	0.7199	0.8883	0.6933	0.7027	0.9007
	<i>str</i>	0.9013	0.8907	0.8043	0.9355	0.9314	0.9159
	<i>Sshr</i>	0.7705	0.7565	0.9484	0.7044	0.7078	0.9271
Randomforest		0.8958	0.8882	0.9228	0.8556	0.8497	0.9146
XGBoost		0.8704	0.8701	0.8826	0.8337	0.8349	0.9011
LightGBM		0.8681	0.8621	0.8597	0.8262	0.8272	0.8912

**Table 5.11:** Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike SARIMA Base Method.

(Notes: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)

Base Method	Forecast Combination	H <sub>3</sub> Cells			Market		
		30min	1h	24h	30min	1h	24h
tcs	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Bottom Up	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	<i>wlsv-shr</i>	0.9676	0.9641	0.9597	0.9613	0.9585	0.9500
	<i>sari-shr</i>	0.9551	0.9509	0.9603	0.9447	0.9431	0.9384
	<i>acov-shr</i>	0.8403	0.8271	0.8995	0.7799	0.7828	0.8829
	<i>acov-wls</i>	0.8389	0.8255	0.8988	0.7772	0.7805	0.8811
	<i>sari-wls</i>	0.9550	0.9508	0.9598	0.9446	0.9429	0.9368
	cst	0.9659	0.9621	0.9521	0.9600	0.9574	0.9437
	<i>wlsv-shr</i>	0.9540	0.9495	0.9572	0.9433	0.9417	0.9426
	<i>sari-shr</i>	0.8790	0.8700	0.9021	0.8331	0.8339	0.8843
Forecast Combination	<i>acov-wls</i>	0.8748	0.8659	0.9021	0.8237	0.8244	0.8832
	<i>sari-wls</i>	0.9547	0.9506	0.9599	0.9433	0.9417	0.9436
	ite	0.9661	0.9622	0.9522	0.9599	0.9572	0.9429
	<i>wlsv-shr</i>	0.9544	0.9501	0.9582	0.9443	0.9426	0.9373
	<i>sari-wls</i>	0.9550	0.9508	0.9598	0.9445	0.9429	0.9367
oct	<i>wlsv</i>	0.9677	0.9641	0.9586	0.9608	0.9579	0.9463
	<i>bdshr</i>	0.9596	0.9550	0.9531	0.9509	0.9481	0.9496
	<i>acov</i>	0.8238	0.8130	0.8981	0.7655	0.7696	0.8792
	<i>str</i>	0.9457	0.9381	0.8660	0.9470	0.9447	0.8940
	<i>Sshr</i>	0.8470	0.8330	0.9506	0.7723	0.7739	0.8995
Randomforest	Randomforest	0.9586	0.9587	0.9718	0.9336	0.9344	0.9548
	XGBoost	0.9271	0.9290	0.9080	0.9032	0.9057	0.9109
	LightGBM	0.9302	0.9316	0.9116	0.9077	0.9114	0.9034

**Table 5.12:** Sensitivity Analysis of Temporal Frequencies: Results for Citi Bike Forecast Combination Base Method.

(Notes: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	0.0939	0.0931	0.0925	0.0918	0.0907	0.0899	0.0884	0.0872	0.0848	0.0815
	Randomforest	0.0793	0.0783	0.0776	0.0769	0.0757	0.0748	0.0733	0.0722	0.0687	0.0653
	XGBoost	0.0851	0.0836	0.0825	0.0818	0.0802	0.0790	0.0773	0.0760	0.0724	0.0684
	LightGBM	0.0838	0.0826	0.0817	0.0811	0.0796	0.0786	0.0769	0.0756	0.0723	0.0683

**Table 5.13:** Energy Load Zones Forecast Results for Naive Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks.)

temporal frequencies were 0.0532 and 0.0382, respectively. At the area level, the results similarly favor SARIMA as the base method and oct (Sshr) as the reconciliation method. Table 5.19 shows that in this case, forecast accuracies for the highest and lowest temporal frequencies were 0.0395 and 0.0307, respectively. At the Italy level, we again observed the best results with SARIMA as the base model and oct (Sshr) as the reconciliation method, where the forecast accuracies for the highest and lowest temporal frequencies were 0.0370 and 0.0314, respectively. Changing the metric to the MASE forecast accuracy index instead of WAPE, similar conclusions can be drawn from Tables D.1 to D.12 (see Appendix D). However, a difference emerges that is in almost all cases, the best results are obtained with the oct (Sshr) reconciliation method. The best forecast reconciliation results for each base forecast (Naive, ETS, SARIMA, Forecast Combination) are highlighted in each table, providing a clear comparison of model performance across different configurations.

### 5.2.2 SENSITIVITY TO FEATURE MATRICES

This section evaluates the impact of different feature matrices on the performance of machine learning-based reconciliation methods for the energy load dataset. As explained earlier, the compact feature matrix includes all 10 temporal frequencies but focuses exclusively on the bottom-level series (30-minute data), where the ML models are trained. In contrast, the complete feature matrix incorporates information across the entire cross-temporal hierarchy, leveraging data from all hierarchical levels. Tables 5.25 to 5.27 present the results for zones, areas, and Italy, respectively.

At the zone level, using the complete feature matrix generally leads to improved forecasting performance for most reconciliation methods. However, an exception is observed with XGBoost when using the forecast combination as the base model, where the compact feature matrix yields better accuracy.

At the middle level, a similar trend is noted, with most reconciliation methods benefiting

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
Base	Base	0.0880	0.0915	0.0891	0.0870	0.0829	0.0604	0.0591	0.0516	0.0484	0.0414	
Bottom Up	Bottom Up	0.0880	0.0873	0.0866	0.0859	0.0847	0.0835	0.0816	0.0802	0.0761	0.0716	
tcs	<i>wlsv-shr</i>	0.0729	0.0721	0.0712	0.0704	0.0690	0.0672	0.0649	0.0622	0.0578	0.0521	
	<i>sari-shr</i>	0.0651	0.0642	0.0632	0.0621	0.0606	0.0579	0.0560	0.0513	0.0449	0.0377	
	<i>acov-shr</i>	0.0534	0.0526	0.0519	0.0513	0.0501	0.0489	0.0474	0.0453	0.0420	0.0367	
	<i>acov-wls</i>	0.0533	0.0525	0.0518	0.0512	0.0500	0.0488	0.0473	0.0452	0.0420	0.0368	
	<i>sari-wls</i>	0.0652	0.0643	0.0633	0.0623	0.0608	0.0580	0.0563	0.0514	0.0450	0.0378	
cst	<i>wlsv-shr</i>	0.0739	0.0731	0.0723	0.0714	0.0700	0.0683	0.0660	0.0629	0.0584	0.0528	
	<i>sari-shr</i>	0.0658	0.0649	0.0639	0.0629	0.0613	0.0587	0.0568	0.0521	0.0457	0.0387	
	<i>acov-shr</i>	0.0595	0.0586	0.0577	0.0568	0.0553	0.0537	0.0517	0.0490	0.0448	0.0387	
	<i>acov-wls</i>	0.0592	0.0584	0.0575	0.0567	0.0552	0.0539	0.0519	0.0495	0.0456	0.0396	
	<i>sari-wls</i>	0.0661	0.0652	0.0642	0.0632	0.0617	0.0590	0.0572	0.0525	0.0463	0.0392	
ETS	ite	<i>wlsv-shr</i>	0.0723	0.0714	0.0706	0.0697	0.0683	0.0664	0.0642	0.0610	0.0563	0.0508
		<i>sari-shr</i>	0.0649	0.0640	0.0631	0.0620	0.0604	0.0577	0.0558	0.0510	0.0445	0.0374
		<i>acov-wls</i>	0.0535	0.0527	0.0519	0.0513	0.0501	0.0489	0.0473	0.0453	0.0421	0.0369
		<i>sari-wls</i>	0.0652	0.0643	0.0633	0.0623	0.0608	0.0580	0.0562	0.0514	0.0450	0.0378
	oct	<i>wlsv</i>	0.0731	0.0723	0.0715	0.0706	0.0692	0.0674	0.0652	0.0625	0.0582	0.0525
		<i>bdshr</i>	0.0689	0.0681	0.0671	0.0662	0.0647	0.0627	0.0603	0.0570	0.0523	0.0462
		<i>acov</i>	0.0533	0.0525	0.0518	0.0512	0.0500	0.0490	0.0473	0.0456	0.0424	0.0377
		<i>str</i>	0.0710	0.0700	0.0691	0.0681	0.0666	0.0645	0.0623	0.0584	0.0534	0.0471
		<i>Sshr</i>	0.0536	0.0528	0.0520	0.0514	0.0502	0.0492	0.0475	0.0458	0.0430	0.0378
	Randomforest	0.0590	0.0579	0.0568	0.0557	0.0543	0.0525	0.0504	0.0481	0.0445	0.0408	
	XGBoost	0.0643	0.0627	0.0612	0.0601	0.0585	0.0565	0.0541	0.0519	0.0481	0.0440	
	LightGBM	0.0613	0.0601	0.0587	0.0578	0.0562	0.0544	0.0522	0.0499	0.0461	0.0423	

**Table 5.14:** Energy Load Zones Forecast Results for ETS Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.0860	0.0860	0.0858	0.0854	0.0828	0.0542	0.0477	0.0469	0.0435	0.0469	
	Bottom Up	0.0860	0.0853	0.0846	0.0839	0.0827	0.0814	0.0795	0.0776	0.0733	0.0690	
tcs	<i>wlsv-shr</i>	0.0694	0.0685	0.0676	0.0666	0.0654	0.0634	0.0613	0.0583	0.0540	0.0499	
	<i>sari-shr</i>	0.0616	0.0607	0.0597	0.0586	0.0571	0.0544	0.0527	0.0484	0.0429	0.0383	
	<i>acov-shr</i>	0.0532	0.0524	0.0517	0.0511	0.0500	0.0487	0.0473	0.0452	0.0419	0.0382	
	<i>acov-wls</i>	0.0534	0.0526	0.0519	0.0512	0.0502	0.0489	0.0475	0.0454	0.0421	0.0385	
	<i>sari-wls</i>	0.0618	0.0609	0.0599	0.0588	0.0574	0.0546	0.0530	0.0486	0.0432	0.0386	
cst	<i>wlsv-shr</i>	0.0702	0.0693	0.0684	0.0674	0.0661	0.0642	0.0620	0.0591	0.0549	0.0508	
	<i>sari-shr</i>	0.0618	0.0608	0.0598	0.0587	0.0572	0.0546	0.0529	0.0486	0.0434	0.0391	
	<i>acov-shr</i>	0.0571	0.0562	0.0553	0.0544	0.0530	0.0516	0.0497	0.0473	0.0434	0.0392	
	<i>acov-wls</i>	0.0578	0.0569	0.0560	0.0552	0.0538	0.0524	0.0505	0.0482	0.0444	0.0402	
	<i>sari-wls</i>	0.0623	0.0614	0.0604	0.0593	0.0579	0.0552	0.0535	0.0493	0.0443	0.0397	
SARIMA												
	ite	<i>wlsv-shr</i>	0.0686	0.0677	0.0668	0.0658	0.0644	0.0624	0.0602	0.0571	0.0528	0.0486
		<i>sari-shr</i>	0.0613	0.0603	0.0593	0.0582	0.0567	0.0540	0.0523	0.0479	0.0423	0.0377
		<i>acov-wls</i>	0.0535	0.0527	0.0520	0.0513	0.0503	0.0489	0.0476	0.0454	0.0421	0.0385
		<i>sari-wls</i>	0.0618	0.0609	0.0599	0.0588	0.0574	0.0546	0.0530	0.0486	0.0432	0.0386
oct	<i>wlsv</i>	0.0700	0.0692	0.0683	0.0674	0.0661	0.0641	0.0620	0.0591	0.0549	0.0508	
	<i>bdskr</i>	0.0660	0.0651	0.0641	0.0630	0.0615	0.0594	0.0573	0.0540	0.0497	0.0452	
	<i>acov</i>	0.0541	0.0533	0.0526	0.0520	0.0509	0.0496	0.0482	0.0463	0.0430	0.0394	
	<i>str</i>	0.0675	0.0665	0.0655	0.0644	0.0629	0.0608	0.0586	0.0548	0.0503	0.0462	
	<i>Sshr</i>	0.0535	0.0527	0.0519	0.0513	0.0501	0.0491	0.0475	0.0462	0.0430	0.0392	
Randomforest												
	XGBoost	0.0597	0.0585	0.0574	0.0563	0.0549	0.0531	0.0510	0.0488	0.0457	0.0423	
	LightGBM	0.0651	0.0636	0.0622	0.0611	0.0595	0.0576	0.0550	0.0530	0.0495	0.0454	

**Table 5.15:** Energy Load Zones Forecast Results for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.0741	0.0746	0.0730	0.0721	0.0697	0.0557	0.0533	0.0515	0.0486	0.0492
	Bottom Up	0.0741	0.0734	0.0728	0.0721	0.0711	0.0700	0.0683	0.0667	0.0638	0.0608
Forecast Combination	tcs <i>wlsv-shr</i>	0.0671	0.0663	0.0656	0.0649	0.0638	0.0624	0.0606	0.0584	0.0553	0.0522
	<i>sari-shr</i>	0.0620	0.0612	0.0604	0.0597	0.0584	0.0566	0.0552	0.0522	0.0484	0.0448
	<i>acov-shr</i>	0.0577	0.0569	0.0562	0.0556	0.0545	0.0535	0.0520	0.0503	0.0476	0.0444
	<i>acov-wls</i>	0.0575	0.0567	0.0561	0.0555	0.0544	0.0533	0.0518	0.0501	0.0474	0.0442
	<i>sari-wls</i>	0.0622	0.0613	0.0606	0.0599	0.0586	0.0568	0.0554	0.0524	0.0486	0.0450
	cst <i>wlsv-shr</i>	0.0676	0.0668	0.0661	0.0654	0.0642	0.0629	0.0612	0.0589	0.0558	0.0528
	<i>sari-shr</i>	0.0623	0.0615	0.0607	0.0600	0.0587	0.0571	0.0555	0.0527	0.0490	0.0456
	<i>acov-shr</i>	0.0596	0.0588	0.0581	0.0574	0.0562	0.0550	0.0533	0.0513	0.0484	0.0450
	<i>acov-wls</i>	0.0597	0.0589	0.0582	0.0576	0.0564	0.0553	0.0535	0.0516	0.0487	0.0453
	<i>sari-wls</i>	0.0627	0.0618	0.0611	0.0604	0.0591	0.0574	0.0559	0.0532	0.0495	0.0460
ite	<i>wlsv-shr</i>	0.0668	0.0660	0.0653	0.0646	0.0634	0.0620	0.0603	0.0580	0.0548	0.0517
	<i>sari-shr</i>	0.0618	0.0609	0.0602	0.0595	0.0582	0.0564	0.0549	0.0519	0.0480	0.0445
	<i>acov-wls</i>	0.0575	0.0567	0.0561	0.0555	0.0544	0.0533	0.0518	0.0500	0.0474	0.0442
	<i>sari-wls</i>	0.0622	0.0613	0.0606	0.0599	0.0586	0.0568	0.0554	0.0524	0.0486	0.0450
oct	<i>wlsv</i>	0.0671	0.0663	0.0656	0.0649	0.0638	0.0624	0.0607	0.0585	0.0553	0.0522
	<i>bdsbr</i>	0.0652	0.0644	0.0636	0.0629	0.0617	0.0603	0.0584	0.0561	0.0529	0.0495
	<i>acov</i>	0.0576	0.0568	0.0562	0.0556	0.0545	0.0535	0.0520	0.0503	0.0478	0.0447
	<i>str</i>	0.0642	0.0633	0.0626	0.0619	0.0606	0.0591	0.0574	0.0547	0.0513	0.0480
	<i>Sshbr</i>	0.0552	0.0544	0.0538	0.0532	0.0520	0.0511	0.0495	0.0482	0.0456	0.0423
	Randomforest	0.0603	0.0593	0.0583	0.0575	0.0564	0.0549	0.0533	0.0514	0.0483	0.0452
	XGBoost	0.0643	0.0628	0.0616	0.0608	0.0592	0.0577	0.0557	0.0538	0.0502	0.0464
	LightGBM	0.0625	0.0613	0.0602	0.0594	0.0580	0.0567	0.0548	0.0530	0.0497	0.0462

**Table 5.16:** Energy Load Zones Forecast Results for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Forecast Combinstion are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	0.0776	0.0772	0.0768	0.0765	0.0760	0.0755	0.0749	0.0744	0.0732	0.0716
	Randomforest	0.0679	0.0673	0.0668	0.0664	0.0657	0.0650	0.0640	0.0636	0.0614	0.0592
	XGBoost	0.0720	0.0711	0.0704	0.0699	0.0691	0.0682	0.0673	0.0665	0.0641	0.0618
	LightGBM	0.0713	0.0706	0.0699	0.0696	0.0688	0.0681	0.0670	0.0664	0.0641	0.0616

**Table 5.17:** Energy Load Areas Forecast Results for Naive Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.0822	0.0861	0.0838	0.0783	0.0766	0.0469	0.0490	0.0404	0.0366	0.0335	
	Bottom Up	0.0815	0.0811	0.0806	0.0802	0.0794	0.0786	0.0771	0.0761	0.0733	0.0703	
tcs	<i>wlsv-shr</i>	0.0633	0.0627	0.0621	0.0615	0.0605	0.0588	0.0572	0.0544	0.0511	0.0474	
	<i>sar1-shr</i>	0.0545	0.0538	0.0532	0.0523	0.0512	0.0480	0.0469	0.0414	0.0359	0.0304	
	<i>acov-shr</i>	0.0411	0.0407	0.0402	0.0398	0.0390	0.0383	0.0372	0.0356	0.0334	0.0294	
	<i>acov-wls</i>	0.0409	0.0404	0.0400	0.0395	0.0388	0.0380	0.0370	0.0354	0.0333	0.0295	
	<i>sar1-wls</i>	0.0547	0.0540	0.0534	0.0525	0.0514	0.0482	0.0472	0.0416	0.0360	0.0305	
cst	<i>wlsv-shr</i>	0.0657	0.0652	0.0647	0.0640	0.0630	0.0614	0.0598	0.0569	0.0536	0.0500	
	<i>sar1-shr</i>	0.0557	0.0550	0.0544	0.0536	0.0524	0.0495	0.0486	0.0430	0.0379	0.0325	
	<i>acov-shr</i>	0.0480	0.0474	0.0469	0.0462	0.0451	0.0437	0.0424	0.0396	0.0370	0.0323	
	<i>acov-wls</i>	0.0476	0.0471	0.0466	0.0460	0.0449	0.0438	0.0425	0.0401	0.0375	0.0331	
	<i>sar1-wls</i>	0.0559	0.0553	0.0546	0.0538	0.0527	0.0496	0.0488	0.0433	0.0382	0.0328	
ETS	ite	<i>wlsv-shr</i>	0.0629	0.0623	0.0617	0.0611	0.0600	0.0581	0.0568	0.0534	0.0499	0.0464
		<i>sar1-shr</i>	0.0544	0.0538	0.0531	0.0522	0.0511	0.0480	0.0470	0.0413	0.0357	0.0304
		<i>acov-wls</i>	0.0410	0.0405	0.0401	0.0396	0.0389	0.0381	0.0370	0.0355	0.0334	0.0296
		<i>sar1-wls</i>	0.0547	0.0540	0.0534	0.0525	0.0514	0.0482	0.0472	0.0416	0.0360	0.0305
	oct	<i>wlsv</i>	0.0634	0.0629	0.0623	0.0617	0.0606	0.0590	0.0574	0.0546	0.0514	0.0476
		<i>bdschr</i>	0.0591	0.0585	0.0579	0.0571	0.0560	0.0539	0.0522	0.0487	0.0454	0.0408
		<i>acov</i>	0.0410	0.0405	0.0401	0.0397	0.0390	0.0383	0.0372	0.0359	0.0338	0.0300
		<i>str</i>	0.0613	0.0607	0.0600	0.0593	0.0582	0.0560	0.0545	0.0504	0.0468	0.0421
		<i>Sshsr</i>	0.0406	0.0401	0.0396	0.0392	0.0385	0.0379	0.0367	0.0357	0.0339	0.0299
	Randomforest	0.0465	0.0457	0.0448	0.0440	0.0430	0.0415	0.0396	0.0376	0.0358	0.0334	
	XGBoost	0.0498	0.0487	0.0476	0.0467	0.0457	0.0441	0.0422	0.0406	0.0380	0.0352	
	LightGBM	0.0474	0.0464	0.0454	0.0446	0.0435	0.0421	0.0404	0.0387	0.0366	0.0337	

**Table 5.18:** Energy Load Areas Forecast Results for ETS Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.0807	0.0803	0.0814	0.0812	0.0790	0.0423	0.0355	0.0361	0.0332	0.0369	
	Bottom Up	0.0803	0.0799	0.0795	0.0791	0.0783	0.0772	0.0758	0.0746	0.0718	0.0689	
tcs	<i>wlsv-shr</i>	0.0604	0.0599	0.0592	0.0585	0.0576	0.0557	0.0540	0.0513	0.0484	0.0459	
	<i>sari-shr</i>	0.0512	0.0506	0.0499	0.0491	0.0479	0.0449	0.0437	0.0388	0.0344	0.0313	
	<i>acov-shr</i>	0.0413	0.0408	0.0404	0.0400	0.0393	0.0384	0.0374	0.0358	0.0339	0.0313	
	<i>acov-wls</i>	0.0413	0.0408	0.0404	0.0400	0.0393	0.0384	0.0374	0.0358	0.0339	0.0316	
	<i>sari-wls</i>	0.0513	0.0506	0.0499	0.0491	0.0480	0.0449	0.0438	0.0389	0.0345	0.0316	
cst	<i>wlsv-shr</i>	0.0624	0.0618	0.0611	0.0605	0.0594	0.0577	0.0559	0.0533	0.0505	0.0480	
	<i>sari-shr</i>	0.0519	0.0513	0.0506	0.0497	0.0486	0.0457	0.0445	0.0396	0.0355	0.0329	
	<i>acov-shr</i>	0.0461	0.0454	0.0448	0.0442	0.0431	0.0419	0.0404	0.0381	0.0356	0.0329	
	<i>acov-wls</i>	0.0467	0.0461	0.0455	0.0449	0.0439	0.0427	0.0412	0.0391	0.0366	0.0339	
	<i>sari-wls</i>	0.0524	0.0518	0.0511	0.0502	0.0492	0.0463	0.0451	0.0403	0.0364	0.0337	
SARIMA	ite	<i>wlsv-shr</i>	0.0596	0.0590	0.0583	0.0576	0.0565	0.0546	0.0528	0.0500	0.0470	0.0445
		<i>sari-shr</i>	0.0508	0.0502	0.0494	0.0486	0.0474	0.0444	0.0432	0.0382	0.0336	0.0307
		<i>acov-wls</i>	0.0414	0.0409	0.0405	0.0401	0.0395	0.0385	0.0375	0.0359	0.0339	0.0315
		<i>sari-wls</i>	0.0512	0.0506	0.0499	0.0491	0.0480	0.0449	0.0438	0.0389	0.0345	0.0315
	oct	<i>wlsv</i>	0.0610	0.0605	0.0598	0.0592	0.0582	0.0564	0.0547	0.0520	0.0492	0.0468
		<i>bdsbr</i>	0.0568	0.0561	0.0554	0.0546	0.0535	0.0513	0.0497	0.0464	0.0435	0.0406
		<i>acov</i>	0.0417	0.0412	0.0408	0.0404	0.0397	0.0388	0.0379	0.0364	0.0344	0.0320
		<i>str</i>	0.0590	0.0584	0.0577	0.0569	0.0558	0.0537	0.0519	0.0481	0.0452	0.0427
		<i>Sshr</i>	0.0395	0.0390	0.0385	0.0381	0.0374	0.0366	0.0357	0.0347	0.0329	0.0307
	Randomforest	0.0470	0.0462	0.0453	0.0444	0.0433	0.0417	0.0397	0.0380	0.0364	0.0340	
	XGBoost	0.0500	0.0490	0.0479	0.0472	0.0459	0.0442	0.0420	0.0404	0.0384	0.0361	
	LightGBM	0.0475	0.0465	0.0455	0.0448	0.0436	0.0422	0.0403	0.0385	0.0368	0.0343	

**Table 5.19:** Energy Load Areas Forecast Results for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.0670	0.0678	0.0664	0.0647	0.0634	0.0443	0.0432	0.0414	0.0390	0.0409
	Bottom Up	0.0667	0.0663	0.0660	0.0656	0.0650	0.0643	0.0630	0.0617	0.0600	0.0584
Forecast Combination	tcs <i>wlsv-shr</i>	0.0578	0.0574	0.0569	0.0565	0.0557	0.0547	0.0534	0.0511	0.0492	0.0475
	<i>sari-shr</i>	0.0515	0.0510	0.0505	0.0500	0.0491	0.0473	0.0466	0.0433	0.0403	0.0382
	<i>acov-shr</i>	0.0459	0.0455	0.0451	0.0448	0.0441	0.0434	0.0425	0.0412	0.0395	0.0378
	<i>acov-wls</i>	0.0457	0.0453	0.0449	0.0446	0.0439	0.0432	0.0423	0.0409	0.0393	0.0376
	<i>sari-wls</i>	0.0516	0.0511	0.0506	0.0501	0.0492	0.0475	0.0467	0.0435	0.0405	0.0383
	cst <i>wlsv-shr</i>	0.0589	0.0585	0.0580	0.0576	0.0569	0.0559	0.0546	0.0524	0.0506	0.0489
	<i>sari-shr</i>	0.0521	0.0516	0.0511	0.0506	0.0497	0.0481	0.0473	0.0442	0.0414	0.0395
	<i>acov-shr</i>	0.0488	0.0483	0.0478	0.0474	0.0466	0.0457	0.0446	0.0427	0.0409	0.0391
	<i>acov-wls</i>	0.0488	0.0483	0.0479	0.0475	0.0467	0.0458	0.0447	0.0430	0.0411	0.0392
	<i>sari-wls</i>	0.0524	0.0519	0.0514	0.0509	0.0501	0.0484	0.0476	0.0446	0.0419	0.0399
ite	<i>wlsv-shr</i>	0.0576	0.0572	0.0567	0.0563	0.0555	0.0544	0.0532	0.0508	0.0488	0.0471
	<i>sari-shr</i>	0.0512	0.0507	0.0502	0.0497	0.0489	0.0471	0.0463	0.0430	0.0400	0.0379
	<i>acov-wls</i>	0.0457	0.0453	0.0449	0.0446	0.0439	0.0432	0.0423	0.0408	0.0392	0.0375
	<i>sari-wls</i>	0.0516	0.0511	0.0506	0.0501	0.0492	0.0475	0.0467	0.0435	0.0405	0.0383
oct	<i>wlsv</i>	0.0578	0.0573	0.0569	0.0565	0.0557	0.0547	0.0534	0.0511	0.0492	0.0476
	<i>bdsbr</i>	0.0558	0.0553	0.0548	0.0543	0.0536	0.0524	0.0511	0.0487	0.0466	0.0447
	<i>acov</i>	0.0457	0.0452	0.0448	0.0445	0.0439	0.0432	0.0424	0.0410	0.0393	0.0376
	<i>str</i>	0.0546	0.0541	0.0536	0.0531	0.0523	0.0509	0.0499	0.0469	0.0448	0.0428
	<i>Sshbr</i>	0.0420	0.0415	0.0412	0.0409	0.0402	0.0397	0.0386	0.0379	0.0364	0.0345
	Randomforest	0.0475	0.0468	0.0461	0.0455	0.0447	0.0436	0.0424	0.0407	0.0389	0.0370
	XGBoost	0.0496	0.0486	0.0478	0.0471	0.0462	0.0450	0.0435	0.0419	0.0398	0.0376
	LightGBM	0.0484	0.0476	0.0468	0.0462	0.0454	0.0441	0.0430	0.0414	0.0395	0.0374

**Table 5.20:** Energy Load Areas Forecast Results for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	0.0740	0.0738	0.0737	0.0735	0.0733	0.0729	0.0725	0.0721	0.0717	0.0702
	Randomforest	0.0683	0.0680	0.0677	0.0673	0.0670	0.0665	0.0658	0.0655	0.0639	0.0614
	XGBoost	0.0710	0.0704	0.0700	0.0698	0.0692	0.0686	0.0682	0.0675	0.0658	0.0635
	LightGBM	0.0714	0.0709	0.0706	0.0703	0.0698	0.0692	0.0685	0.0682	0.0665	0.0637

**Table 5.21:** Energy Load Italy Forecast Results for Naive Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
Base	Base	0.0987	0.1099	0.1068	0.0920	0.0892	0.0443	0.0398	0.0401	0.0346	0.0338	
Bottom Up	Bottom Up	0.0965	0.0965	0.0963	0.0962	0.0958	0.0955	0.0946	0.0941	0.0921	0.0903	
tcs	<i>wlsv-shr</i>	0.0676	0.0674	0.0670	0.0668	0.0660	0.0645	0.0634	0.0603	0.0581	0.0558	
	<i>sar1-shr</i>	0.0541	0.0538	0.0534	0.0529	0.0523	0.0484	0.0487	0.0411	0.0353	0.0309	
	<i>acov-shr</i>	0.0387	0.0385	0.0383	0.0381	0.0376	0.0370	0.0365	0.0351	0.0335	0.0308	
	<i>acov-wls</i>	0.0382	0.0380	0.0378	0.0376	0.0372	0.0366	0.0362	0.0348	0.0333	0.0308	
	<i>sar1-wls</i>	0.0543	0.0540	0.0536	0.0532	0.0525	0.0487	0.0491	0.0414	0.0354	0.0309	
cst	<i>wlsv-shr</i>	0.0734	0.0732	0.0730	0.0727	0.0720	0.0707	0.0697	0.0667	0.0642	0.0623	
	<i>sar1-shr</i>	0.0563	0.0560	0.0556	0.0551	0.0544	0.0514	0.0516	0.0446	0.0386	0.0350	
	<i>acov-shr</i>	0.0461	0.0458	0.0455	0.0451	0.0446	0.0434	0.0426	0.0401	0.0380	0.0352	
	<i>acov-wls</i>	0.0456	0.0454	0.0451	0.0448	0.0444	0.0434	0.0428	0.0410	0.0393	0.0366	
	<i>sar1-wls</i>	0.0566	0.0563	0.0559	0.0556	0.0549	0.0516	0.0520	0.0451	0.0394	0.0357	
ETS	ite	<i>wlsv-shr</i>	0.0669	0.0667	0.0663	0.0660	0.0653	0.0635	0.0625	0.0588	0.0560	0.0540
		<i>sar1-shr</i>	0.0540	0.0536	0.0532	0.0527	0.0520	0.0485	0.0487	0.0410	0.0350	0.0306
		<i>acov-wls</i>	0.0383	0.0381	0.0378	0.0377	0.0372	0.0366	0.0362	0.0349	0.0334	0.0310
		<i>sar1-wls</i>	0.0544	0.0540	0.0536	0.0532	0.0525	0.0487	0.0491	0.0414	0.0354	0.0309
	oct	<i>wlsv</i>	0.0676	0.0674	0.0670	0.0668	0.0661	0.0645	0.0634	0.0604	0.0581	0.0558
		<i>bdsbr</i>	0.0628	0.0625	0.0621	0.0617	0.0610	0.0589	0.0578	0.0540	0.0508	0.0485
		<i>acov</i>	0.0390	0.0388	0.0386	0.0384	0.0380	0.0374	0.0371	0.0358	0.0343	0.0320
		<i>str</i>	0.0662	0.0659	0.0656	0.0653	0.0645	0.0625	0.0617	0.0570	0.0541	0.0513
		<i>Sshbr</i>	0.0382	0.0380	0.0378	0.0377	0.0373	0.0368	0.0362	0.0353	0.0341	0.0315
Randomforest		0.0426	0.0421	0.0412	0.0406	0.0398	0.0383	0.0371	0.0349	0.0340	0.0314	
XGBoost		0.0439	0.0433	0.0424	0.0419	0.0412	0.0395	0.0384	0.0366	0.0353	0.0327	
LightGBM		0.0424	0.0417	0.0409	0.0404	0.0396	0.0382	0.0367	0.0354	0.0338	0.0311	

**Table 5.22:** Energy Load Italy Forecast Results for ETS Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.0936	0.0918	0.0920	0.0914	0.0891	0.0402	0.0336	0.0371	0.0329	0.0355	
	Bottom Up	0.0928	0.0928	0.0926	0.0924	0.0920	0.0913	0.0904	0.0893	0.0875	0.0858	
tcs	<i>wlsv-shr</i>	0.0640	0.0638	0.0634	0.0630	0.0624	0.0608	0.0594	0.0565	0.0546	0.0531	
	<i>sari-shr</i>	0.0517	0.0514	0.0509	0.0505	0.0498	0.0461	0.0463	0.0396	0.0347	0.0322	
	<i>acov-shr</i>	0.0403	0.0400	0.0398	0.0396	0.0393	0.0384	0.0381	0.0362	0.0349	0.0329	
	<i>acov-wls</i>	0.0398	0.0396	0.0394	0.0393	0.0390	0.0382	0.0378	0.0359	0.0348	0.0330	
	<i>sari-wls</i>	0.0514	0.0511	0.0507	0.0502	0.0496	0.0458	0.0462	0.0394	0.0346	0.0323	
cst	<i>wlsv-shr</i>	0.0692	0.0690	0.0687	0.0684	0.0676	0.0662	0.0650	0.0621	0.0603	0.0589	
	<i>sari-shr</i>	0.0531	0.0528	0.0524	0.0520	0.0515	0.0483	0.0483	0.0417	0.0375	0.0355	
	<i>acov-shr</i>	0.0458	0.0454	0.0450	0.0448	0.0443	0.0432	0.0422	0.0395	0.0378	0.0358	
	<i>acov-wls</i>	0.0461	0.0459	0.0457	0.0454	0.0450	0.0439	0.0429	0.0406	0.0389	0.0370	
	<i>sari-wls</i>	0.0535	0.0532	0.0529	0.0525	0.0520	0.0487	0.0489	0.0424	0.0385	0.0365	
SARIMA												
	ite	<i>wlsv-shr</i>	0.0633	0.0630	0.0625	0.0621	0.0614	0.0597	0.0582	0.0552	0.0531	0.0516
		<i>sari-shr</i>	0.0511	0.0508	0.0503	0.0498	0.0491	0.0455	0.0457	0.0390	0.0337	0.0313
		<i>acov-wls</i>	0.0400	0.0398	0.0396	0.0394	0.0391	0.0383	0.0378	0.0360	0.0348	0.0331
		<i>sari-wls</i>	0.0514	0.0511	0.0507	0.0502	0.0496	0.0458	0.0462	0.0394	0.0346	0.0323
oct	<i>wlsv</i>	0.0642	0.0640	0.0636	0.0633	0.0628	0.0611	0.0597	0.0569	0.0551	0.0536	
	<i>bdskr</i>	0.0600	0.0597	0.0593	0.0588	0.0581	0.0561	0.0548	0.0513	0.0492	0.0478	
	<i>acov</i>	0.0409	0.0407	0.0405	0.0403	0.0401	0.0393	0.0389	0.0372	0.0360	0.0343	
	<i>str</i>	0.0634	0.0631	0.0628	0.0624	0.0619	0.0599	0.0587	0.0539	0.0518	0.0503	
	<i>Sshr</i>	0.0370	0.0368	0.0365	0.0363	0.0360	0.0354	0.0350	0.0341	0.0328	0.0314	
Randomforest												
		0.0433	0.0428	0.0420	0.0413	0.0405	0.0388	0.0374	0.0353	0.0346	0.0322	
		0.0445	0.0438	0.0431	0.0425	0.0416	0.0400	0.0384	0.0368	0.0355	0.0335	
XGBoost												
		0.0425	0.0419	0.0411	0.0405	0.0398	0.0382	0.0368	0.0354	0.0341	0.0317	
LightGBM												

**Table 5.23:** Energy Load Italy Forecast Results for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.0760	0.0786	0.0765	0.0713	0.0700	0.0425	0.0400	0.0408	0.0377	0.0405
	Bottom Up	0.0753	0.0752	0.0751	0.0750	0.0747	0.0744	0.0738	0.0727	0.0719	0.0712
Forecast Combination	tcs <i>wlsv-shr</i>	0.0611	0.0610	0.0607	0.0605	0.0601	0.0592	0.0586	0.0563	0.0552	0.0542
	<i>sari-shr</i>	0.0513	0.0511	0.0508	0.0505	0.0501	0.0479	0.0480	0.0435	0.0406	0.0391
	<i>acov-shr</i>	0.0450	0.0449	0.0446	0.0445	0.0442	0.0435	0.0432	0.0416	0.0405	0.0393
	<i>acov-wls</i>	0.0447	0.0446	0.0444	0.0442	0.0440	0.0432	0.0429	0.0412	0.0401	0.0389
	<i>sari-wls</i>	0.0512	0.0510	0.0506	0.0503	0.0499	0.0478	0.0479	0.0435	0.0406	0.0390
	cst <i>wlsv-shr</i>	0.0640	0.0638	0.0636	0.0635	0.0631	0.0624	0.0618	0.0596	0.0586	0.0577
	<i>sari-shr</i>	0.0525	0.0523	0.0520	0.0517	0.0513	0.0496	0.0495	0.0456	0.0431	0.0418
	<i>acov-shr</i>	0.0484	0.0482	0.0479	0.0477	0.0473	0.0465	0.0460	0.0439	0.0426	0.0414
	<i>acov-wls</i>	0.0480	0.0479	0.0476	0.0474	0.0471	0.0463	0.0458	0.0438	0.0426	0.0414
	<i>sari-wls</i>	0.0526	0.0524	0.0521	0.0519	0.0515	0.0496	0.0496	0.0458	0.0433	0.0420
ite	<i>wlsv-shr</i>	0.0610	0.0608	0.0605	0.0603	0.0600	0.0590	0.0584	0.0560	0.0549	0.0539
	<i>sari-shr</i>	0.0510	0.0508	0.0505	0.0502	0.0497	0.0476	0.0478	0.0432	0.0402	0.0387
	<i>acov-wls</i>	0.0447	0.0445	0.0443	0.0441	0.0439	0.0431	0.0428	0.0412	0.0400	0.0388
	<i>sari-wls</i>	0.0512	0.0510	0.0506	0.0503	0.0499	0.0478	0.0479	0.0435	0.0406	0.0390
	oct <i>wlsv</i>	0.0608	0.0606	0.0604	0.0601	0.0598	0.0589	0.0583	0.0560	0.0549	0.0539
	<i>bdsbr</i>	0.0588	0.0586	0.0583	0.0580	0.0576	0.0566	0.0559	0.0533	0.0522	0.0510
	<i>acov</i>	0.0453	0.0452	0.0450	0.0448	0.0445	0.0439	0.0436	0.0421	0.0409	0.0399
	<i>str</i>	0.0565	0.0562	0.0560	0.0558	0.0554	0.0539	0.0536	0.0503	0.0487	0.0477
	<i>Sshbr</i>	0.0402	0.0400	0.0398	0.0397	0.0395	0.0390	0.0385	0.0379	0.0370	0.0358
	Randomforest	0.0454	0.0449	0.0443	0.0439	0.0434	0.0424	0.0416	0.0401	0.0387	0.0366
	XGBoost	0.0454	0.0448	0.0442	0.0438	0.0431	0.0422	0.0410	0.0397	0.0383	0.0361
	LightGBM	0.0452	0.0447	0.0441	0.0437	0.0432	0.0422	0.0412	0.0400	0.0386	0.0367

**Table 5.24:** Energy Load Italy Forecast Results for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in WAPE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Zones Results											
Naive	Randomforest	0.9776	0.9773	0.9771	0.9782	0.9781	0.9788	0.9805	0.9778	0.9789	0.9791
	XGBoost	0.9879	0.9923	0.9931	0.9943	0.9954	0.9955	0.9971	0.9929	0.9928	1.0016
	LightGBM	0.9721	0.9744	0.9745	0.9755	0.9757	0.9750	0.9775	0.9765	0.9709	0.9768
ETS	Randomforest	0.9705	0.9678	0.9655	0.9671	0.9641	0.9681	0.9622	0.9669	0.9681	0.9643
	XGBoost	0.9586	0.9614	0.9616	0.9607	0.9569	0.9578	0.9511	0.9515	0.9453	0.9314
	LightGBM	0.9591	0.9595	0.9597	0.9597	0.9562	0.9585	0.9548	0.9612	0.9581	0.9535
SARIMA	Randomforest	0.9793	0.9774	0.9753	0.9769	0.9731	0.9762	0.9761	0.9800	0.9729	0.9701
	XGBoost	0.9665	0.9692	0.9688	0.9703	0.9674	0.9679	0.9677	0.9645	0.9570	0.9427
	LightGBM	0.9637	0.9655	0.9641	0.9650	0.9618	0.9610	0.9653	0.9615	0.9586	0.9496
Forecast Combination	Randomforest	0.9971	0.9959	0.9958	0.9971	0.9937	0.9965	0.9976	0.9992	0.9985	0.9948
	XGBoost	1.0052	1.0095	1.0103	1.0106	1.0096	1.0089	1.0081	1.0063	1.0115	1.0143
	LightGBM	0.9972	0.9993	1.0000	1.0002	1.0001	0.9985	0.9994	0.9983	0.9950	0.9907

**Table 5.25:** Sensitivity Analysis of Features Matrix for Zones: Results for Energy Load data.  
 (Note: This table presents relative forecast accuracy based on WAPE, comparing the complete features matrix to the compact features matrix. A value less than one signifies improved performance with the complete matrix.)

from the complete feature matrix. However, the compact feature matrix performs better for XGBoost with Naive and forecast combination as the base models, as well as LightGBM with the forecast combination.

At the Italy level, for almost all reconciliation models, the compact feature matrix yields better results. This shift in performance suggests that, at higher levels of aggregation, reducing the complexity of input features may enhance the ability of ML-based models to generalize more effectively.

These findings reinforce the observation that the relative effectiveness of complete versus compact feature matrices is application dependent, varying across different hierarchical levels and datasets.

### 5.2.3 SENSITIVITY TO TEMPORAL AGGREGATION ORDERS

In this section, we conduct a sensitivity analysis on the temporal hierarchy to assess the impact of different aggregation levels on forecast reconciliation performance. Similarly to the Citi Bike dataset, we selected the 30-minute, 1-hour, and 24-hour temporal frequencies and compared them against the case where all 10 temporal frequencies were used. The relative WAPE values summarizing these results are presented in Tables 5.28 to 5.31.

For most reconciliation methods and base model combinations, using all temporal frequen-

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Areas Results											
Naive	Randomforest	0.9916	0.9907	0.9904	0.9915	0.9907	0.9924	0.9935	0.9925	0.9941	0.9928
	XGBoost	1.0043	1.0079	1.0078	1.0083	1.0077	1.0112	1.0106	1.0079	1.0127	1.0177
	LightGBM	0.9844	0.9865	0.9866	0.9878	0.9866	0.9877	0.9889	0.9877	0.9896	0.9976
ETS	Randomforest	0.9669	0.9628	0.9605	0.9608	0.9581	0.9603	0.9533	0.9571	0.9497	0.9432
	XGBoost	0.9706	0.9731	0.9721	0.9754	0.9704	0.9703	0.9668	0.9698	0.9768	0.9667
	LightGBM	0.9646	0.9650	0.9660	0.9671	0.9656	0.9653	0.9602	0.9619	0.9626	0.9623
SARIMA	Randomforest	0.9750	0.9709	0.9680	0.9691	0.9680	0.9669	0.9683	0.9670	0.9597	0.9643
	XGBoost	0.9686	0.9705	0.9699	0.9683	0.9697	0.9712	0.9737	0.9728	0.9664	0.9497
	LightGBM	0.9713	0.9725	0.9729	0.9715	0.9729	0.9686	0.9718	0.9697	0.9636	0.9622
Forecast Combination	Randomforest	0.9886	0.9844	0.9841	0.9861	0.9835	0.9856	0.9854	0.9857	0.9906	0.9861
	XGBoost	1.0172	1.0207	1.0207	1.0225	1.0211	1.0216	1.0206	1.0160	1.0198	1.0185
	LightGBM	1.0042	1.0053	1.0047	1.0063	1.0045	1.0093	1.0064	1.0040	1.0054	1.0048

**Table 5.26: Sensitivity Analysis of Features Matrix for Areas: Results for Energy Load data.**  
 (Note: This table presents relative forecast accuracy based on WAPE, comparing the complete features matrix to the compact features matrix. A value less than one signifies improved performance with the complete matrix.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Italy Results											
Naive	Randomforest	1.0053	1.0042	1.0045	1.0069	1.0063	1.0068	1.0114	1.0062	1.0068	1.0113
	XGBoost	1.0303	1.0340	1.0332	1.0338	1.0355	1.0339	1.0331	1.0320	1.0333	1.0441
	LightGBM	1.0034	1.0055	1.0057	1.0070	1.0057	1.0073	1.0098	1.0048	1.0001	1.0174
ETS	Randomforest	1.0027	0.9987	0.9979	0.9993	0.9979	1.0009	0.9922	1.0168	1.0019	1.0070
	XGBoost	1.0284	1.0298	1.0303	1.0319	1.0271	1.0329	1.0244	1.0361	1.0362	1.0400
	LightGBM	1.0121	1.0140	1.0148	1.0163	1.0144	1.0189	1.0260	1.0213	1.0374	1.0454
SARIMA	Randomforest	0.9962	0.9914	0.9886	0.9918	0.9914	0.9946	0.9934	1.0037	0.9991	1.0061
	XGBoost	0.9882	0.9919	0.9898	0.9890	0.9914	0.9950	0.9997	1.0012	1.0017	0.9927
	LightGBM	1.0035	1.0043	1.0045	1.0070	1.0080	1.0118	1.0157	1.0111	1.0136	1.0188
Forecast Combination	Randomforest	1.0048	1.0008	0.9994	1.0023	0.9981	0.9994	1.0021	0.9997	1.0039	1.0113
	XGBoost	1.0632	1.0666	1.0679	1.0689	1.0702	1.0687	1.0800	1.0723	1.0734	1.0863
	LightGBM	1.0467	1.0475	1.0488	1.0503	1.0487	1.0508	1.0564	1.0482	1.0479	1.0579

**Table 5.27: Sensitivity Analysis of Features Matrix for Italy: Results for Energy Load data.**  
 (Note: This table presents relative forecast accuracy based on WAPE, comparing the complete features matrix to the compact features matrix. A value less than one signifies improved performance with the complete matrix.)

Base Method	Forecast Combination	Zones			Areas			Italy		
		30min	1h	24h	30min	1h	24h	30min	1h	24h
Naive	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Randomforest	1.0106	1.0117	1.0208	1.0096	1.0109	1.0202	1.0118	1.0121	1.0199
	XGBoost	0.9861	0.9891	0.9753	0.9854	0.9881	0.9796	0.9934	0.9961	0.9863
	LightGBM	0.9887	0.9907	0.9819	0.9836	0.9850	0.9747	0.9945	0.9956	0.9774

**Table 5.28:** Sensitivity Analysis of Temporal Frequencies: Results for Energy Load Naive Base Method.

(Note: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)

cies results in better forecast accuracy, as indicated by relative WAPE values below one (which means that retaining the full temporal structure leads to lower forecast errors). However, a few notable exceptions exist. For example, Random Forest with Naive as the base forecast performed better when using only the selected temporal frequencies (30-minute, 1-hour, and 24-hour) instead of the full hierarchy, across all cross-sectional levels (zones, areas, and Italy).

Also, the improvements appear consistent across different base forecasts (ETS, SARIMA, etc.) and the magnitude of improvement for linear reconciliation methods is not necessarily larger than that of machine learning based methods. In some cases, XGBoost and LightGBM exhibit only marginal improvements or even slight deteriorations in certain cross-sectional levels (e.g., some zones or areas). Notably, for oct with Sshr, we see some of the best relative improvements.

It also should be noted again that base forecasts and the bottom-up for the reconciliation method remain unaffected for all base method forecasts by the choice of temporal frequencies, as these models do not rely on temporal aggregation for reconciliation. Nevertheless, they are included in the tables for completeness. These findings further reinforce the dataset dependent nature of hierarchical forecasting, where the effectiveness of reconciliation strategies may vary based on temporal structure and base forecasting method.

Base Method	Forecast Combination	Zones			Areas			Italy		
		30min	1h	24h	30min	1h	24h	30min	1h	24h
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Bottom Up	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ETS	tcs <i>wlsv-shr</i>	0.9579	0.9572	1.0359	0.9240	0.9232	1.0182	0.8756	0.8747	0.9921
	<i>sari-shr</i>	0.9156	0.9141	0.9390	0.8761	0.8741	0.9103	0.8079	0.8062	0.8513
	<i>acov-shr</i>	0.6792	0.6799	0.7964	0.5794	0.5799	0.7105	0.4803	0.4811	0.6266
	<i>acov-wls</i>	0.6799	0.6799	0.7875	0.5756	0.5752	0.6981	0.4715	0.4714	0.6125
	<i>sari-wls</i>	0.9186	0.9172	0.9396	0.8793	0.8772	0.9070	0.8120	0.8102	0.8476
	cst <i>wlsv-shr</i>	0.9663	0.9656	1.0132	0.9491	0.9482	1.0166	0.9316	0.9310	1.0160
	<i>sari-shr</i>	0.9158	0.9146	0.9359	0.8784	0.8765	0.9181	0.8164	0.8152	0.8773
	<i>acov-shr</i>	0.7796	0.7771	0.8156	0.6979	0.6948	0.7427	0.5926	0.5914	0.6668
	<i>acov-wls</i>	0.7769	0.7749	0.8332	0.6932	0.6908	0.7621	0.5844	0.5839	0.6885
	<i>sari-wls</i>	0.9214	0.9200	0.9442	0.8836	0.8820	0.9248	0.8208	0.8194	0.8898
ite	<i>wlsv-shr</i>	0.9508	0.9498	1.0050	0.9182	0.9172	0.9884	0.8684	0.8680	0.9497
	<i>sari-shr</i>	0.9121	0.9108	0.9325	0.8728	0.8708	0.9066	0.8055	0.8036	0.8471
	<i>acov-wls</i>	0.6819	0.6819	0.7888	0.5772	0.5767	0.7006	0.4725	0.4723	0.6168
	<i>sari-wls</i>	0.9186	0.9171	0.9393	0.8794	0.8772	0.9078	0.8122	0.8103	0.8482
oct	<i>wlsv</i>	0.9560	0.9553	1.0216	0.9200	0.9192	0.9970	0.8687	0.8679	0.9652
	<i>bdshr</i>	0.9129	0.9116	0.9589	0.8755	0.8736	0.9378	0.8294	0.8286	0.9346
	<i>acov</i>	0.6569	0.6559	0.7243	0.5572	0.5554	0.6167	0.4607	0.4593	0.5376
	<i>str</i>	0.8880	0.8850	0.8184	0.8482	0.8451	0.7736	0.7944	0.7924	0.7449
	<i>Sshr</i>	0.7111	0.7105	0.8848	0.6043	0.6037	0.8122	0.5112	0.5107	0.7559
Randomforest	Randomforest	0.8768	0.8733	0.8364	0.8140	0.8091	0.7667	0.7250	0.7203	0.6613
	XGBoost	0.8821	0.8814	0.8395	0.8317	0.8300	0.7833	0.7556	0.7541	0.7130
	LightGBM	0.8728	0.8717	0.8195	0.8152	0.8116	0.7580	0.7380	0.7349	0.6783

**Table 5.29:** Sensitivity Analysis of Temporal Frequencies: Results for Energy Load ETS Base Method.

(Note: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)

Base Method	Forecast Combination	Zones			Areas			Italy		
		30min	1h	24h	30min	1h	24h	30min	1h	24h
	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Bottom Up	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SARIMA	tcs <i>wlsv-shr</i>	0.9028	0.9008	0.9423	0.8715	0.8697	0.9299	0.8364	0.8358	0.9342
	<i>sari-shr</i>	0.8473	0.8442	0.8587	0.8092	0.8064	0.8407	0.7750	0.7732	0.8625
	<i>acov-shr</i>	0.6733	0.6713	0.7703	0.5775	0.5760	0.7029	0.5040	0.5034	0.6608
	<i>acov-wls</i>	0.6754	0.6731	0.7715	0.5781	0.5763	0.7072	0.5006	0.5003	0.6674
	<i>sari-wls</i>	0.8465	0.8434	0.8640	0.8061	0.8034	0.8504	0.7695	0.7679	0.8732
	cst <i>wlsv-shr</i>	0.9102	0.9081	0.9345	0.8950	0.8924	0.9342	0.8907	0.8899	0.9585
	<i>sari-shr</i>	0.8438	0.8406	0.8620	0.8104	0.8072	0.8555	0.7801	0.7785	0.8777
	<i>acov-shr</i>	0.7400	0.7358	0.7692	0.6629	0.6587	0.7052	0.5929	0.5902	0.6550
	<i>acov-wls</i>	0.7449	0.7416	0.7786	0.6684	0.6652	0.7200	0.5952	0.5941	0.6724
	<i>sari-wls</i>	0.8464	0.8431	0.8673	0.8135	0.8105	0.8687	0.7806	0.7789	0.8910
Randomforest	ite <i>wlsv-shr</i>	0.8946	0.8920	0.9259	0.8618	0.8590	0.9101	0.8310	0.8295	0.9235
	<i>sari-shr</i>	0.8440	0.8407	0.8560	0.8037	0.8005	0.8350	0.7708	0.7688	0.8583
	<i>acov-wls</i>	0.6770	0.6748	0.7723	0.5801	0.5784	0.7070	0.5027	0.5024	0.6683
	<i>sari-wls</i>	0.8464	0.8433	0.8640	0.8061	0.8033	0.8505	0.7694	0.7678	0.8733
	oct <i>wlsv</i>	0.9064	0.9047	0.9496	0.8758	0.8740	0.9440	0.8375	0.8367	0.9442
	<i>bdshsr</i>	0.8730	0.8701	0.9226	0.8388	0.8354	0.9231	0.8041	0.8027	0.9480
	<i>acov</i>	0.6553	0.6542	0.7126	0.5620	0.5599	0.6161	0.4934	0.4923	0.5614
	<i>str</i>	0.8619	0.8580	0.8302	0.8330	0.8298	0.8076	0.8004	0.7989	0.7903
	<i>Sshr</i>	0.6996	0.6971	0.8259	0.5797	0.5774	0.7387	0.5001	0.4999	0.7047
	XGBoost	0.8633	0.8597	0.8235	0.8058	0.8011	0.7556	0.7440	0.7396	0.6922
	LightGBM	0.8527	0.8523	0.8033	0.8125	0.8112	0.7614	0.7776	0.7757	0.7467
	Randomforest	0.8523	0.8502	0.7945	0.8015	0.7979	0.7480	0.7681	0.7652	0.7277

**Table 5.30:** Sensitivity Analysis of Temporal Frequencies: Results for Energy Load SARIMA Base Method.

(Note: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)

Base Method	Forecast Combination	Zones			Areas			Italy			
		30min	1h	24h	30min	1h	24h	30min	1h	24h	
	Base	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	Bottom Up	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
tcs	<i>wlsv-shr</i>	0.9599	0.9591	0.9790	0.9412	0.9402	0.9702	0.9186	0.9181	0.9637	
	<i>sari-shr</i>	0.9324	0.9307	0.9508	0.9109	0.9094	0.9478	0.8842	0.8840	0.9483	
	<i>acov-shr</i>	0.8242	0.8224	0.8897	0.7517	0.7506	0.8564	0.6864	0.6861	0.8056	
	<i>acov-wls</i>	0.8230	0.8213	0.8838	0.7509	0.7498	0.8513	0.6869	0.6868	0.8043	
	<i>sari-wls</i>	0.9290	0.9273	0.9424	0.9060	0.9045	0.9385	0.8775	0.8770	0.9405	
	cst	<i>wlsv-shr</i>	0.9632	0.9625	0.9746	0.9546	0.9536	0.9745	0.9490	0.9486	0.9842
		<i>sari-shr</i>	0.9299	0.9282	0.9493	0.9094	0.9081	0.9503	0.8870	0.8864	0.9617
		<i>acov-shr</i>	0.8637	0.8614	0.8960	0.8116	0.8093	0.8713	0.7538	0.7529	0.8289
		<i>acov-wls</i>	0.8651	0.8630	0.8987	0.8122	0.8104	0.8738	0.7506	0.7501	0.8338
		<i>sari-wls</i>	0.9318	0.9301	0.9516	0.9114	0.9101	0.9542	0.8871	0.8865	0.9674
Forecast Combination	ite	<i>wlsv-shr</i>	0.9547	0.9538	0.9682	0.9364	0.9350	0.9586	0.9146	0.9138	0.9544
		<i>sari-shr</i>	0.9262	0.9245	0.9394	0.9031	0.9015	0.9344	0.8765	0.8761	0.9356
		<i>acov-wls</i>	0.8228	0.8210	0.8825	0.7505	0.7493	0.8494	0.6859	0.6858	0.8024
		<i>sari-wls</i>	0.9289	0.9273	0.9422	0.9059	0.9044	0.9383	0.8774	0.8770	0.9404
	oct	<i>wlsv</i>	0.9607	0.9599	0.9788	0.9425	0.9416	0.9713	0.9184	0.9179	0.9649
		<i>bdshr</i>	0.9394	0.9381	0.9552	0.9189	0.9173	0.9496	0.8981	0.8969	0.9526
		<i>acov</i>	0.8101	0.8084	0.8571	0.7327	0.7313	0.7917	0.6711	0.6704	0.7405
		<i>str</i>	0.9164	0.9137	0.8936	0.8878	0.8857	0.8647	0.8468	0.8449	0.8314
		<i>Sshr</i>	0.7968	0.7946	0.8600	0.7012	0.6995	0.8055	0.6370	0.6370	0.7964
	Randomforest	0.9563	0.9560	0.9488	0.9316	0.9307	0.9263	0.8994	0.8982	0.8721	
	XGBoost	0.9399	0.9421	0.9180	0.9231	0.9242	0.9119	0.8862	0.8874	0.8756	
	LightGBM	0.9404	0.9415	0.9219	0.9160	0.9170	0.9038	0.8949	0.8966	0.8849	

**Table 5.31:** Sensitivity Analysis of Temporal Frequencies: Results for Energy Load Forecast Combination Base Method.

(Note: This table shows relative forecast accuracy measured in WAPE when using all ten temporal frequencies for reconciliation compared to using only the temporal frequencies of interest. A value below one indicates better performance when using the former.)



# 6

## Conclusion

### 6.1 ADDRESSING THE STUDY OBJECTIVES

This study introduced several modifications to the original framework proposed by [3] to enhance the robustness of forecast reconciliation across different domains. One improvement was the adoption of the latest FoReco package version, which reduced computational time and resources. Additionally, to ensure consistency in non-negativity and rounding adjustments, we applied the same treatment to both linear and machine learning-based reconciliation methods, eliminating inconsistencies in how constraints were handled.

A major contribution of this work was the expansion of covariance matrix selection for linear reconciliation methods. To assess whether different covariance structures could influence the reconciliation performance, we systematically evaluated additional covariance estimators. These extensions allowed us to examine a broader set of reconciliation models. While the most significant performance changes were later observed when changing from in-sample to validation error, this expansion provided a more comprehensive evaluation.

Another key refinement in this study was considering a validation set in both linear and ML-based reconciliation models. While the original work used in-sample errors from the training set for linear methods, a validation set was implemented for ML-based reconciliation. This discrepancy likely contributed to the superior performance of ML models. By using a validation set for both methods, we aimed to determine whether this adjustment could enhance the accu-

racy of linear models or even enable them to outperform ML-based methods. The results confirmed that this modification significantly improved the performance of linear reconciliation models, leading to consistent outperformance over ML based methods across all evaluations. Among the ML models, Random Forest continued to deliver the best results, but overall, linear reconciliation models consistently achieved superior accuracy. These findings highlight the critical role of a validation set in reconciliation and demonstrate that linear models can outperform ML based approaches in hierarchical forecasting.

Expanding the analysis to a new industry with a more complex hierarchy allowed us to assess the different aspects of forecast reconciliation methods. Despite the structural differences between the Citi Bike and energy load datasets, the overall conclusions remained consistent. Linear reconciliation models continued to outperform ML-based methods, demonstrating their robustness across different forecasting applications. However, some variations in model performance were observed, which will be discussed in detail in Section 6.2.

## 6.2 SUMMARY OF RESULTS AND DATASET COMPARISONS

This section provides a comparative analysis of the forecasting results across the two datasets, Citi Bike and Energy Load, highlighting key performance trends and methodological insights.

*Overall Forecast Performance:* Across both datasets, linear reconciliation methods outperformed machine learning-based reconciliation methods. The best performing models for Citi Bike data include tcs (acov-wls, acov-shr) and oct (Sshr, acov). Additionally, linear reconciliation methods demonstrated greater improvements across different base forecasts, whereas ML based methods exhibited more stability across configurations. For the energy load dataset, methods tcs (acov-wls, acov-shr) and oct (Sshr) emerged as the top performing approaches across different hierarchical levels. At the bottom level, which consists of individual geographical zones, the best accuracy was obtained when SARIMA was used as the base forecast and reconciled with tcs (acov-shr). At the regional level, SARIMA with oct (Sshr) achieved the highest accuracy, a trend that continued at the Italy level, where this combination again provided the best results. When evaluating results under the MASE accuracy metric instead of WAPE, similar conclusions were reached. However, a key distinction was that in nearly all cases, the oct (Sshr) model consistently produced the best results, further reinforcing its robustness across hierarchical levels.

*Impact of Feature Matrix Selection:* For the Citi Bike dataset, results indicated that using the complete feature matrix, which incorporates information across all hierarchical levels, im-

proved forecast accuracy at the bottom-level H<sub>3</sub> cells. However, at the market level, the impact of using the complete feature matrix was either negligible or slightly negative, suggesting that incorporating additional hierarchical information does not necessarily enhance ML-based reconciliation at higher aggregation levels. This finding highlights that the effectiveness of feature selection is highly dependent on the hierarchical structure and dataset characteristics.

A similar analysis was conducted for the Energy Load dataset, where the impact of feature matrix selection varied across hierarchical levels. At the zone level, using the complete feature matrix led to improved accuracy for most reconciliation methods. However, an exception was observed with XGBoost using the forecast combination as the base model, which performed better with the compact feature matrix. At the middle level, most reconciliation methods again benefited from the complete feature matrix, except for XGBoost with Naive and forecast combination base forecasts, as well as LightGBM with forecast combination, which showed better performance using the compact feature matrix. Interestingly, at the highest aggregation level (Italy), the trend reversed, with the compact feature matrix yielding better results for nearly all models. This suggests that at higher aggregation levels, reducing feature complexity may enhance model generalization, allowing ML-based reconciliation to focus on the most relevant patterns.

Overall, the choice between complete and compact feature matrices had a dataset-dependent effect. In both datasets, ML models generally performed better when trained with the complete feature matrix at the bottom level, reinforcing the importance of incorporating hierarchical information for granular forecasting. However, at higher levels of aggregation, the impact of feature matrix selection diverged. These findings emphasize the importance of adapting feature selection strategies based on dataset characteristics and hierarchical structure.

*Sensitivity to Temporal Aggregation Orders:* For the Citi Bike dataset, forecast accuracy improved when all ten temporal frequencies were used instead of a subset (30min, 1h, and 24h). The best improvements were observed in linear reconciliation models: tcs and cst when combined with covariance matrices such as acov-shr and acov-wls; ite with acov-wls; and oct with acov and Ssh. These findings emphasize the benefits of incorporating all temporal frequencies into the reconciliation process, especially for linear models.

For the Energy Load dataset, the results again confirmed that using all available temporal frequencies generally led to better forecasting accuracy. However, a few exceptions were noted, particularly for Random Forest with Naive as the base forecast, where using only the selected temporal frequencies yielded better accuracy across all hierarchical levels. The relative improvements varied across different base forecasts and methods, with no consistent pattern favoring

either linear or ML-based models. In some cases, XGBoost and LightGBM exhibited only marginal improvements or slight deteriorations, particularly at some zone and area levels. However, oct (Sshr) continued to show strong improvements, reinforcing its effectiveness in cross-temporal reconciliation.

These results reinforce that while retaining full temporal granularity is generally beneficial, the extent of improvement varies depending on the dataset and forecasting method.

### 6.3 LIMITATIONS AND FUTURE WORKS

Several extensions could enhance the findings of this study. For example, the choice of features in ML based reconciliation was primarily based on existing literature and prior work. Future research could explore automated feature selection techniques, deep learning feature extraction, or dynamic feature importance assessment to further enhance ML based reconciliation. Additionally, our machine learning based reconciliation approach was implemented using tree based models (Random Forest, XGBoost, LightGBM). However, the methodology is adaptable to other machine learning algorithms. Exploring neural network-based reconciliation approaches could provide additional insights into the flexibility of ML based reconciliation.

Moreover, future works could also explore real-time reconciliation techniques, including online learning methods that adapt dynamically as new data becomes available, reducing the need for full model retraining and improving computational efficiency.

# References

- [1] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. OTexts, 2018.
- [2] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, “Optimal combination forecasts for hierarchical time series,” *Computational Statistics Data Analysis*, vol. 55, no. 9, pp. 2579–2589, 2011.
- [3] J. Rombouts, M. Ternes, and I. Wilms, “Cross-temporal forecast reconciliation at digital platforms with machine learning,” *International Journal of Forecasting*, vol. 41, no. 1, pp. 321–344, 2025.
- [4] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and A. Panagiotelis, “Forecast reconciliation: A review,” *International Journal of Forecasting*, vol. 40, no. 2, pp. 430–456, 2024.
- [5] R. Hollyman, F. Petropoulos, and M. E. Tipping, “Understanding forecast reconciliation,” *European Journal of Operational Research*, vol. 294, no. 1, pp. 149–160, 2021.
- [6] T. Di Fonzo and D. Girolimetto, “Forecast combination-based forecast reconciliation: Insights and extensions,” *International Journal of Forecasting*, vol. 40, no. 2, pp. 490–514, 2024.
- [7] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and A. Panagiotelis, “Editorial: Innovations in hierarchical forecasting,” *International Journal of Forecasting*, vol. 40, no. 2, pp. 427–429, 2024.
- [8] X. Wang, R. J. Hyndman, and S. L. Wickramasuriya, “Optimal forecast reconciliation with time series selection,” *European Journal of Operational Research*, 2024.
- [9] L. Zambon, D. Azzimonti, N. Rubattu, and G. Corani, “Probabilistic reconciliation of mixed-type hierarchical time series,” in *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024.

- [10] R. Mattera, G. Athanasopoulos, and R. Hyndman, “Improving out-of-sample forecasts of stock price indexes with forecast reconciliation and clustering,” *Quantitative Finance*, vol. 24, no. 11, pp. 1641–1667, 2024.
- [11] S. L. Wickramasuriya, G. Athanasopoulos, and R. J. Hyndman, “Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization,” *Journal of the American Statistical Association*, vol. 114, no. 526, pp. 804–819, 2019.
- [12] R. J. Hyndman, A. J. Lee, and E. Wang, “Fast computation of reconciled forecasts for hierarchical and grouped time series,” *Computational Statistics Data Analysis*, vol. 97, pp. 16–32, 2016.
- [13] A. Panagiotelis, G. Athanasopoulos, P. Gamakumara, and R. J. Hyndman, “Forecast reconciliation: A geometric view with new insights on bias correction,” *International Journal of Forecasting*, vol. 37, no. 1, pp. 343–359, 2021.
- [14] T. van Erven and J. Cugliari, “Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts,” in *Modeling and Stochastic Learning for Forecasting in High Dimensions*, Cham, 2015.
- [15] D. Girolimetto and T. Di Fonzo, “Point and probabilistic forecast reconciliation for general linearly constrained multiple time series,” *Statistical Methods Applications*, vol. 33, no. 2, pp. 581–607, 2023.
- [16] E. Spiliotis, M. Abolghasemi, R. J. Hyndman, F. Petropoulos, and V. Assimakopoulos, “Hierarchical forecast reconciliation with machine learning,” *Applied Soft Computing*, vol. 112, pp. 1568–4946, 2021.
- [17] M. Anderer and F. Li, “Hierarchical forecasting with a top-down alignment of independent-level forecasts,” *International Journal of Forecasting*, vol. 38, no. 4, pp. 1405–1414, 2022.
- [18] M. Abolghasemi, G. Tarr, and C. Bergmeir, “Machine learning applications in hierarchical time series forecasting: Investigating the impact of promotions,” *International Journal of Forecasting*, vol. 40, no. 2, pp. 597–615, 2024.
- [19] F. Theodosiou and N. Kourentzes, “Forecasting with deep temporal hierarchies,” *Available at SSRN 3918315*, 2021.

- [20] S. Wang, F. Zhou, Y. Sun, L. Ma, J. Zhang, and Y. Zheng, “End-to-end modeling of hierarchical time series using autoregressive transformer and conditional normalizing flow-based reconciliation,” in *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2022.
- [21] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, K. Kuroptev, M. O’Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmeen, F. Garza, D. Girolimetto, R. Ihaka, R Core Team, D. Reid, D. Shaub, Y. Tang, X. Wang, and Z. Zhou, “Forecasting functions for time series and linear models,” <https://CRAN.R-project.org/package=forecast>, accessed: 20.06.2024.
- [22] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, “A state space framework for automatic forecasting using exponential smoothing methods,” *International Journal of forecasting*, vol. 18, no. 3, pp. 439–454, 2002.
- [23] R. J. Hyndman and Y. Khandakar, “Automatic time series forecasting: The forecast package for r,” *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.
- [24] B. Rostami-Tabar and R. J. Hyndman, “Hierarchical time series forecasting in emergency medical services,” *Journal of Service Research*, 2024.
- [25] G. Elliott and A. Timmerman, *Economic Forecasting*. Princeton University Press, Princeton, New Jersey, 2016.
- [26] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “M5 accuracy competition: Results, findings, and conclusions,” *International Journal of Forecasting*, vol. 38, no. 4, pp. 1346–1364, 2022a.
- [27] T. Januschowski, Y. Wang, K. Torkkola, T. Erkkilä, H. Hasson, and J. Gasthaus, “Forecasting with trees,” *International Journal of Forecasting*, vol. 38, no. 4, pp. 1473–1481, 2022.
- [28] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [29] A. Liaw and M. Wiener, “Forecasting with trees,” *R News*, vol. 2, pp. 18–22, 2002.
- [30] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, Aug. 2016.

- [31] Y. Shi, G. Ke, D. Soukhavong, J. Lamb, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, N. Titov, Y. Yan, Microsoft Corporation, Dropbox, Inc., A. Ferreira, D. Lemire, V. Zverovich, IBM Corporation, D. Cortes, and M. Mayer, “Lightgbm: Light gradient boosting machine,” <https://CRAN.R-project.org/package=lightgbm>, accessed: 26.07.2024.
- [32] T. Di Fonzo and D. Girolimetto, “Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives,” *International Journal of Forecasting*, vol. 39, no. 1, pp. 39–57, 2023.
- [33] R. C. Team, “R: A language and environment for statistical computing. vienna, austria: R foundation for statistical computing,” <https://www.r-project.org/>, accessed: 16.02.2023.
- [34] N. Kourentzes and G. Athanasopoulos, “Cross-temporal coherent forecasts for australian tourism,” *Annals of Tourism Research, Elsevier*, vol. 75, no. C, pp. 393–409, 2019.
- [35] T. Di Fonzo and D. Girolimetto, “Spatio-temporal reconciliation of solar forecasts,” *Solar Energy*, vol. 251, pp. 13–29, 2023.
- [36] D. Girolimetto and T. Di Fonzo, “Foreco: Forecast reconciliation,” <https://cran.r-project.org/package=FoReco>, accessed: 20.08.2024.
- [37] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and F. Petropoulos, “Forecasting with temporal hierarchies,” *European Journal of Operational Research*, vol. 262, no. 1, pp. 60–74, 2017.
- [38] P. Nystrup, E. Lindström, P. Pinson, and H. Madsen, “Temporal hierarchies with autocorrelation for load forecasting,” *European Journal of Operational Research*, vol. 280, no. 3, pp. 876–888, 2020.
- [39] J. Hoover, “Measuring forecast accuracy: Omissions in today’s forecasting engines and demand-planning software,” *Foresight: The International Journal of Applied Forecasting*, no. 4, pp. 32–35, 2006.
- [40] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

# Acknowledgments

I would like to express my deepest gratitude to my supervisor, Professor Luisa Bisaglia, for her continuous support and guidance throughout my thesis. Her encouragement and valuable insights have been instrumental in shaping this work, and I am truly grateful for her patience and dedication.

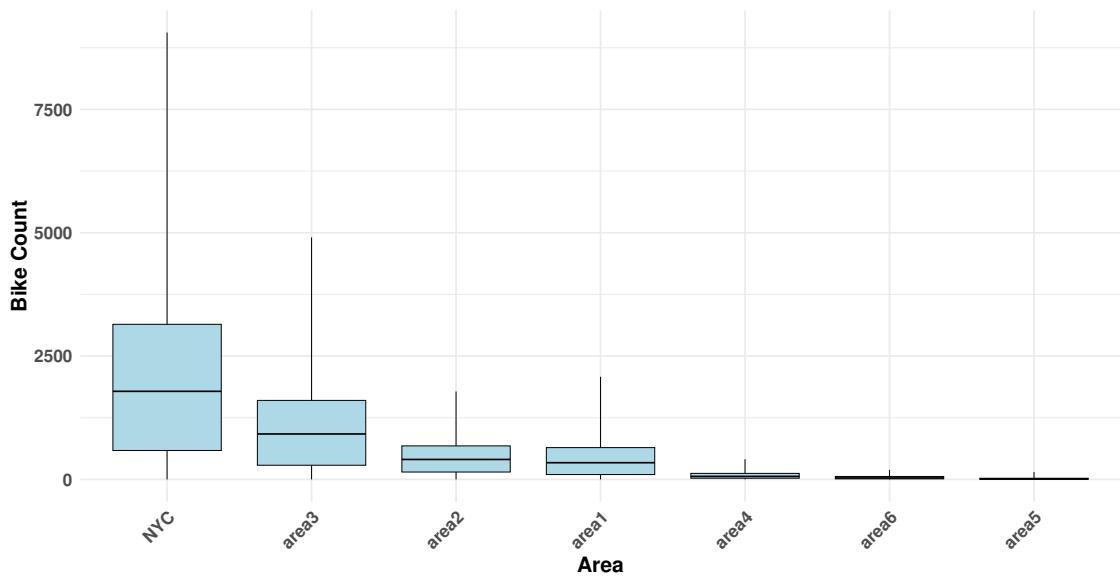
I also extend my sincere appreciation to my co-supervisor, Dr. Daniele Girolimetto, for his constructive feedback and invaluable suggestions. His support and commitment have greatly contributed to the development of this thesis, and I deeply appreciate his guidance throughout this process.

I am truly grateful for their expertise, technical guidance, and willingness to share their knowledge to assist me throughout this journey.

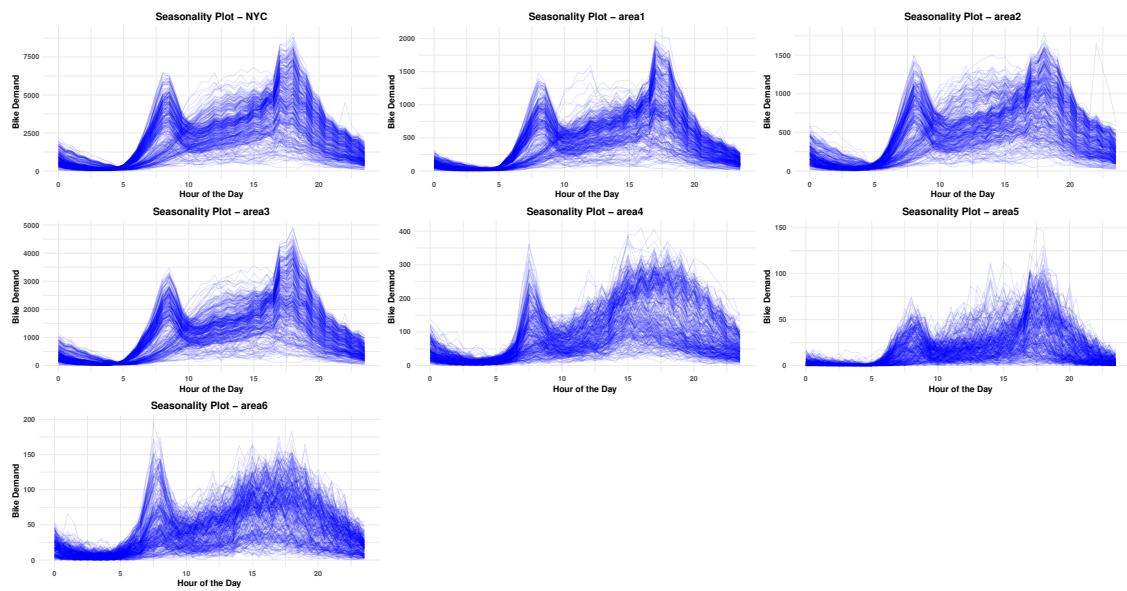


# A

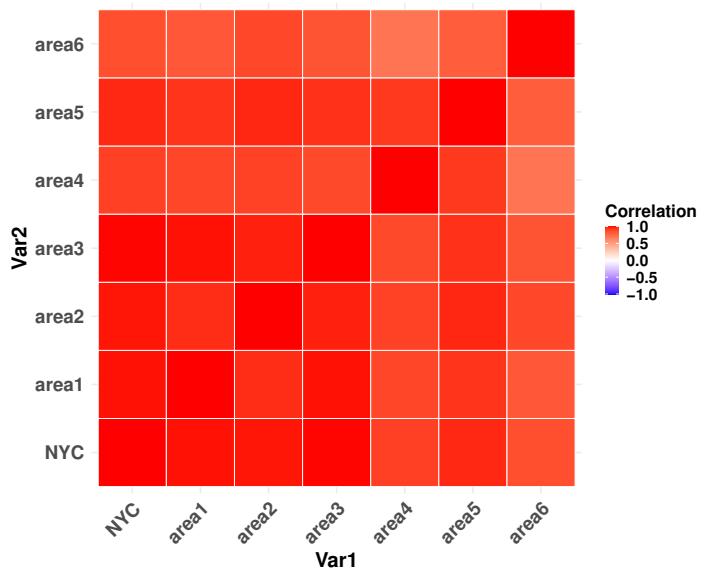
## Additional Figures for Describing Citi Bike Dataset



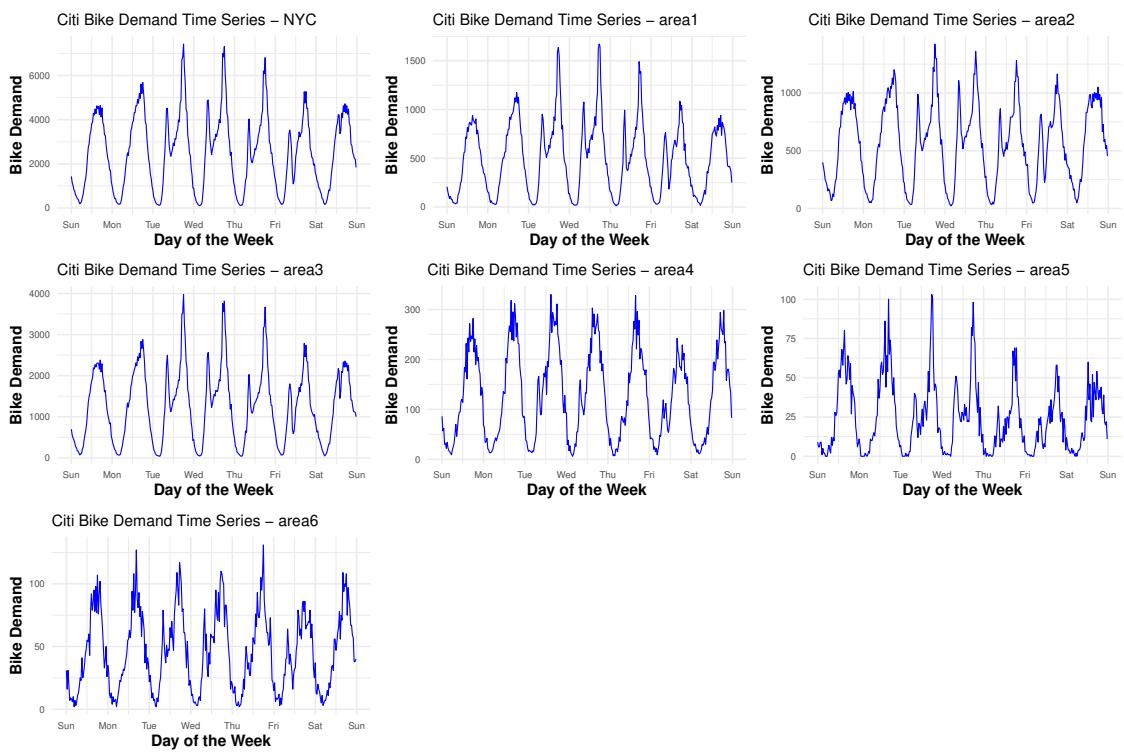
**Figure A.1:** Boxplot of Citi Bike data by area.



**Figure A.2:** Hourly bike demand variability across different regions of NYC.



**Figure A.3:** Correlation matrix of Citi Bike.

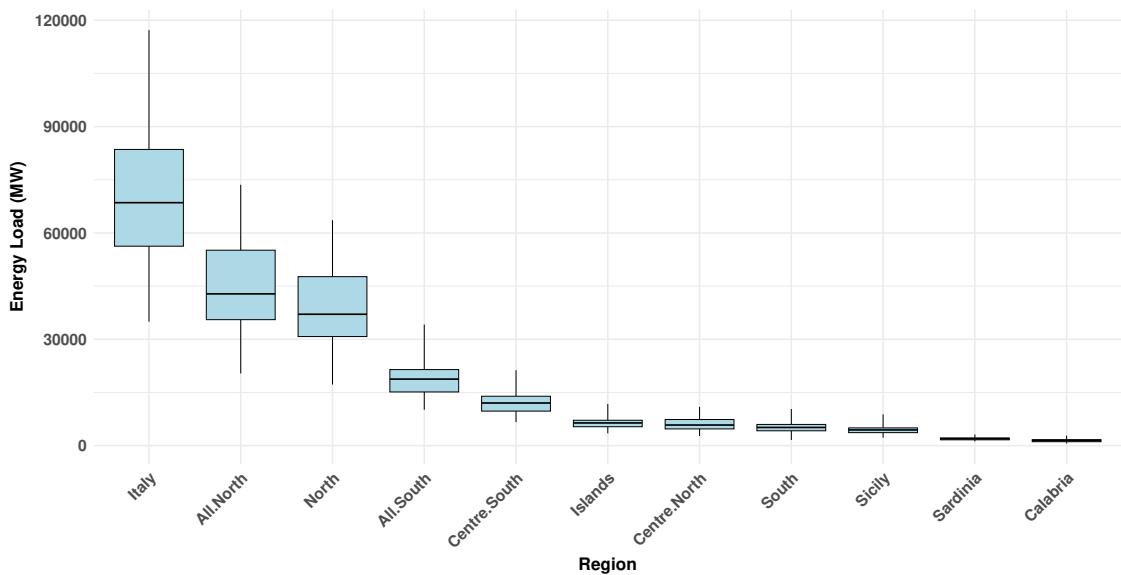


**Figure A.4:** Daily Bike demand patterns across different regions of NYC (January 1–7, 2023).

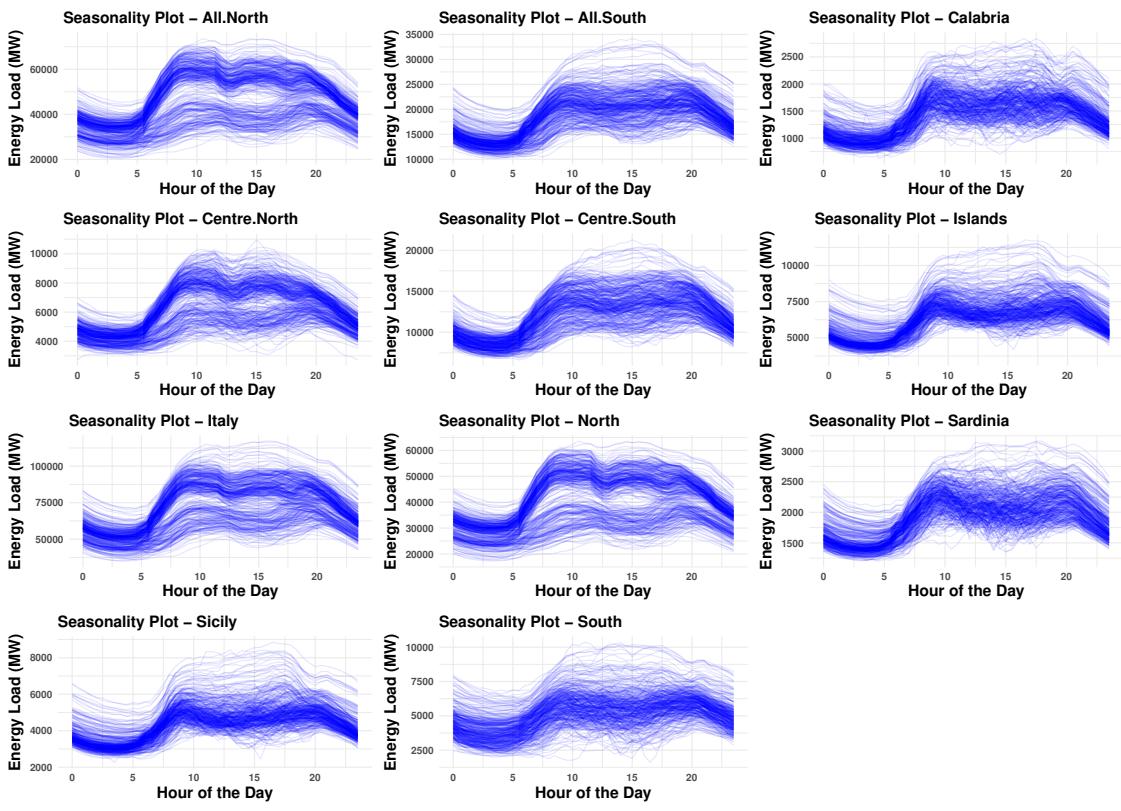


# B

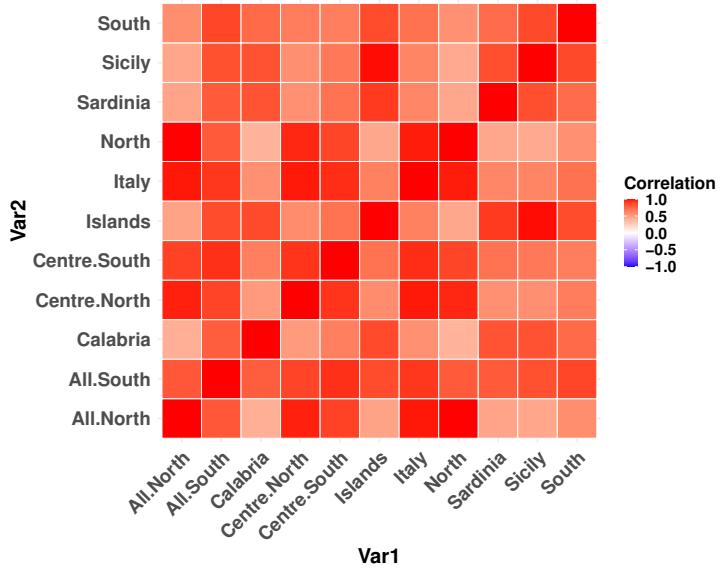
## Additional Figures for Describing Energy Load Dataset



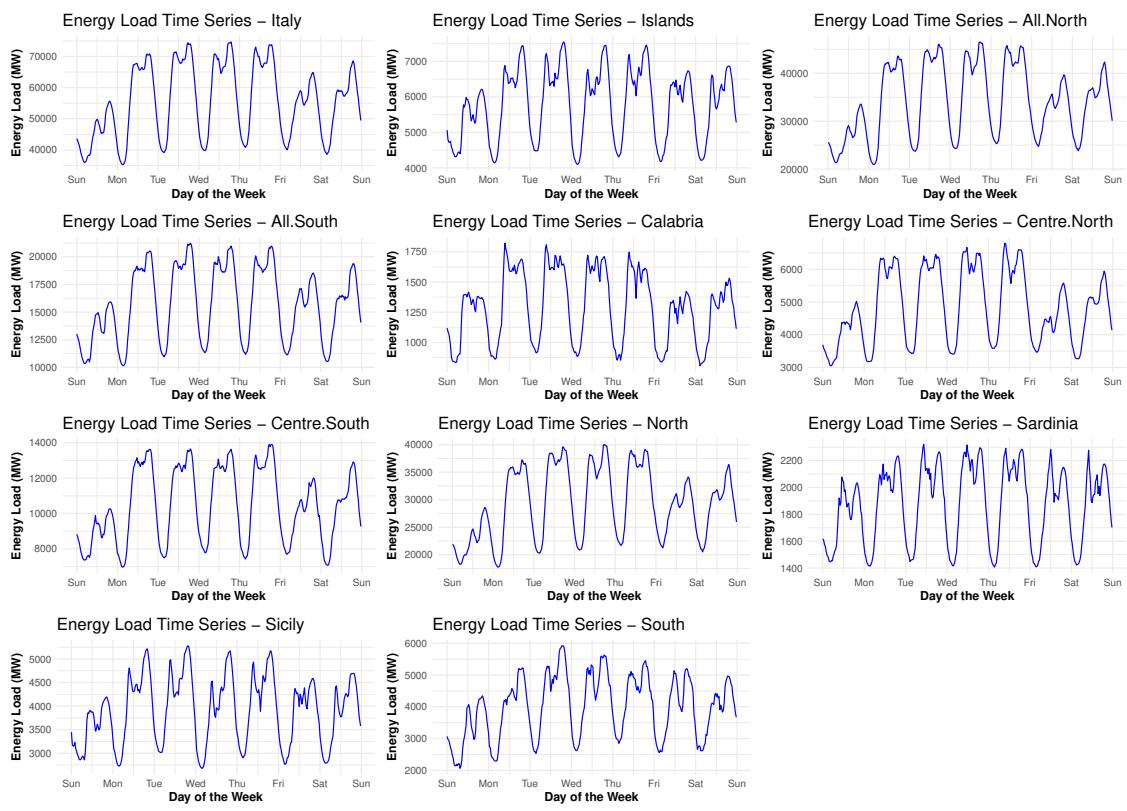
**Figure B.1:** Boxplot of Energy Load data by region.



**Figure B.2:** Hourly Energy Load variability across different regions of Italy.



**Figure B.3:** Correlation matrix of Energy Load.



**Figure B.4:** Daily Energy Load patterns across different regions of Italy (January 1–7, 2023).





## MASE Accuracy Index for Citi Bike Forecasts

This appendix provides the detailed forecast results for the Citi Bike dataset, focusing on the Mean Absolute Scaled Error (MASE) accuracy index. Tables C.1 to C.4 present the results for the H<sub>3</sub> Cells using various base methods: Naive, ETS, SARIMA, and Forecast Combination. Tables C.5 to C.8 report the results for the Citi Bike Market using the same base methods.

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	1.0521	1.0544	1.0573	1.0596	1.0612	1.0677	1.0693	1.0756	1.0838	1.0894
	Randomforest	0.9883	1.0030	1.0099	1.0112	1.0135	1.0151	1.0032	0.9886	1.0017	0.9797
	XGBoost	1.0878	1.0996	1.1018	1.1025	1.0979	1.0969	1.0798	1.0559	1.0682	1.0334
	LightGBM	1.0653	1.0816	1.0862	1.0895	1.0873	1.0913	1.0694	1.0577	1.0654	1.0302

**Table C.1:** Citi Bike H<sub>3</sub> Cells Forecast Results of MASE index for Naive Base Method.  
 (Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	1.4020	1.3554	1.3757	1.3809	1.3639	1.2197	1.1765	1.1796	1.1408	0.8585	
	Bottom Up	1.4020	1.4561	1.4847	1.5013	1.5163	1.5216	1.5216	1.5272	1.5378	1.4401	
tcs	<i>wlsv-shr</i>	1.2337	1.2738	1.2949	1.3052	1.3058	1.2855	1.2822	1.2537	1.2201	1.0551	
	<i>sari-shr</i>	1.1756	1.2113	1.2273	1.2344	1.2268	1.1964	1.1749	1.1600	1.1020	0.8710	
	<i>acov-shr</i>	0.9212	0.9419	0.9528	0.9582	0.9669	0.9607	0.9523	0.9564	0.9415	0.8548	
	<i>acov-wls</i>	0.9148	0.9353	0.9461	0.9513	0.9598	0.9544	0.9459	0.9504	0.9363	0.8524	
	<i>sari-wls</i>	1.1765	1.2124	1.2284	1.2356	1.2280	1.1978	1.1761	1.1615	1.1031	0.8715	
cst	<i>wlsv-shr</i>	1.2231	1.2616	1.2815	1.2908	1.2872	1.2654	1.2559	1.2287	1.1890	1.0202	
	<i>sari-shr</i>	1.1765	1.2119	1.2278	1.2345	1.2260	1.1955	1.1708	1.1554	1.0968	0.8643	
	<i>acov-shr</i>	1.0365	1.0649	1.0780	1.0858	1.0876	1.0840	1.0704	1.0749	1.0375	0.8746	
	<i>acov-wls</i>	1.0155	1.0436	1.0566	1.0644	1.0680	1.0661	1.0542	1.0582	1.0270	0.8699	
	<i>sari-wls</i>	1.1796	1.2158	1.2320	1.2396	1.2322	1.2017	1.1810	1.1637	1.1043	0.8736	
ETS	ite	<i>wlsv-shr</i>	1.2213	1.2598	1.2794	1.2886	1.2846	1.2625	1.2524	1.2264	1.1861	1.0184
		<i>sari-shr</i>	1.1732	1.2083	1.2240	1.2306	1.2223	1.1916	1.1672	1.1538	1.0962	0.8642
		<i>acov-wls</i>	0.9174	0.9380	0.9488	0.9541	0.9626	0.9573	0.9484	0.9533	0.9386	0.8535
		<i>sari-wls</i>	1.1766	1.2124	1.2285	1.2357	1.2281	1.1978	1.1762	1.1615	1.1032	0.8715
	oct	<i>wlsv</i>	1.2362	1.2766	1.2979	1.3085	1.3097	1.2896	1.2870	1.2583	1.2254	1.0613
		<i>bdschr</i>	1.2047	1.2411	1.2601	1.2663	1.2603	1.2344	1.2173	1.2008	1.1595	0.9744
		<i>acov</i>	0.9094	0.9277	0.9383	0.9428	0.9507	0.9462	0.9367	0.9411	0.9303	0.8552
		<i>str</i>	1.2051	1.2423	1.2591	1.2643	1.2645	1.2299	1.2261	1.1997	1.1450	0.9471
		<i>Sshr</i>	0.9060	0.9217	0.9299	0.9339	0.9413	0.9402	0.9315	0.9411	0.9319	0.8849
	Randomforest	0.9687	0.9795	0.9838	0.9832	0.9793	0.9702	0.9458	0.9496	0.9462	0.9296	
	XGBoost	1.0483	1.0571	1.0548	1.0538	1.0473	1.0408	1.0111	1.0111	1.0066	0.9757	
	LightGBM	1.0193	1.0326	1.0339	1.0350	1.0296	1.0274	0.9964	0.9918	0.9960	0.9642	

**Table C.2:** Citi Bike H3 Cells Forecast Results of MASE index for ETS Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	1.2908	1.2452	1.2528	1.2680	1.1935	1.0547	0.9969	0.9857	0.8651	0.8590
	Bottom Up	1.2908	1.3389	1.3626	1.3761	1.3778	1.3732	1.3529	1.3471	1.3231	1.2031
tcs	<i>wlsv-shr</i>	1.1438	1.1730	1.1865	1.1892	1.1739	1.1394	1.1126	1.0939	1.0464	0.9101
	<i>sari-shr</i>	1.0853	1.1088	1.1183	1.1194	1.1030	1.0591	1.0281	1.0126	0.9594	0.8398
	<i>acov-shr</i>	0.9080	0.9210	0.9276	0.9277	0.9335	0.9181	0.9041	0.9019	0.8821	0.8365
	<i>acov-wls</i>	0.9093	0.9226	0.9294	0.9298	0.9358	0.9205	0.9075	0.9044	0.8853	0.8399
	<i>sari-wls</i>	1.0849	1.1083	1.1178	1.1189	1.1022	1.0583	1.0277	1.0119	0.9588	0.8401
cst	<i>wlsv-shr</i>	1.1396	1.1685	1.1812	1.1832	1.1669	1.1318	1.1000	1.0853	1.0363	0.8984
	<i>sari-shr</i>	1.0840	1.1073	1.1164	1.1170	1.0997	1.0561	1.0232	1.0088	0.9543	0.8342
	<i>acov-shr</i>	0.9897	1.0094	1.0180	1.0204	1.0186	0.9946	0.9764	0.9744	0.9309	0.8396
	<i>acov-wls</i>	0.9845	1.0046	1.0132	1.0158	1.0165	0.9968	0.9818	0.9797	0.9379	0.8453
	<i>sari-wls</i>	1.0867	1.1104	1.1199	1.1212	1.1042	1.0601	1.0295	1.0126	0.9585	0.8401
SARIMA											
ite	<i>wlsv-shr</i>	1.1384	1.1671	1.1797	1.1816	1.1651	1.1295	1.0975	1.0829	1.0341	0.8967
	<i>sari-shr</i>	1.0824	1.1054	1.1144	1.1149	1.0979	1.0546	1.0221	1.0084	0.9550	0.8350
	<i>acov-wls</i>	0.9123	0.9257	0.9323	0.9331	0.9390	0.9230	0.9105	0.9070	0.8874	0.8414
	<i>sari-wls</i>	1.0850	1.1083	1.1178	1.1189	1.1022	1.0583	1.0277	1.0119	0.9589	0.8401
oct	<i>wlsv</i>	1.1440	1.1732	1.1867	1.1896	1.1744	1.1404	1.1143	1.0950	1.0482	0.9131
	<i>bdsr</i>	1.1312	1.1585	1.1713	1.1711	1.1547	1.1165	1.0829	1.0725	1.0238	0.8860
	<i>acov</i>	0.9167	0.9287	0.9344	0.9344	0.9394	0.9262	0.9095	0.9082	0.8897	0.8466
	<i>str</i>	1.1359	1.1635	1.1721	1.1700	1.1506	1.1044	1.0762	1.0581	1.0045	0.8817
	<i>Sshr</i>	0.9314	0.9448	0.9493	0.9536	0.9566	0.9503	0.9392	0.9427	0.9288	0.9031
Randomforest											
XGBoost		0.9676	0.9781	0.9799	0.9806	0.9732	0.9557	0.9252	0.9287	0.9193	0.8939
		1.0383	1.0474	1.0427	1.0418	1.0302	1.0220	0.9874	0.9882	0.9876	0.9485
LightGBM											
		1.0073	1.0187	1.0189	1.0188	1.0094	1.0012	0.9668	0.9679	0.9636	0.9246

**Table C.3:** Citi Bike H3 Cells Forecast Results of MASE index for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	1.0942	1.0615	1.0664	1.0744	1.0440	0.9599	0.9541	0.9418	0.8970	0.8460
	Bottom Up	1.0942	1.1268	1.1424	1.1536	1.1631	1.1600	1.1551	1.1558	1.1574	1.0833
Forecast Combination	tcs <i>wlsv-shr</i>	1.0173	1.0402	1.0507	1.0560	1.0561	1.0396	1.0317	1.0223	1.0003	0.9080
	<i>sari-shr</i>	0.9876	1.0078	1.0165	1.0193	1.0159	0.9956	0.9831	0.9731	0.9411	0.8389
	<i>acov-shr</i>	0.8785	0.8890	0.8946	0.8974	0.9023	0.8986	0.8907	0.8928	0.8834	0.8271
	<i>acov-wls</i>	0.8778	0.8882	0.8939	0.8966	0.9015	0.8980	0.8898	0.8919	0.8830	0.8278
	<i>sari-wls</i>	0.9876	1.0078	1.0166	1.0193	1.0160	0.9957	0.9831	0.9731	0.9409	0.8388
	cst <i>wlsv-shr</i>	1.0137	1.0360	1.0464	1.0511	1.0505	1.0327	1.0225	1.0143	0.9895	0.8973
	<i>sari-shr</i>	0.9854	1.0052	1.0136	1.0164	1.0129	0.9922	0.9788	0.9694	0.9361	0.8350
	<i>acov-shr</i>	0.9185	0.9334	0.9398	0.9437	0.9454	0.9393	0.9305	0.9336	0.9103	0.8282
	<i>acov-wls</i>	0.9145	0.9296	0.9362	0.9405	0.9425	0.9387	0.9310	0.9347	0.9139	0.8311
	<i>sari-wls</i>	0.9873	1.0075	1.0162	1.0191	1.0159	0.9950	0.9833	0.9725	0.9397	0.8397
Forecast Combination	ite <i>wlsv-shr</i>	1.0135	1.0356	1.0460	1.0506	1.0500	1.0321	1.0218	1.0137	0.9889	0.8965
	<i>sari-shr</i>	0.9858	1.0057	1.0142	1.0169	1.0132	0.9931	0.9792	0.9703	0.9379	0.8349
	<i>acov-wls</i>	0.8789	0.8894	0.8952	0.8978	0.9028	0.8991	0.8911	0.8930	0.8838	0.8284
	<i>sari-wls</i>	0.9876	1.0078	1.0166	1.0193	1.0160	0.9957	0.9831	0.9731	0.9409	0.8388
	oct <i>wlsv</i>	1.0178	1.0407	1.0513	1.0566	1.0567	1.0405	1.0328	1.0232	1.0013	0.9095
	<i>bdshb</i>	1.0085	1.0297	1.0399	1.0435	1.0424	1.0238	1.0129	1.0055	0.9812	0.8856
	<i>acov</i>	0.8796	0.8894	0.8942	0.8965	0.9011	0.8974	0.8884	0.8923	0.8830	0.8332
	<i>str</i>	1.0009	1.0217	1.0305	1.0320	1.0288	1.0053	0.9963	0.9846	0.9543	0.8568
	<i>Sshr</i>	0.8849	0.8955	0.8996	0.9022	0.9071	0.9053	0.9011	0.9082	0.8992	0.8785
	Randomforest	0.9312	0.9412	0.9450	0.9480	0.9484	0.9456	0.9280	0.9357	0.9353	0.9173
	XGBoost	1.0171	1.0258	1.0244	1.0247	1.0186	1.0195	0.9982	1.0034	1.0051	0.9798
	LightGBM	0.9961	1.0098	1.0105	1.0133	1.0114	1.0134	0.9904	1.0006	1.0046	0.9770

**Table C.4:** Citi Bike H3 Cells Forecast Results of MASE index for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	1.0962	1.0970	1.0989	1.1011	1.1026	1.1088	1.1124	1.1226	1.1300	1.1368
	Randomforest	1.0760	1.0720	1.0706	1.0674	1.0645	1.0613	1.0489	1.0318	1.0370	1.0186
	XGBoost	1.1294	1.1222	1.1169	1.1119	1.1034	1.0988	1.0868	1.0664	1.0646	1.0399
	LightGBM	1.1380	1.1321	1.1290	1.1270	1.1194	1.1211	1.1018	1.0822	1.0809	1.0497

**Table C.5:** Citi Bike Market Forecast Results of MASE accuracy index for Naive Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
Base	Base	1.8760	1.5007	1.4829	1.4706	1.4051	1.2403	1.1622	1.1718	1.1097	0.8897	
Bottom Up	Bottom Up	1.6934	1.7012	1.7100	1.7126	1.7005	1.6991	1.6914	1.6866	1.7156	1.6726	
tcs	wlsv-shr	1.4007	1.3981	1.3980	1.3950	1.3674	1.3306	1.3258	1.2623	1.2155	1.1069	
	sari-shr	1.3158	1.3098	1.3047	1.3021	1.2633	1.2137	1.1838	1.1519	1.0709	0.8837	
	acov-shr	0.9643	0.9664	0.9676	0.9659	0.9615	0.9547	0.9435	0.9432	0.9226	0.8760	
	acov-wls	0.9526	0.9555	0.9568	0.9551	0.9498	0.9444	0.9331	0.9325	0.9151	0.8746	
	sari-wls	1.3175	1.3116	1.3066	1.3040	1.2654	1.2160	1.1855	1.1540	1.0726	0.8842	
cst	wlsv-shr	1.3845	1.3788	1.3787	1.3739	1.3424	1.3058	1.2923	1.2309	1.1749	1.0731	
	sari-shr	1.3165	1.3100	1.3049	1.3025	1.2625	1.2142	1.1804	1.1482	1.0669	0.8862	
	acov-shr	1.0962	1.0916	1.0920	1.0905	1.0738	1.0635	1.0492	1.0421	0.9953	0.8855	
	acov-wls	1.0593	1.0565	1.0575	1.0560	1.0424	1.0335	1.0240	1.0129	0.9794	0.8821	
	sari-wls	1.3204	1.3150	1.3106	1.3086	1.2707	1.2222	1.1967	1.1602	1.0783	0.8998	
ETS	ite	wlsv-shr	1.3791	1.3734	1.3729	1.3676	1.3353	1.2975	1.2833	1.2243	1.1669	1.0596
		sari-shr	1.3125	1.3055	1.3002	1.2977	1.2577	1.2085	1.1730	1.1459	1.0651	0.8772
		acov-wls	0.9566	0.9593	0.9607	0.9589	0.9538	0.9484	0.9363	0.9368	0.9175	0.8755
		sari-wls	1.3175	1.3116	1.3066	1.3041	1.2654	1.2161	1.1855	1.1540	1.0726	0.8843
	oct	wlsv	1.4054	1.4031	1.4033	1.4005	1.3737	1.3371	1.3326	1.2693	1.2241	1.1146
	bdshr	1.3585	1.3521	1.3508	1.3416	1.3055	1.2643	1.2417	1.1982	1.1414	1.0133	
	acov	0.9464	0.9494	0.9502	0.9481	0.9430	0.9386	0.9274	0.9261	0.9124	0.8830	
	str	1.3467	1.3399	1.3363	1.3282	1.2951	1.2398	1.2304	1.1795	1.1035	0.9588	
	Sshr	0.9511	0.9526	0.9546	0.9528	0.9509	0.9501	0.9467	0.9451	0.9427	0.9058	
Randomforest		1.0518	1.0400	1.0360	1.0283	1.0126	0.9974	0.9691	0.9593	0.9691	0.9530	
XGBoost		1.0877	1.0765	1.0701	1.0626	1.0459	1.0359	1.0060	0.9939	1.0071	0.9792	
LightGBM		1.0722	1.0622	1.0561	1.0525	1.0341	1.0281	0.9942	0.9873	1.0004	0.9863	

**Table C.6:** Citi Bike Market Forecast Results of MASE index for ETS Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	1.4241	1.3784	1.3821	1.3950	1.2470	1.0891	1.0604	1.0529	0.9037	0.9511
	Bottom Up	1.3988	1.3967	1.3964	1.3937	1.3598	1.3250	1.3018	1.2380	1.2006	1.1089
tcs	<i>wlsv-shr</i>	1.2896	1.2837	1.2794	1.2705	1.2271	1.1785	1.1423	1.1072	1.0439	0.9251
	<i>sari-shr</i>	1.2255	1.2175	1.2112	1.2058	1.1629	1.1094	1.0702	1.0509	0.9786	0.8680
	<i>acov-shr</i>	0.9612	0.9634	0.9636	0.9608	0.9544	0.9386	0.9241	0.9224	0.9003	0.8618
	<i>acov-wls</i>	0.9611	0.9634	0.9636	0.9610	0.9546	0.9389	0.9257	0.9238	0.9019	0.8644
	<i>sari-wls</i>	1.2252	1.2171	1.2109	1.2056	1.1622	1.1088	1.0700	1.0503	0.9782	0.8682
cst	<i>wlsv-shr</i>	1.2844	1.2782	1.2734	1.2647	1.2194	1.1713	1.1266	1.0977	1.0320	0.9092
	<i>sari-shr</i>	1.2237	1.2154	1.2092	1.2032	1.1597	1.1092	1.0670	1.0504	0.9774	0.8665
	<i>acov-shr</i>	1.0662	1.0631	1.0611	1.0584	1.0396	1.0127	0.9912	0.9878	0.9373	0.8675
	<i>acov-wls</i>	1.0586	1.0560	1.0544	1.0521	1.0357	1.0116	0.9945	0.9911	0.9421	0.8741
	<i>sari-wls</i>	1.2266	1.2185	1.2125	1.2071	1.1637	1.1120	1.0732	1.0540	0.9813	0.8717
SARIMA											
ite	<i>wlsv-shr</i>	1.2844	1.2782	1.2733	1.2646	1.2196	1.1705	1.1262	1.0974	1.0319	0.9088
	<i>sari-shr</i>	1.2222	1.2140	1.2075	1.2018	1.1581	1.1061	1.0641	1.0470	0.9747	0.8638
	<i>acov-wls</i>	0.9693	0.9712	0.9716	0.9690	0.9627	0.9458	0.9345	0.9302	0.9074	0.8682
	<i>sari-wls</i>	1.2252	1.2172	1.2109	1.2056	1.1622	1.1088	1.0699	1.0503	0.9782	0.8682
oct	<i>wlsv</i>	1.2894	1.2836	1.2793	1.2703	1.2273	1.1786	1.1443	1.1079	1.0456	0.9279
	<i>bdsr</i>	1.2794	1.2725	1.2678	1.2565	1.2129	1.1627	1.1145	1.0923	1.0312	0.9057
	<i>acov</i>	0.9616	0.9637	0.9629	0.9603	0.9534	0.9389	0.9219	0.9214	0.8981	0.8625
	<i>str</i>	1.2515	1.2422	1.2372	1.2255	1.1816	1.1216	1.0878	1.0657	0.9954	0.8846
	<i>Sshr</i>	0.9515	0.9516	0.9518	0.9522	0.9462	0.9402	0.9324	0.9332	0.9191	0.8852
Randomforest		1.0462	1.0320	1.0253	1.0179	0.9958	0.9751	0.9422	0.9381	0.9293	0.9102
XGBoost		1.0529	1.0382	1.0283	1.0207	1.0030	0.9932	0.9592	0.9514	0.9562	0.9238
LightGBM		1.0404	1.0276	1.0178	1.0089	0.9898	0.9786	0.9479	0.9425	0.9424	0.9158

**Table C.7:** Citi Bike Market Forecast Results of MASE index for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	1.2718	1.1534	1.1475	1.1505	1.0788	0.9933	0.9871	0.9771	0.9168	0.9045
	Bottom Up	1.2289	1.2324	1.2330	1.2357	1.2257	1.2148	1.2013	1.1833	1.1873	1.1493
Forecast Combination	tcs <i>wlsv-shr</i>	1.1313	1.1300	1.1292	1.1277	1.1117	1.0849	1.0746	1.0490	1.0161	0.9568
	<i>sari-shr</i>	1.0929	1.0917	1.0895	1.0866	1.0649	1.0354	1.0146	0.9978	0.9511	0.8703
	<i>acov-shr</i>	0.9214	0.9229	0.9230	0.9227	0.9197	0.9137	0.9050	0.9072	0.8969	0.8604
	<i>acov-wls</i>	0.9202	0.9219	0.9218	0.9218	0.9186	0.9127	0.9034	0.9064	0.8967	0.8615
	<i>sari-wls</i>	1.0930	1.0918	1.0896	1.0867	1.0651	1.0355	1.0148	0.9977	0.9507	0.8701
	cst <i>wlsv-shr</i>	1.1253	1.1239	1.1229	1.1206	1.1041	1.0759	1.0609	1.0386	1.0012	0.9425
	<i>sari-shr</i>	1.0898	1.0884	1.0861	1.0834	1.0622	1.0336	1.0126	0.9982	0.9514	0.8728
	<i>acov-shr</i>	0.9804	0.9799	0.9786	0.9787	0.9694	0.9572	0.9474	0.9454	0.9141	0.8576
	<i>acov-wls</i>	0.9722	0.9714	0.9704	0.9710	0.9623	0.9532	0.9449	0.9432	0.9171	0.8610
	<i>sari-wls</i>	1.0920	1.0909	1.0887	1.0862	1.0653	1.0360	1.0180	1.0014	0.9557	0.8777
ite	<i>wlsv-shr</i>	1.1256	1.1243	1.1233	1.1209	1.1046	1.0766	1.0616	1.0394	1.0023	0.9438
	<i>sari-shr</i>	1.0906	1.0892	1.0869	1.0841	1.0622	1.0334	1.0103	0.9953	0.9478	0.8670
	<i>acov-wls</i>	0.9221	0.9235	0.9237	0.9236	0.9200	0.9142	0.9058	0.9079	0.8980	0.8618
	<i>sari-wls</i>	1.0930	1.0918	1.0896	1.0867	1.0651	1.0356	1.0148	0.9977	0.9507	0.8702
oct	<i>wlsv</i>	1.1321	1.1307	1.1300	1.1285	1.1128	1.0860	1.0762	1.0502	1.0179	0.9581
	<i>bdshb</i>	1.1199	1.1181	1.1175	1.1138	1.0956	1.0677	1.0523	1.0328	0.9958	0.9348
	<i>acov</i>	0.9186	0.9204	0.9204	0.9206	0.9174	0.9139	0.9032	0.9068	0.8997	0.8657
	<i>str</i>	1.1023	1.0999	1.0976	1.0923	1.0719	1.0365	1.0237	1.0052	0.9613	0.8866
	<i>Sshr</i>	0.9128	0.9137	0.9142	0.9143	0.9135	0.9093	0.9092	0.9067	0.9002	0.8842
Randomforest	Randomforest	1.0025	0.9950	0.9919	0.9904	0.9827	0.9719	0.9533	0.9647	0.9557	0.9565
	XGBoost	1.0365	1.0272	1.0236	1.0203	1.0127	1.0008	0.9833	0.9909	0.9844	0.9805
	LightGBM	1.0303	1.0222	1.0171	1.0189	1.0079	1.0020	0.9786	0.9889	0.9895	0.9786

**Table C.8:** Citi Bike Market Forecast Results of MASE index for Forecast Combination Base Method. (Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)



# D

## MASE Accuracy Index for Energy Load Forecasts

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
Naive	Base	1.0268	1.0292	1.0306	1.0327	1.0356	1.0377	1.0398	1.0459	1.0518	1.0570
	Randomforest	0.8741	0.8730	0.8713	0.8715	0.8697	0.8691	0.8659	0.8697	0.8564	0.8499
	XGBoost	0.9380	0.9317	0.9268	0.9258	0.9217	0.9173	0.9136	0.9164	0.9016	0.8916
	LightGBM	0.9226	0.9198	0.9166	0.9173	0.9140	0.9121	0.9085	0.9105	0.9004	0.8899

**Table D.1:** Energy Load Zones Forecast Results of MASE index for Naive Base Method.  
(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.9948	1.0480	1.0247	1.0068	0.9696	0.6910	0.6832	0.6174	0.5920	0.5339	
	Bottom Up	0.9948	0.9969	0.9966	0.9960	0.9956	0.9930	0.9859	0.9867	0.9676	0.9434	
tcs	<i>wlsv-shr</i>	0.8108	0.8102	0.8067	0.8036	0.7989	0.7864	0.7729	0.7532	0.7237	0.6761	
	<i>sar1-shr</i>	0.7196	0.7174	0.7123	0.7065	0.6988	0.6741	0.6645	0.6169	0.5573	0.4819	
	<i>acov-shr</i>	0.5824	0.5801	0.5769	0.5746	0.5697	0.5631	0.5541	0.5410	0.5185	0.4676	
	<i>acov-wls</i>	0.5803	0.5780	0.5749	0.5727	0.5680	0.5612	0.5524	0.5397	0.5177	0.4687	
	<i>sar1-wls</i>	0.7216	0.7193	0.7143	0.7086	0.7010	0.6761	0.6671	0.6187	0.5585	0.4832	
cst	<i>wlsv-shr</i>	0.8262	0.8255	0.8227	0.8194	0.8140	0.8034	0.7893	0.7663	0.7333	0.6870	
	<i>sar1-shr</i>	0.7288	0.7268	0.7220	0.7163	0.7082	0.6854	0.6753	0.6286	0.5688	0.4962	
	<i>acov-shr</i>	0.6531	0.6502	0.6458	0.6411	0.6333	0.6228	0.6095	0.5877	0.5555	0.4939	
	<i>acov-wls</i>	0.6508	0.6485	0.6446	0.6406	0.6333	0.6250	0.6129	0.5959	0.5676	0.5068	
	<i>sar1-wls</i>	0.7320	0.7300	0.7253	0.7201	0.7121	0.6883	0.6801	0.6333	0.5758	0.5025	
ETS	ite	<i>wlsv-shr</i>	0.8027	0.8018	0.7984	0.7950	0.7895	0.7760	0.7628	0.7374	0.7016	0.6555
		<i>sar1-shr</i>	0.7180	0.7157	0.7109	0.7047	0.6964	0.6725	0.6617	0.6140	0.5520	0.4777
		<i>acov-wls</i>	0.5817	0.5795	0.5763	0.5739	0.5693	0.5622	0.5531	0.5407	0.5186	0.4692
		<i>sar1-wls</i>	0.7215	0.7193	0.7143	0.7085	0.7009	0.6759	0.6670	0.6186	0.5583	0.4830
	oct	<i>wlsv</i>	0.8129	0.8125	0.8090	0.8059	0.8012	0.7890	0.7758	0.7563	0.7277	0.6802
		<i>bdsbr</i>	0.7647	0.7632	0.7585	0.7542	0.7467	0.7316	0.7162	0.6881	0.6520	0.5932
		<i>acov</i>	0.5806	0.5782	0.5752	0.5730	0.5684	0.5629	0.5531	0.5441	0.5215	0.4801
		<i>str</i>	0.7914	0.7896	0.7855	0.7808	0.7740	0.7585	0.7438	0.7103	0.6711	0.6122
		<i>Sshbr</i>	0.5820	0.5791	0.5756	0.5735	0.5682	0.5624	0.5528	0.5435	0.5265	0.4793
	Randomforest	0.6451	0.6395	0.6321	0.6251	0.6176	0.6030	0.5883	0.5729	0.5480	0.5232	
	XGBoost	0.7029	0.6928	0.6816	0.6755	0.6669	0.6502	0.6337	0.6194	0.5931	0.5656	
	LightGBM	0.6698	0.6627	0.6531	0.6477	0.6394	0.6252	0.6106	0.5948	0.5682	0.5427	

**Table D.2:** Energy Load Zones Forecast Results of MASE index for ETS Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.9696	0.9811	0.9852	0.9893	0.9708	0.6198	0.5540	0.5568	0.5296	0.6026
	Bottom Up	0.9696	0.9714	0.9713	0.9704	0.9701	0.9649	0.9583	0.9537	0.9296	0.9088
tcs	<i>wlsv-shr</i>	0.7707	0.7691	0.7651	0.7603	0.7557	0.7411	0.7284	0.7064	0.6752	0.6491
	<i>sari-shr</i>	0.6789	0.6758	0.6700	0.6635	0.6557	0.6309	0.6220	0.5794	0.5292	0.4902
	<i>acov-shr</i>	0.5797	0.5768	0.5735	0.5712	0.5672	0.5586	0.5522	0.5383	0.5148	0.4876
	<i>acov-wls</i>	0.5799	0.5772	0.5743	0.5719	0.5679	0.5595	0.5537	0.5400	0.5167	0.4912
	<i>sari-wls</i>	0.6798	0.6768	0.6712	0.6652	0.6577	0.6323	0.6245	0.5817	0.5324	0.4941
cst	<i>wlsv-shr</i>	0.7835	0.7818	0.7776	0.7730	0.7675	0.7540	0.7409	0.7195	0.6899	0.6632
	<i>sari-shr</i>	0.6811	0.6779	0.6721	0.6657	0.6576	0.6337	0.6242	0.5819	0.5359	0.5007
	<i>acov-shr</i>	0.6252	0.6216	0.6168	0.6120	0.6050	0.5949	0.5826	0.5648	0.5354	0.5022
	<i>acov-wls</i>	0.6324	0.6296	0.6253	0.6208	0.6143	0.6048	0.5928	0.5767	0.5494	0.5151
	<i>sari-wls</i>	0.6869	0.6839	0.6785	0.6723	0.6650	0.6410	0.6321	0.5915	0.5473	0.5101
SARIMA											
ite	<i>wlsv-shr</i>	0.7625	0.7604	0.7557	0.7507	0.7450	0.7297	0.7157	0.6924	0.6601	0.6326
	<i>sari-shr</i>	0.6744	0.6711	0.6650	0.6588	0.6502	0.6251	0.6163	0.5724	0.5199	0.4814
	<i>acov-wls</i>	0.5811	0.5784	0.5756	0.5731	0.5692	0.5603	0.5545	0.5403	0.5169	0.4913
	<i>sari-wls</i>	0.6797	0.6767	0.6711	0.6651	0.6576	0.6322	0.6243	0.5815	0.5322	0.4938
oct	<i>wlsv</i>	0.7764	0.7751	0.7713	0.7669	0.7631	0.7481	0.7364	0.7147	0.6853	0.6594
	<i>bdskr</i>	0.7317	0.7289	0.7233	0.7175	0.7101	0.6931	0.6801	0.6519	0.6197	0.5850
	<i>acov</i>	0.5882	0.5857	0.5825	0.5805	0.5764	0.5681	0.5624	0.5511	0.5284	0.5042
	<i>str</i>	0.7514	0.7486	0.7436	0.7370	0.7298	0.7127	0.6982	0.6656	0.6303	0.6017
	<i>Sshr</i>	0.5795	0.5764	0.5729	0.5703	0.5650	0.5595	0.5505	0.5459	0.5242	0.4979
ML Benchmarks											
Randomforest		0.6521	0.6465	0.6389	0.6318	0.6254	0.6104	0.5955	0.5810	0.5636	0.5437
XGBoost		0.7097	0.7006	0.6896	0.6834	0.6747	0.6595	0.6409	0.6303	0.6077	0.5840
LightGBM		0.6750	0.6680	0.6588	0.6534	0.6450	0.6328	0.6155	0.6040	0.5819	0.5583

**Table D.3:** Energy Load Zones Forecast Results of MASE index for SARIMA Base Method.  
 (Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	0.8288	0.8445	0.8302	0.8270	0.8102	0.6395	0.6229	0.6170	0.5995	0.6396	
	Bottom Up	0.8288	0.8297	0.8289	0.8282	0.8279	0.8240	0.8175	0.8141	0.8050	0.7989	
tcs	wlsv-shr	0.7429	0.7424	0.7404	0.7388	0.7360	0.7280	0.7196	0.7071	0.6910	0.6804	
	sari-shr	0.6819	0.6802	0.6774	0.6749	0.6702	0.6563	0.6511	0.6273	0.6000	0.5801	
	acov-shr	0.6295	0.6274	0.6256	0.6241	0.6206	0.6161	0.6095	0.6008	0.5880	0.5728	
	acov-wls	0.6273	0.6253	0.6236	0.6221	0.6188	0.6137	0.6071	0.5982	0.5858	0.5702	
	sari-wls	0.6832	0.6815	0.6789	0.6764	0.6718	0.6580	0.6528	0.6298	0.6026	0.5823	
cst	wlsv-shr	0.7501	0.7496	0.7475	0.7460	0.7427	0.7356	0.7275	0.7142	0.6990	0.6888	
	sari-shr	0.6858	0.6841	0.6813	0.6790	0.6742	0.6616	0.6555	0.6338	0.6080	0.5912	
	acov-shr	0.6535	0.6515	0.6490	0.6468	0.6422	0.6360	0.6273	0.6158	0.6004	0.5821	
	acov-wls	0.6546	0.6529	0.6506	0.6487	0.6445	0.6387	0.6298	0.6196	0.6046	0.5861	
	sari-wls	0.6897	0.6881	0.6855	0.6834	0.6787	0.6661	0.6604	0.6396	0.6147	0.5975	
Forecast Combination	ite	wlsv-shr	0.7397	0.7390	0.7368	0.7352	0.7317	0.7236	0.7150	0.7007	0.6845	0.6729
	sari-shr	0.6790	0.6772	0.6743	0.6718	0.6668	0.6530	0.6476	0.6234	0.5950	0.5753	
	acov-wls	0.6270	0.6251	0.6234	0.6218	0.6185	0.6134	0.6066	0.5976	0.5851	0.5694	
	sari-wls	0.6832	0.6815	0.6788	0.6764	0.6718	0.6580	0.6527	0.6297	0.6025	0.5823	
	oct	wlsv	0.7428	0.7423	0.7404	0.7388	0.7361	0.7281	0.7197	0.7073	0.6915	0.6805
	bdsbr	0.7208	0.7196	0.7167	0.7147	0.7104	0.7021	0.6919	0.6773	0.6590	0.6428	
	acov	0.6277	0.6256	0.6239	0.6227	0.6195	0.6150	0.6088	0.6006	0.5889	0.5745	
	str	0.7090	0.7072	0.7047	0.7023	0.6978	0.6877	0.6793	0.6598	0.6388	0.6242	
	Ssbr	0.5981	0.5955	0.5935	0.5919	0.5870	0.5836	0.5748	0.5714	0.5579	0.5398	
	Randomforest	0.6568	0.6525	0.6473	0.6435	0.6394	0.6292	0.6210	0.6108	0.5931	0.5792	
	XGBoost	0.7005	0.6915	0.6830	0.6796	0.6712	0.6610	0.6490	0.6397	0.6172	0.5960	
	LightGBM	0.6793	0.6733	0.6663	0.6631	0.6566	0.6477	0.6370	0.6277	0.6097	0.5906	

**Table D.4:** Energy Load Zones Forecast Results of MASE index for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	1.0633	1.0646	1.0659	1.0680	1.0697	1.0727	1.0733	1.0772	1.0828	1.0835
Naive	Randomforest	0.9246	0.9229	0.9209	0.9204	0.9183	0.9173	0.9112	0.9135	0.9001	0.8891
	XGBoost	0.9775	0.9713	0.9674	0.9663	0.9626	0.9584	0.9535	0.9518	0.9361	0.9224
	LightGBM	0.9674	0.9638	0.9603	0.9610	0.9575	0.9556	0.9496	0.9496	0.9357	0.9201

**Table D.5:** Energy Load Areas Forecast Results for Naive Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	1.1158	1.1716	1.1436	1.0772	1.0615	0.6606	0.6885	0.5815	0.5390	0.5016	
	Bottom Up	1.1055	1.1065	1.1056	1.1040	1.1001	1.0987	1.0836	1.0812	1.0553	1.0277	
tcs	<i>wlsv-shr</i>	0.8590	0.8573	0.8533	0.8491	0.8392	0.8247	0.8056	0.7754	0.7392	0.6941	
	<i>sari-shr</i>	0.7426	0.7391	0.7338	0.7260	0.7147	0.6791	0.6665	0.5968	0.5272	0.4529	
	<i>acov-shr</i>	0.5604	0.5577	0.5549	0.5517	0.5456	0.5403	0.5282	0.5119	0.4882	0.4352	
	<i>acov-wls</i>	0.5562	0.5536	0.5507	0.5478	0.5424	0.5366	0.5250	0.5095	0.4864	0.4363	
	<i>sari-wls</i>	0.7455	0.7419	0.7367	0.7290	0.7179	0.6820	0.6706	0.5997	0.5292	0.4541	
cst	<i>wlsv-shr</i>	0.8915	0.8898	0.8867	0.8820	0.8729	0.8594	0.8401	0.8085	0.7726	0.7304	
	<i>sari-shr</i>	0.7584	0.7550	0.7499	0.7425	0.7311	0.6986	0.6885	0.6182	0.5535	0.4808	
	<i>acov-shr</i>	0.6529	0.6491	0.6451	0.6391	0.6277	0.6162	0.6008	0.5676	0.5382	0.4755	
	<i>acov-wls</i>	0.6493	0.6462	0.6430	0.6378	0.6269	0.6180	0.6027	0.5764	0.5479	0.4888	
	<i>sari-wls</i>	0.7612	0.7581	0.7531	0.7458	0.7350	0.7006	0.6920	0.6222	0.5592	0.4859	
ETS	ite	<i>wlsv-shr</i>	0.8529	0.8509	0.8470	0.8424	0.8328	0.8151	0.7990	0.7614	0.7205	0.6791
		<i>sari-shr</i>	0.7418	0.7382	0.7329	0.7246	0.7131	0.6790	0.6670	0.5952	0.5246	0.4513
		<i>acov-wls</i>	0.5576	0.5549	0.5520	0.5491	0.5436	0.5374	0.5257	0.5100	0.4871	0.4372
		<i>sari-wls</i>	0.7455	0.7419	0.7367	0.7290	0.7179	0.6819	0.6706	0.5996	0.5291	0.4540
	oct	<i>wlsv</i>	0.8607	0.8591	0.8550	0.8509	0.8413	0.8269	0.8082	0.7780	0.7424	0.6972
		<i>bdsbr</i>	0.8030	0.8000	0.7948	0.7885	0.7772	0.7563	0.7366	0.6939	0.6576	0.5961
		<i>acov</i>	0.5586	0.5560	0.5533	0.5506	0.5448	0.5407	0.5290	0.5164	0.4938	0.4440
		<i>str</i>	0.8337	0.8308	0.8265	0.8208	0.8094	0.7878	0.7694	0.7212	0.6798	0.6207
		<i>Sshbr</i>	0.5485	0.5458	0.5425	0.5397	0.5338	0.5297	0.5167	0.5077	0.4908	0.4389
	Randomforest	0.6337	0.6267	0.6172	0.6094	0.6000	0.5848	0.5623	0.5396	0.5225	0.4977	
	XGBoost	0.6776	0.6670	0.6559	0.6473	0.6377	0.6206	0.5989	0.5817	0.5547	0.5243	
	LightGBM	0.6439	0.6350	0.6246	0.6171	0.6065	0.5922	0.5733	0.5533	0.5336	0.5028	

**Table D.6:** Energy Load Areas Forecast Results of MASE index for ETS Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	1.0977	1.0970	1.1180	1.1210	1.0992	0.5994	0.5051	0.5171	0.4863	0.5510	
	Bottom Up	1.0905	1.0918	1.0909	1.0896	1.0859	1.0814	1.0673	1.0620	1.0358	1.0110	
tcs	<i>wlsv-shr</i>	0.8239	0.8220	0.8171	0.8113	0.8027	0.7860	0.7651	0.7370	0.7045	0.6818	
	<i>sari-shr</i>	0.7003	0.6968	0.6906	0.6823	0.6705	0.6364	0.6227	0.5611	0.5044	0.4685	
	<i>acov-shr</i>	0.5632	0.5606	0.5573	0.5551	0.5493	0.5421	0.5319	0.5155	0.4951	0.4661	
	<i>acov-wls</i>	0.5617	0.5590	0.5560	0.5540	0.5488	0.5420	0.5311	0.5149	0.4955	0.4697	
	<i>sari-wls</i>	0.6997	0.6963	0.6903	0.6822	0.6710	0.6361	0.6230	0.5624	0.5058	0.4721	
cst	<i>wlsv-shr</i>	0.8477	0.8452	0.8406	0.8350	0.8248	0.8096	0.7888	0.7614	0.7306	0.7074	
	<i>sari-shr</i>	0.7080	0.7042	0.6982	0.6900	0.6782	0.6465	0.6315	0.5697	0.5185	0.4885	
	<i>acov-shr</i>	0.6271	0.6228	0.6178	0.6122	0.6012	0.5904	0.5718	0.5473	0.5192	0.4885	
	<i>acov-wls</i>	0.6355	0.6323	0.6280	0.6227	0.6124	0.6022	0.5843	0.5616	0.5346	0.5042	
	<i>sari-wls</i>	0.7149	0.7114	0.7058	0.6978	0.6874	0.6551	0.6408	0.5809	0.5324	0.5014	
SARIMA												
	ite	<i>wlsv-shr</i>	0.8125	0.8097	0.8040	0.7982	0.7877	0.7702	0.7483	0.7188	0.6842	0.6603
		<i>sari-shr</i>	0.6939	0.6899	0.6832	0.6748	0.6623	0.6288	0.6140	0.5519	0.4915	0.4575
		<i>acov-wls</i>	0.5632	0.5606	0.5576	0.5557	0.5505	0.5429	0.5321	0.5156	0.4955	0.4691
		<i>sari-wls</i>	0.6996	0.6961	0.6901	0.6821	0.6708	0.6359	0.6228	0.5620	0.5054	0.4715
oct	<i>wlsv</i>	0.8308	0.8290	0.8246	0.8190	0.8114	0.7945	0.7742	0.7465	0.7155	0.6945	
	<i>bdskr</i>	0.7717	0.7679	0.7617	0.7545	0.7431	0.7216	0.7012	0.6633	0.6310	0.5996	
	<i>acov</i>	0.5663	0.5636	0.5607	0.5584	0.5535	0.5462	0.5377	0.5229	0.5014	0.4745	
	<i>str</i>	0.8037	0.8008	0.7963	0.7886	0.7778	0.7575	0.7355	0.6907	0.6582	0.6336	
	<i>Sshr</i>	0.5333	0.5300	0.5261	0.5234	0.5179	0.5121	0.5021	0.4942	0.4762	0.4520	
Randomforest												
		0.6410	0.6343	0.6254	0.6166	0.6058	0.5902	0.5653	0.5457	0.5335	0.5107	
	XGBoost	0.6818	0.6721	0.6609	0.6543	0.6419	0.6250	0.5977	0.5809	0.5624	0.5408	
LightGBM												
		0.6458	0.6377	0.6268	0.6207	0.6087	0.5952	0.5716	0.5523	0.5367	0.5118	

**Table D.7:** Energy Load Areas Forecast Results of MASE index for SARIMA Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	0.9108	0.9261	0.9113	0.8945	0.8843	0.6293	0.6171	0.5999	0.5771	0.6214
	Bottom Up	0.9061	0.9068	0.9060	0.9054	0.9026	0.9025	0.8893	0.8808	0.8713	0.8652
tcs	wlsv-shr	0.7877	0.7868	0.7846	0.7827	0.7775	0.7710	0.7579	0.7334	0.7191	0.7088
	sari-shr	0.7042	0.7021	0.6990	0.6962	0.6890	0.6713	0.6649	0.6266	0.5950	0.5761
	acov-shr	0.6259	0.6240	0.6220	0.6212	0.6161	0.6123	0.6040	0.5923	0.5792	0.5664
	acov-wls	0.6236	0.6219	0.6199	0.6189	0.6141	0.6106	0.6020	0.5888	0.5759	0.5636
	sari-wls	0.7054	0.7035	0.7004	0.6977	0.6905	0.6732	0.6668	0.6291	0.5979	0.5787
cst	wlsv-shr	0.8020	0.8010	0.7988	0.7967	0.7920	0.7862	0.7731	0.7501	0.7360	0.7262
	sari-shr	0.7111	0.7092	0.7063	0.7035	0.6962	0.6807	0.6737	0.6373	0.6089	0.5941
	acov-shr	0.6652	0.6629	0.6604	0.6583	0.6511	0.6455	0.6342	0.6143	0.5992	0.5854
	acov-wls	0.6660	0.6640	0.6620	0.6600	0.6533	0.6480	0.6369	0.6191	0.6041	0.5889
	sari-wls	0.7159	0.7142	0.7115	0.7088	0.7015	0.6862	0.6792	0.6440	0.6163	0.6005
Forecast Combination	ite	0.7847	0.7834	0.7810	0.7790	0.7734	0.7666	0.7532	0.7277	0.7124	0.7013
	wlsv-shr	0.7005	0.6983	0.6950	0.6922	0.6849	0.6671	0.6610	0.6219	0.5894	0.5717
	sari-shr	0.6232	0.6214	0.6194	0.6184	0.6136	0.6099	0.6012	0.5878	0.5749	0.5623
	acov-wls	0.7054	0.7035	0.7003	0.6977	0.6905	0.6732	0.6667	0.6290	0.5978	0.5786
	oct	0.7874	0.7865	0.7844	0.7826	0.7774	0.7712	0.7580	0.7335	0.7194	0.7096
	wlsv	0.7602	0.7585	0.7554	0.7525	0.7467	0.7385	0.7239	0.6974	0.6802	0.6650
	bdsbr	0.6215	0.6198	0.6178	0.6169	0.6127	0.6096	0.6015	0.5887	0.5756	0.5623
	acov	0.7451	0.7434	0.7404	0.7379	0.7316	0.7203	0.7097	0.6753	0.6567	0.6423
	str	0.5683	0.5660	0.5637	0.5628	0.5575	0.5557	0.5451	0.5410	0.5288	0.5113
	Ssbr	0.6469	0.6417	0.6356	0.6303	0.6238	0.6139	0.6016	0.5826	0.5681	0.5522
	Randomforest	0.6755	0.6659	0.6581	0.6532	0.6446	0.6338	0.6176	0.6013	0.5820	0.5622
	XGBoost	0.6580	0.6512	0.6441	0.6393	0.6327	0.6207	0.6093	0.5921	0.5762	0.5576

**Table D.8:** Energy Load Areas Forecast Results of MASE index for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	1.1031	1.1037	1.1042	1.1054	1.1064	1.1070	1.1078	1.1074	1.1123	1.1089
Naive	Randomforest	1.0132	1.0112	1.0093	1.0075	1.0061	1.0048	1.0003	1.0004	0.9877	0.9663
	XGBoost	1.0477	1.0422	1.0390	1.0385	1.0340	1.0309	1.0312	1.0252	1.0107	0.9909
	LightGBM	1.0549	1.0514	1.0486	1.0479	1.0447	1.0425	1.0373	1.0374	1.0241	0.9969

**Table D.9:** Energy Load Italy Forecast Results of MASE index for Naive Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Naive are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
Base	Base	1.4490	1.6157	1.5593	1.3539	1.3231	0.6647	0.6059	0.6107	0.5390	0.5327	
Bottom Up	Bottom Up	1.4158	1.4188	1.4194	1.4210	1.4195	1.4240	1.4182	1.4174	1.3942	1.3899	
tcs	<i>wlsv-shr</i>	0.9918	0.9914	0.9880	0.9871	0.9787	0.9619	0.9483	0.9085	0.8788	0.8556	
	<i>sari-shr</i>	0.7976	0.7953	0.7914	0.7864	0.7789	0.7282	0.7348	0.6255	0.5430	0.4800	
	<i>acov-shr</i>	0.5724	0.5709	0.5683	0.5669	0.5618	0.5561	0.5518	0.5339	0.5126	0.4775	
	<i>acov-wls</i>	0.5640	0.5624	0.5602	0.5595	0.5548	0.5493	0.5458	0.5288	0.5093	0.4762	
	<i>sari-wls</i>	0.8018	0.7997	0.7956	0.7907	0.7834	0.7320	0.7400	0.6301	0.5444	0.4810	
cst	<i>wlsv-shr</i>	1.0733	1.0737	1.0728	1.0713	1.0635	1.0523	1.0395	1.0000	0.9681	0.9519	
	<i>sari-shr</i>	0.8293	0.8274	0.8231	0.8188	0.8105	0.7718	0.7763	0.6772	0.5899	0.5400	
	<i>acov-shr</i>	0.6767	0.6748	0.6711	0.6678	0.6621	0.6482	0.6401	0.6054	0.5780	0.5413	
	<i>acov-wls</i>	0.6717	0.6707	0.6673	0.6659	0.6606	0.6508	0.6442	0.6208	0.5982	0.5637	
	<i>sari-wls</i>	0.8346	0.8328	0.8287	0.8258	0.8177	0.7748	0.7824	0.6846	0.6012	0.5508	
ETS	ite	<i>wlsv-shr</i>	0.9798	0.9797	0.9771	0.9749	0.9660	0.9469	0.9338	0.8841	0.8461	0.8256
		<i>sari-shr</i>	0.7956	0.7929	0.7887	0.7835	0.7757	0.7294	0.7349	0.6245	0.5380	0.4761
		<i>acov-wls</i>	0.5651	0.5635	0.5612	0.5604	0.5557	0.5501	0.5464	0.5299	0.5110	0.4789
		<i>sari-wls</i>	0.8018	0.7998	0.7956	0.7908	0.7834	0.7320	0.7400	0.6300	0.5445	0.4809
	oct	<i>wlsv</i>	0.9917	0.9914	0.9877	0.9869	0.9792	0.9617	0.9489	0.9088	0.8789	0.8548
		<i>bdsbr</i>	0.9198	0.9190	0.9143	0.9111	0.9030	0.8780	0.8653	0.8121	0.7681	0.7420
		<i>acov</i>	0.5756	0.5740	0.5720	0.5711	0.5668	0.5621	0.5589	0.5435	0.5241	0.4950
		<i>str</i>	0.9737	0.9727	0.9696	0.9674	0.9586	0.9354	0.9258	0.8620	0.8224	0.7897
		<i>Sshbr</i>	0.5589	0.5575	0.5559	0.5550	0.5521	0.5482	0.5415	0.5296	0.5164	0.4811
Randomforest		0.6293	0.6228	0.6117	0.6044	0.5947	0.5750	0.5593	0.5305	0.5193	0.4884	
XGBoost		0.6468	0.6391	0.6281	0.6215	0.6137	0.5922	0.5770	0.5543	0.5381	0.5061	
LightGBM		0.6245	0.6168	0.6063	0.5999	0.5915	0.5734	0.5531	0.5378	0.5161	0.4836	

**Table D.10:** Energy Load Italy Forecast Results of MASE index for ETS Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast ETS are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency										
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h	
	Base	1.3726	1.3529	1.3572	1.3542	1.3279	0.6077	0.5128	0.5675	0.5111	0.5592	
	Bottom Up	1.3616	1.3642	1.3643	1.3655	1.3629	1.3623	1.3545	1.3460	1.3255	1.3194	
tcs	<i>wlsv-shr</i>	0.9444	0.9439	0.9399	0.9371	0.9303	0.9138	0.8961	0.8579	0.8335	0.8226	
	<i>sari-shr</i>	0.7657	0.7632	0.7588	0.7536	0.7461	0.6970	0.7013	0.6074	0.5357	0.5037	
	<i>acov-shr</i>	0.5972	0.5956	0.5931	0.5926	0.5900	0.5801	0.5783	0.5533	0.5367	0.5136	
	<i>acov-wls</i>	0.5885	0.5873	0.5852	0.5846	0.5822	0.5740	0.5709	0.5470	0.5334	0.5148	
	<i>sari-wls</i>	0.7597	0.7572	0.7534	0.7482	0.7416	0.6912	0.6986	0.6020	0.5328	0.5034	
cst	<i>wlsv-shr</i>	1.0166	1.0162	1.0136	1.0120	1.0034	0.9891	0.9738	0.9370	0.9139	0.9070	
	<i>sari-shr</i>	0.7854	0.7831	0.7789	0.7747	0.7684	0.7268	0.7292	0.6357	0.5736	0.5509	
	<i>acov-shr</i>	0.6753	0.6723	0.6683	0.6659	0.6606	0.6482	0.6358	0.6007	0.5788	0.5543	
	<i>acov-wls</i>	0.6808	0.6795	0.6773	0.6754	0.6711	0.6594	0.6476	0.6181	0.5952	0.5740	
	<i>sari-wls</i>	0.7917	0.7894	0.7861	0.7821	0.7765	0.7339	0.7387	0.6462	0.5903	0.5670	
SARIMA												
	ite	<i>wlsv-shr</i>	0.9328	0.9315	0.9269	0.9236	0.9149	0.8960	0.8769	0.8371	0.8099	0.7991
		<i>sari-shr</i>	0.7567	0.7536	0.7489	0.7436	0.7352	0.6864	0.6914	0.5962	0.5185	0.4887
		<i>acov-wls</i>	0.5905	0.5894	0.5874	0.5866	0.5840	0.5758	0.5714	0.5479	0.5349	0.5151
		<i>sari-wls</i>	0.7596	0.7572	0.7533	0.7482	0.7416	0.6911	0.6985	0.6018	0.5326	0.5032
oct	<i>wlsv</i>	0.9460	0.9455	0.9418	0.9390	0.9336	0.9157	0.8977	0.8611	0.8384	0.8276	
	<i>bdsbr</i>	0.8812	0.8792	0.8741	0.8691	0.8605	0.8380	0.8207	0.7737	0.7457	0.7356	
	<i>acov</i>	0.6023	0.6012	0.5995	0.5989	0.5972	0.5895	0.5867	0.5648	0.5513	0.5332	
	<i>str</i>	0.9357	0.9348	0.9322	0.9287	0.9227	0.9000	0.8853	0.8196	0.7914	0.7812	
	<i>Sshsr</i>	0.5429	0.5416	0.5389	0.5374	0.5352	0.5293	0.5259	0.5154	0.4998	0.4854	
Randomforest												
		0.6425	0.6365	0.6271	0.6170	0.6078	0.5861	0.5679	0.5396	0.5329	0.5053	
	XGBoost	0.6587	0.6497	0.6407	0.6338	0.6240	0.6033	0.5813	0.5605	0.5453	0.5248	
LightGBM												
		0.6292	0.6225	0.6121	0.6045	0.5955	0.5763	0.5569	0.5394	0.5242	0.4975	

**Table D.11: Energy Load Italy Forecast Results of MASE index for SARIMA Base Method.**

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast SARIMA are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)

Base Method	Forecast Combination	Temporal Frequency									
		30min	1h	1.5h	2h	3h	4h	6h	8h	12h	24h
	Base	1.1165	1.1567	1.1242	1.0554	1.0425	0.6442	0.6124	0.6254	0.5846	0.6439
	Bottom Up	1.1060	1.1082	1.1087	1.1102	1.1090	1.1120	1.1079	1.0967	1.0938	1.1038
Forecast Combination	tcs <i>wlsv-shr</i>	0.9003	0.9008	0.8984	0.8978	0.8945	0.8878	0.8822	0.8524	0.8442	0.8430
	<i>sari-shr</i>	0.7603	0.7595	0.7564	0.7538	0.7502	0.7235	0.7279	0.6651	0.6268	0.6133
	<i>acov-shr</i>	0.6670	0.6663	0.6646	0.6640	0.6626	0.6563	0.6543	0.6352	0.6231	0.6152
	<i>acov-wls</i>	0.6621	0.6616	0.6600	0.6595	0.6582	0.6515	0.6494	0.6289	0.6169	0.6090
	<i>sari-wls</i>	0.7576	0.7567	0.7537	0.7513	0.7476	0.7211	0.7257	0.6639	0.6251	0.6115
	cst <i>wlsv-shr</i>	0.9402	0.9410	0.9398	0.9398	0.9369	0.9328	0.9279	0.8989	0.8929	0.8946
	<i>sari-shr</i>	0.7759	0.7749	0.7722	0.7705	0.7665	0.7459	0.7476	0.6934	0.6609	0.6528
	<i>acov-shr</i>	0.7148	0.7138	0.7113	0.7103	0.7076	0.6996	0.6943	0.6673	0.6547	0.6463
	<i>acov-wls</i>	0.7097	0.7092	0.7073	0.7064	0.7038	0.6966	0.6916	0.6659	0.6542	0.6466
	<i>sari-wls</i>	0.7779	0.7769	0.7748	0.7730	0.7693	0.7469	0.7497	0.6974	0.6654	0.6565
ite	<i>wlsv-shr</i>	0.8983	0.8985	0.8959	0.8952	0.8923	0.8847	0.8791	0.8469	0.8389	0.8374
	<i>sari-shr</i>	0.7553	0.7543	0.7511	0.7487	0.7446	0.7191	0.7235	0.6603	0.6200	0.6069
	<i>acov-wls</i>	0.6609	0.6603	0.6587	0.6581	0.6569	0.6501	0.6477	0.6273	0.6153	0.6072
	<i>sari-wls</i>	0.7576	0.7567	0.7538	0.7514	0.7476	0.7211	0.7257	0.6638	0.6251	0.6115
oct	<i>wlsv</i>	0.8955	0.8960	0.8937	0.8930	0.8901	0.8830	0.8778	0.8480	0.8393	0.8382
	<i>bdsbr</i>	0.8657	0.8651	0.8621	0.8606	0.8571	0.8475	0.8402	0.8066	0.7967	0.7907
	<i>acov</i>	0.6694	0.6692	0.6677	0.6673	0.6660	0.6612	0.6588	0.6400	0.6282	0.6232
	<i>str</i>	0.8334	0.8326	0.8302	0.8297	0.8260	0.8109	0.8087	0.7647	0.7468	0.7426
	<i>Sshbr</i>	0.5919	0.5913	0.5897	0.5890	0.5881	0.5847	0.5796	0.5747	0.5668	0.5571
Randomforest		0.6728	0.6680	0.6615	0.6569	0.6514	0.6414	0.6317	0.6118	0.5965	0.5747
XGBoost		0.6723	0.6651	0.6581	0.6540	0.6461	0.6360	0.6216	0.6043	0.5890	0.5656
LightGBM		0.6704	0.6656	0.6588	0.6548	0.6495	0.6379	0.6272	0.6121	0.5962	0.5772

**Table D.12:** Energy Load Italy Forecast Results of MASE index for Forecast Combination Base Method.

(Note: This table shows forecast accuracy measured in MASE. The best forecast reconciliation results for the base forecast Forecast Combination are highlighted in gray. Section 4.3 explains the ML benchmarks, while Section 4.4 details the linear benchmarks.)