FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Hierarchical Framework for Electric Load Forecasting with Self-Consumption Impact Analysis: A Case Study for Portugal

Rita Portela Marques



Master's Degree in Electrical and Computer Engineering

Supervisor: Sergio Asanza, PhD

October 13, 2026

© Rita Portela Marques, 2026

Resumo

A previsão de carga elétrica requer, cada vez mais, uma desagregação espacial e uma coerência interna entre os diferentes níveis administrativos, um requisito que se torna mais premente à medida que o autoconsumo cresce de forma heterogénea em todo o território, com consequências para o planeamento da rede de distribuição. Esta dissertação desenvolve um framework hierárquico de previsão desta carga para Portugal continental, estruturado em torno dos níveis nacional, distrital, municipal e de freguesia, utilizando dados de consumo publicados pela E-REDES. O modelo AutoARIMA é ajustado a todas as séries da hierarquia, enquanto as previsões Naive e Seasonal Naive são calculadas para cada série. As previsões resultantes são reconciliadas através dos métodos bottom-up, top-down e middle-out.

Ao aplicar o framework separadamente ao consumo em baixa tensão, em média/alta tensão, e nos dois níveis combinados, verifica-se que não há um método de reconciliação consistentemente superior: o mais adequado depende da hierarquia e do nível de tensão analisado. Os dois segmentos, baixa e média/alta tensão, respondem também de forma diferente perante uma anomalia de procura registada apenas num deles, o que sugere que são influenciados, pelo menos em parte, por fatores distintos.

Uma análise exploratória do impacto do autoconsumo revela que a capacidade instalada cresceu rapidamente e está agora presente em todos os municípios de Portugal continental, enquanto o consumo subjacente continuou a crescer em vez de diminuir. A penetração do autoconsumo, em relação ao consumo local, é mais baixa nas maiores áreas metropolitanas e mais elevada nos municípios mais pequenos do interior e do sul, um padrão apenas visível a um nível de granularidade espacial elevado. Apesar desta variação espacial, a precisão das previsões ao nível das freguesias não revelou uma associação sistemática relevante com a densidade local de autoconsumo, sugerindo que a penetração fotovoltaica, por si só, não foi um determinante importante da previsibilidade nos dados analisados. Estes resultados demonstram o valor prático da previsão hierárquica, desagregada ao nível da freguesia, para o planeamento da rede de distribuição num contexto de autoconsumo em crescimento.

Palavras-chave: Previsão Espacial de Carga, Previsão Hieráquica, Reconciliação de Previsões, Autoconsumo, Planeamento de Redes de Distribuição

Abstract

Electric load forecasting increasingly needs to be both spatially disaggregated and internally coherent across administrative levels, a requirement that becomes more pressing as distributed self-consumption grows unevenly across a territory, with consequences for distribution network planning. This dissertation develops a hierarchical electric load forecasting framework for mainland Portugal, structured around its national, district, municipality, and parish levels, using open consumption data published by E-REDES. `AutoARIMA` is fitted to every series in the hierarchy, while `Naive` and **Seasonal Naive** forecasts are computed for each series. The resulting forecasts are reconciled using bottom-up, top-down, and middle-out approaches.

Applying the framework separately to low voltage, medium, high, and very high voltage, and all-voltage consumption reveals that no single reconciliation method performs best across the board: the optimal choice depends on the specific hierarchy and voltage scope being forecast. The low voltage and medium/high voltage segments also respond differently to a demand anomaly observed in one but not the other, suggesting they are driven, at least in part, by different underlying factors.

An exploratory analysis of self-consumption impact shows that installed capacity has grown rapidly and is now present in every municipality in mainland Portugal, while underlying consumption continued to grow rather than decline. Self-consumption penetration, relative to local consumption, is lowest in the largest metropolitan areas and highest in smaller interior and southern municipalities, a pattern only visible at fine spatial granularity. Despite this spatial variation, parish-level forecast accuracy showed no substantive systematic association with local self-consumption density, suggesting that photovoltaic penetration alone was not a major driver of forecastability in the analyzed data. These results demonstrate the practical value of hierarchical forecasting, disaggregated to the parish level, for distribution network planning under growing self-consumption.

Keywords: Spatial Load Forecasting, Hierarchical Forecasting, Forecast Reconciliation, Self-Consumption, Distribution Network Planning

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework to achieve a better and more sustainable future for all. They comprise 17 goals[1] that address global challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.



Figure 1: United Nations Sustainable Development Goals (SDG).

Table 1 highlights the specific goals and targets to which this dissertation contributes, detailing the nature of the contribution and the performance indicators used to assess its impact.

[1] https://sdgs.un.org/goals

Table 1: Sustainable Development Goals Addressed by the Research.

SDG	Target	Contribution	Performance Indicators and Metrics
7	7.2	Spatial analysis of self-consumption growth supports the integration of distributed solar photovoltaic generation in mainland Portugal, contributing to an increase in the share of renewable sources in electricity consumption.	- Increase in self-consumption capacity density (kW/GWh) - Number of municipalities with self-consumption installations
	7.b	Hierarchical load forecasting at fine spatial granularity supports the expansion and modernization of energy infrastructure by enabling data-driven distribution network planning, addressing the infrastructure needs highlighted by Target 7.b.	- Forecast accuracy at parish level (MASE, RMSE) - Number of parishes, municipalities, and districts covered by the framework
9	9.1	Coherent forecasts across national, district, municipality, and parish levels support the development of reliable and sustainable distribution network infrastructure, contributing to economic development and human well-being.	- Reduction in forecast error relative to the naive benchmark (MASE) - Structural forecast coherence guaranteed by the summing matrix formulation

"Cença."

Sérgio Fernandes

Contents

1 Introduction **1**
- 1.1 Motivation . . . 2
- 1.2 Goals . . . 2
- 1.3 Dissertation Structure . . . 2

2 Literature Review and State-of-the-Art **4**
- 2.1 Spatial Load Forecasting . . . 4
  - 2.1.1 Classification of SLF Methods . . . 5
- 2.2 Hierarchical Forecasting Approaches . . . 9
- 2.3 Hybrid Approaches: Simulation and Hierarchical Methods . . . 10
- 2.4 Probabilistic Approaches, Uncertainty Quantification and Model Validation . . . 11
- 2.5 Self-Consumption and Distributed Energy Resource Adoption . . . 12
- 2.6 Research Gaps and Positioning of Current Work . . . 13
- 2.7 Summary . . . 14

3 Methodology **16**
- 3.1 Base Forecasting Models . . . 16
  - 3.1.1 ARIMA and Seasonal ARIMA . . . 17
  - 3.1.2 AutoARIMA . . . 18
  - 3.1.3 Naive . . . 18
- 3.2 Time Series Decomposition . . . 19
- 3.3 Forecast Reconciliation . . . 19
  - 3.3.1 Hierarchical Structure of the Model . . . 19
  - 3.3.2 Single-level Reconciliation Approaches . . . 20
- 3.4 Evaluation Metrics . . . 23
- 3.5 Correlation Analysis . . . 24
- 3.6 Summary . . . 25

4 Case Study, Results and Discussion **27**
- 4.1 Case Study and Data Sources . . . 27
  - 4.1.1 Study Area . . . 27
  - 4.1.2 Data Sources and Preprocessing . . . 27
- 4.2 Exploratory Data Analysis . . . 31
  - 4.2.1 Time Series Decomposition . . . 31
  - 4.2.2 Self-Consumption Growth Analysis . . . 35
- 4.3 Implementation . . . 38
  - 4.3.1 Pipeline . . . 38
  - 4.3.2 Model Configuration . . . 39

4.4 Low Voltage Results . . . 40
4.4.1 Accuracy Metrics . . . 40
4.4.2 National Forecast Plots . . . 41
4.5 Medium, High, and Very High Voltage Results . . . 43
4.5.1 Accuracy Metrics . . . 43
4.5.2 National Forecast Plots . . . 44
4.6 All Voltage Results . . . 46
4.6.1 Accuracy Metrics . . . 46
4.6.2 National Forecast Plots . . . 47
4.7 Discussion . . . 49
4.7.1 Cross-Voltage Comparison . . . 49
4.7.2 The January 2026 Anomaly and Forecastability . . . 50
4.7.3 Parish-Level Error Patterns and Hierarchical Propagation . . . 51
4.7.4 Implications for Self-Consumption and Grid Planning . . . 55
4.7.5 Forecasting Beyond the Evaluation Period . . . 55

5 Conclusion and Future Work **58**
5.1 Conclusion . . . 58
5.2 Limitations . . . 60
5.2.1 Data Limitations . . . 60
5.2.2 Methodological Limitations . . . 60
5.2.3 Scope Limitations . . . 61
5.3 Future Work . . . 61

References **63**

A District-Level Self-Consumption Figures **67**

B Pipeline Pseudocode **68**

C Spatial Error Maps **71**

D Projected Changes in Parish-Level Consumption **75**

List of Figures

1 United Nations Sustainable Development Goals (SDG). iii

2.1 Geographic distribution of predicted demand in spatial load forecasting. 5
2.2 The three phases of the S-curve model of load growth. 8
2.3 Hybrid framework linking hierarchical forecasting to grid-cell disaggregation. . . 11

3.1 Illustrative example of a four-level hierarchical structure and its summing matrix S. 20
3.2 The three single-level reconciliation approaches. 21

4.1 Administrative hierarchy of mainland Portugal: districts (left), municipalities (center), and parishes (right). 28
4.2 MSTL decomposition of the national low-voltage consumption series (November 2020 - April 2025). 32
4.3 MSTL decomposition of the national medium, high, and very high voltage consumption series (November 2020 - April 2025). 33
4.4 MSTL decomposition of the national all-voltage consumption series (November 2020 - April 2025). 34
4.5 Evolution of installed low-voltage UPAC solar PV capacity for the five districts with the highest total installed capacity (January 2023 - June 2026). 35
4.6 National low-voltage monthly consumption and low-voltage installed PV UPAC capacity, January 2023 - April 2026. 37
4.7 Overview of the forecasting pipeline implemented. 38
4.8 National low-voltage total: reconciled `AutoARIMA` forecasts versus observed values, May 2025 - April 2026. 42
4.9 National low-voltage total: Naive and Seasonal Naive forecasts versus observed values, May 2025 - April 2026. 43
4.10 National medium, high, and very high voltage total: reconciled `AutoARIMA` forecasts versus observed values, May 2025 - April 2026. 45
4.11 National medium, high, and very high voltage total: Naive and Seasonal Naive forecasts versus observed values, May 2025 - April 2026. 46
4.12 National all-voltage total: reconciled `AutoARIMA` forecasts versus observed values, May 2025 - April 2026. 48
4.13 National all-voltage total: Naive and Seasonal Naive forecasts versus observed values, May 2025 - April 2026. 49
4.14 Parish-level MASE under AutoARIMA and Bottom-Up reconciliation for low voltage (left), medium, high, very-high voltage (center), and all voltage levels (right). 52
4.15 Parish-level MASE versus UPAC installed capacity density for low voltage (left), medium, high, very-high voltage (center), and all voltage levels (right). 53

4.16 Historical national consumption and forecasts for low voltage, medium, high, and very high voltage, and all voltage levels, from top to bottom. 56
4.17 Projected percentage change in parish-level consumption for low voltage, medium, high, and very high voltage, and all voltage levels, from left to right. 57

C.1 Parish-level MASE obtained with AutoARIMA and Bottom-Up reconciliation, low voltage. 72
C.2 Parish-level MASE obtained with AutoARIMA and Bottom-Up reconciliation, medium, high, and very high voltage. 73
C.3 Parish-level MASE obtained with AutoARIMA and Bottom-Up reconciliation, all voltage. 74

D.1 Projected percentage change in parish-level consumption, low voltage. 76
D.2 Projected percentage change in parish-level consumption, medium, high, and very high voltage. 77
D.3 Projected percentage change in parish-level consumption, all voltage. 78

List of Tables

1 Sustainable Development Goals Addressed by the Research. iv

2.1 Overview of simulation-based SLF methodologies. 7
2.2 Overview of trend-based SLF methodologies. 9
2.3 Research gaps identified in the SLF literature. 14

3.1 Conventional interpretation of correlation coefficient magnitude, following [32]. . 25

4.1 Parishes absent from the low-voltage analysis and corresponding cooperative operators. 30
4.2 Ten low-voltage consumers with the highest monthly consumption in April 2026 and their UPAC capacity. 36
4.3 Forecast accuracy for the low-voltage analysis by hierarchical level. 40
4.4 Forecast accuracy for the medium, high and very high voltage analysis by hierarchical level. 44
4.5 Forecast accuracy for the all-voltage analysis by hierarchical level. 47
4.6 AutoARIMA MASE with and without January 2026 in the test period, by hierarchical level. 51
4.7 Correlation between parish-level MASE and UPAC installed-capacity density, for low voltage (LV), medium, high, and very high voltage (MHV), and all voltage (AV). 53

A.1 Solar PV self-consumption capacity, consumption, and density by district (April 2026). 67

List of Acronyms

ANN	Artificial Neural Network
ARIMA	AutoRegressive Integrated Moving Average
CA	Cellular Automata
DER	Distributed Energy Resources
DICO	District-Municipality Code
DICOFRE	District-Municipality-Parish Code
DNO	Distribution Network Operator
EV	Electric Vehicle
FoM	Figure-of-Merit
GIS	Geographic Information System
GLS	Generalized Least Squares
HLF	Hierarchical Load Forecasting
HYL	Horizon Year Load
LTLF	Long-Term Load Forecasting
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
MCS	Monte Carlo Simulation
MSTL	Multiple Seasonal-Trend decomposition using Loess
p-RMSE	Probabilistic Root-Mean-Squared Error
P-SLF	Probabilistic Spatial Load Forecasting
PI	Prediction Interval
PICP	Prediction Interval Coverage Probability
PV	Photovoltaic
RMSE	Root-Mean-Squared Error
SARIMA	Seasonal AutoRegressive Integrated Moving Average
SLF	Spatial Load Forecasting
STLF	Short-Term Load Forecasting
UPAC	Self-Consumption Production Unit (*Unidade de Produção para Autoconsumo*)

Chapter 1

Introduction

The global energy sector is undergoing a profound transformation driven by decarbonization policies, the rapid adoption of distributed generation technologies, and the increasing electrification of end uses. Residential consumers are increasingly adopting distributed generation technologies such as rooftop solar, together with new electrical loads, including electric vehicles and heat pumps, thereby fundamentally changing their demand profiles [1], [2]. At the distribution network level, these changes translate into greater spatial heterogeneity in demand, with consumption patterns varying significantly across geographical areas depending on the local penetration of these technologies [3].

This heterogeneity poses a specific challenge for distribution network operators. Traditional load forecasting methods, applied at the level of an entire service territory or utility, can obscure substantial variation between individual areas. This variation is directly relevant to planning decisions such as substation and feeder sizing, since these are made at the local, not the national, level. Spatial load forecasting addresses this by producing forecasts disaggregated across the geography of the network. However, existing approaches typically face a trade-off between the coherence of hierarchical, trend-based methods and the fine spatial resolution of simulation-based methods, as discussed in Chapter 2.

The growth of distributed self-consumption adds a further layer of complexity. Distributed solar photovoltaic adoption is highly uneven across a territory, shaped by socio-economic factors such as income, age, and household composition, as well as peer effects [1]. As a result, its impact on net demand cannot be inferred from national or district-level trends alone. Understanding where self-consumption is growing, and how it relates to demand growth, is therefore relevant to distribution network planning in the same way that the spatial distribution of demand itself is.

To address the spatial heterogeneity of demand and self-consumption growth, this dissertation develops a hierarchical electric load forecasting framework at fine spatial granularity and pairs it with a spatially explicit analysis of self-consumption growth, using mainland Portugal as a case study. The work adopts the hierarchical side of the trade-off between coherence and fine spatial resolution, leaving integration with simulation-based methods for future work.

1.1 Motivation

This dissertation is motivated by several needs in distribution network planning. Distribution networks are planned and operated at multiple administrative levels simultaneously, from the national system down to individual parishes. Forecasts produced independently at each level can be mutually inconsistent, which a hierarchical approach that reconciles forecasts across levels can avoid. At the same time, the uneven geographic adoption of distributed generation, such as solar photovoltaics, means that national or regional consumption trends alone are insufficient to anticipate local network impacts, motivating forecasts at fine spatial granularity. The increasing availability of open, fine-grained consumption and self-consumption data, such as that published by E-REDES for mainland Portugal, creates an opportunity to apply and evaluate hierarchical forecasting methods at a spatial resolution not commonly reported in the literature.

1.2 Goals

The primary goal of this dissertation is to develop a hierarchical electric load forecasting framework for mainland Portugal, capable of producing forecasts that are coherent across the national, district, municipality, and parish levels of its administrative hierarchy. The framework compares bottom-up, top-down, and middle-out reconciliation, all of which guarantee this coherence by construction, to establish which approach minimizes forecast error at each level.

Beyond the forecasting framework itself, this dissertation pursues two further objectives. It analyses the spatial and temporal growth of distributed self-consumption in mainland Portugal, examining how it relates to the evolution of electricity demand and to spatial patterns not visible at coarser levels of aggregation. It also evaluates the sensitivity of the best-performing reconciliation method to the specific hierarchy and voltage scope being forecast, providing guidance on how such a choice should be made when applying hierarchical forecasting to a new case study.

Together, these goals establish a framework that supports distribution network planning at a spatial resolution not commonly reported in the literature, while providing a spatially explicit view of how distributed self-consumption is reshaping the demand that this planning must anticipate.

1.3 Dissertation Structure

In order to present this work in a structured and comprehensive manner, the document is divided into five chapters. Chapter 2, the Literature Review and State-of-the-Art, examines the existing work on spatial and hierarchical load forecasting and on self-consumption adoption, identifying the gaps that motivate and position this dissertation. Chapter 3, Methodology, presents the forecasting framework in detail, from the base forecasting models and the time series decomposition method to the reconciliation approaches used to combine forecasts across the hierarchy and the metrics used to evaluate their accuracy. Chapter 4, Case Study, Results and Discussion, applies this

framework to mainland Portugal as a case study, exploring the growth of self-consumption alongside the forecasting results obtained for the low voltage, medium, high, and very high voltage, and all voltage segments, and discussing what these results reveal when brought together. Chapter 5, Conclusion and Future Work, closes the dissertation by drawing together its main findings, reflecting on its limitations, and pointing to directions for future work.

Chapter 2

Literature Review and State-of-the-Art

This chapter presents a comprehensive literature review, providing the necessary theoretical background and contextual foundation for this research. Section 2.1 introduces spatial load forecasting, exploring its definitions, classifications, and relevance to distribution network planning. Section 2.2 provides an overview of hierarchical forecasting approaches. Section 2.3 examines hybrid approaches that combine simulation and hierarchical methods, differentiating the contributions of existing works to each component of a comprehensive forecasting framework. Section 2.4 addresses probabilistic approaches, uncertainty quantification, and model validation. Section 2.5 reviews the literature on self-consumption and distributed energy resource adoption. Finally, Section 2.6 identifies key research gaps in the current state of the art and positions the current work within the existing literature.

2.1 Spatial Load Forecasting

The global transition towards decarbonized and more distributed electricity systems is placing new demands on distribution network planning, as the growing adoption of distributed generation and new electrical loads reshapes where and how demand grows across a territory. Anticipating this growth at fine spatial resolution is a prerequisite for planning the infrastructure required by this transition, which is the specific problem addressed by spatial load forecasting.

As discussed in Chapter 1, accurate demand forecasting is essential for distribution network planning. Forecast errors may adversely affect the planning of substations and feeders, and the overall quality of demand forecasts has a direct influence on subsequent delivery system planning [4].

The need for robust forecasting tools arose from planning requirements that demanded accurate predictions of load evolution over extended time horizons. Planning is commonly divided into short-range and long-range horizons. Short-range planning extends only far enough into the future to account for equipment delivery and construction lead times, with the objective of determining the single best plan for present construction commitments. Long-range planning, by contrast, extends eight to twenty years ahead and aims to ensure that all elements of the short-range plan are

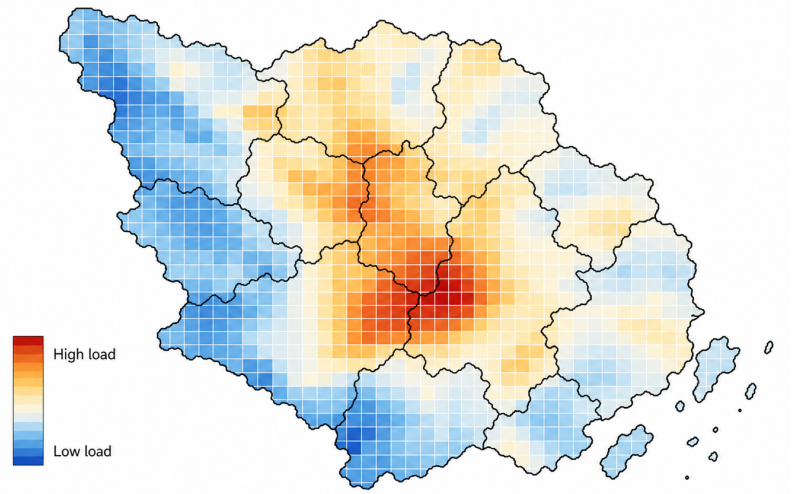


Figure 2.1: Geographic distribution of predicted demand in spatial load forecasting.

efficiently utilized [4]. In terms of temporal scope, load forecasting methodologies are commonly classified as short-term load forecasting (STLF) and long-term load forecasting (LTLF). STLF covers periods from one day to two weeks and supports operational decisions, while LTLF extends beyond three years and serves networks planning, expansion, and maintenance requirements [5].

In the late 1980s, the growing operational complexity of distribution systems led electrical engineers to develop methods for predicting load variations over short time windows for immediate operational needs. More recently, particularly in rapidly developing urban areas, LTLF has become increasingly important, highlighting the need for more detailed spatial forecasting approaches [6]. Willis [4] defines spatial load forecasting (SLF) as an LTLF process that forecasts the future geographic distribution of electric demand within the service territory of a distribution network. Through SLF, anticipated load is assigned to specific small areas or distribution assets within the network. In a later work, Willis [7] refined this concept by distinguishing between three types of forecasts. Small area forecasts divide the service territory into areas sufficiently small to meet distribution planning needs. Spatial forecasts extend this idea by forecasting all small areas on a common, coordinated basis. Geographic forecasts, by contrast, divide the service territory into areas too large for distribution planning purposes. In distribution networks, SLF is predominantly influenced by land-use change, whereas in transmission planning it is driven mainly by economic growth [8]. This concept is illustrated in Figure 2.1, which shows how predicted load can be distributed across a geographic territory.

2.1.1 Classification of SLF Methods

Numerous methodologies have been developed to forecast load growth from a spatial perspective. All these approaches partition the service territory into smaller areas to more effectively determine the geographic location of load growth [7]. According to Willis [7] and Evangelopoulos and

Georgilakis [9], SLF methodologies can be broadly classified into two principal categories: simulation methods, which model the load growth process by replicating spatial dynamics, and trend methods, which extrapolate historical load data into the future.

2.1.1.1 Simulation Methods

Simulation approaches model the load growth process to replicate the load history, thereby determining the timing, location, and magnitude of future loads. Their architecture typically relies on artificial intelligence methodologies including cellular automata theory, multi-agent systems, and evolutionary algorithms, enabling the capture of complex spatial dynamics that trend methods cannot easily replicate.

Wu and Lu [10] applied a data mining technique based on a Knowledge Discovery in Database (KDD) procedure in Kaohsiung, Taiwan, automatically identifying preferential scores associated with land use transitions for each small area. Miranda and Monteiro [11] developed a fuzzy inference model integrated with Geographic Information Systems (GIS) to capture the influence of spatial factors on load growth patterns in Porto, Portugal, producing maps of load growth potential for distribution network planning. Carreno and Padilha-Feltrin [12] proposed an evolutionary algorithm to identify future land use for vacant land in the utility service zone in Brazil, reporting a high rate of success when compared to city planning data. This work was subsequently extended by Carreno et al. [13], who implemented a cellular automata (CA) model for the spatiotemporal allocation of new loads on a grid resolution basis, tested on a real distribution system from a mid-size Brazilian city, demonstrating good performance with a reduced dataset.

He et al. [14] extended the CA approach by incorporating risk factors (including natural environment, land-use management, economic development, population, and urban policies) to forecast and allocate future loads in a Chinese urban area. Melo et al. [15] addressed the SLF problem using a multi-agent system applied to a mid-size Brazilian city, in which the service zone was divided into sub-zones each treated as an independent agent, requiring only basic information such as historical consumption data. Results showed a spatial error below 6% when compared to real data.

While simulation methodologies exhibit considerable flexibility and can capture complex spatial dynamics, they necessitate extensive training data and well-structured databases, and significant domain expertise is required for data extraction and model calibration. Additionally, computational complexity may escalate with higher spatial resolution, depending on the method adopted, which may constrain practical application in extensive service territories. Table 2.1 summarizes the approaches reviewed.

Table 2.1: Overview of simulation-based SLF methodologies.

Reference	Method	Case study	Key results	Limitations
Wu & Lu [10]	Data mining (KDD)	Urban area, Kaohsiung, Taiwan	Automatically identifies land use transition scores	Requires significant domain expertise and large historical databases
Miranda & Monteiro [11]	Fuzzy inference + GIS	Porto, Portugal	Produces spatial maps of load growth potential	Results depend on quality of expert-defined fuzzy rules
Carreno & Feltrin [12]	Evolutionary algorithms	Brazil	High rate of success compared to city planning data	Parameters not directly transferable between cities
Carreno et al. [13]	Cellular automata	Mid-size city, Brazil (∼200,000 residents)	Good performance with reduced dataset	Requires large historical datasets for more complex behaviours
He et al. [14]	Cellular Automata with risk factors	Chinese urban area	Improved accuracy by incorporating risk factors	Not all risk factors affecting urban development are considered
Melo et al. [15]	Multi-agent systems	Mid-size city, Brazil	Spatial error below 6% with basic input data	Requires minimum spatial information for each sub-zone, including development and redevelopment probabilities

2.1.1.2 Trend Methods

In contrast to simulation methodologies, trend approaches are grounded in historical data of annual peak load and extrapolate a time series into the future by fitting past load patterns to an appropriate mathematical function [7]. The choice of function and its parameters depends on factors such as weather conditions, land use characteristics, and seasonal patterns, which must be accounted for through normalization procedures before fitting [16], [17]. It is frequently necessary to determine the horizon year load (HYL), defined as the estimated load level that a small area will eventually reach over a long-term horizon of fifteen to twenty years [7]. The HYL represents the overall growth expected in an area, which may result from new consumer additions, from increasing per capita consumption among existing consumers, or from a combination of both [7].

A widely used function for modelling load growth in small areas is the S-curve, which captures the characteristic three-phase behaviour of development: a dormant period with no or minimal growth before the area begins to develop; a rapid growth phase as the area develops; and

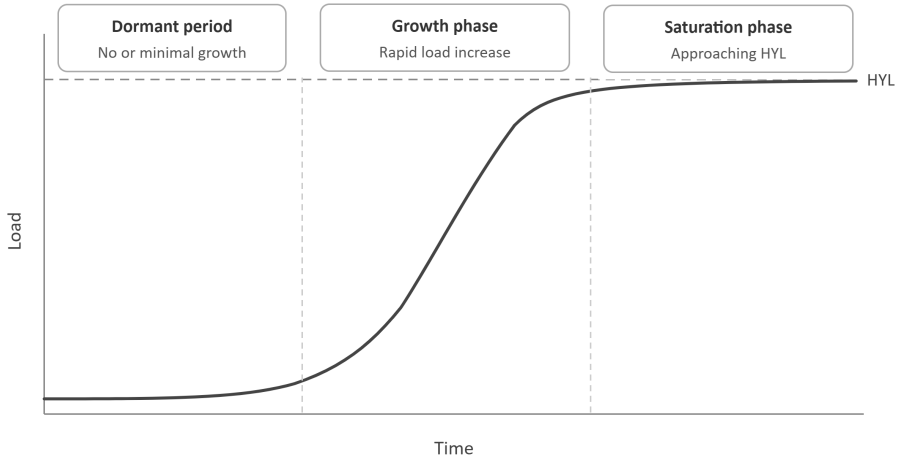


Figure 2.2: The three phases of the S-curve model of load growth.

a saturation phase as the area approaches its HYL [18], [16]. The S-curve naturally incorporates both sources of load growth, namely new consumers entering the area during the growth phase and increasing per capita consumption as the area matures, converging towards the HYL at saturation [7], as illustrated in Figure 2.2.

Vasquez-Arnez et al. [18] developed a methodology that adjusts historical load data to construct S-curves for small areas in a developing borough in the southeast of Brazil, reporting very satisfactory results compared to statistically allocated forecasts. Morales et al. [17] created an SLF tool combining S-curves, historical load data, and land use information, applied to distribution planning in Cuenca, Ecuador. Hong et al. [16] proposed a human-machine methodology that combines land use data with planners' domain knowledge to calculate the HYL for each small area in the service territory, incorporating planners' expertise to provide insights and heuristics that purely automated approaches cannot capture.

Trend methodologies offer simplicity and ease of implementation, making them appealing for utilities with constrained computational resources. However, they perform better at larger geographical scales, where aggregate patterns are more stable, and may struggle at the small area level where local variability and limited historical data reduce the reliability of the fitted functions. Furthermore, the dependence on historical patterns assumes that future development will follow similar trajectories, which may not hold in rapidly evolving contexts where consumption patterns change significantly over time. Table 2.2 summarizes the trend-based methodologies reviewed.

Table 2.2: Overview of trend-based SLF methodologies.

Reference	Method	Case study	Key results	Limitations
Vasquez-Arnez et al. [18]	S-curve fitting	Developing borough, SE Brazil	Very satisfactory results compared to statistically allocated forecasts	Methodology developed for developing boroughs; applicability to other contexts not fully demonstrated
Morales et al. [17]	S-curve (Gompertz) + land use	Cuenca, Ecuador	Execution time of approximately 2 hours compared to weeks manually, with higher accuracy	Results vary widely based on weights assigned to each factor; limited handling of empty areas
Hong et al. [16]	Human-machine methodology	Multiple US utility companies	Incorporates planners' expertise and heuristics that automated approaches cannot capture	Further development needed for grouping small areas and handling planners' inputs systematically

Simulation-based and trend-based methods represent a genuine trade-off rather than a strict hierarchy of quality. Simulation-based methods can capture complex spatial dynamics at the small-area level that trend-based methods cannot, but they require extensive data and domain expertise. Trend-based methods avoid these requirements but are generally less reliable at the small-area level, where simulation-based methods are strongest. Willis [7] argued that an ideal framework would combine the accuracy of simulation-based methods over long planning horizons with the simplicity and responsiveness of trend-based methods, suggesting that neither category is sufficient on its own. This conclusion is revisited in Section 2.3.

2.2 Hierarchical Forecasting Approaches

This section examines in detail the hierarchical methods within the trend-based category, which forecast multiple administrative or spatial levels simultaneously and reconcile the resulting forecasts to ensure mutual consistency.

The massive rollout of smart metering infrastructure has provided the utility industry with datasets of unprecedented granularity [5], [19]. Together with improvements in computational capacity and forecasting methods, this has created the conditions for hierarchical load forecasting (HLF) to emerge as a prominent methodology, encompassing forecasting at multiple levels of aggregation, from the household level to the corporate level, across various planning horizons [5].

Hierarchical forecasting can be implemented through two principal approaches: bottom-up and top-down [6]. In the bottom-up approach, forecasts are first generated at the lowest level of the hierarchy and subsequently aggregated to obtain forecasts for higher levels. By contrast, the top-down approach begins with a forecast at the aggregate level, which is then disaggregated to lower levels through weighting factors that allocate the total forecast across the hierarchy, typically on the basis of land use and geospatial characteristics [6]. A third approach, referred to as middle-out, combines both directions around a selected pivot level, aggregating forecasts upward above that level and disaggregating them downward below it. These approaches are examined in greater detail in Chapter 3.

Hyndman et al. [20] proposed an optimal combination method for hierarchical time series, in which all series at all levels are forecast independently and then revised using a generalized least squares (GLS) regression to produce coherent forecasts that are consistent across the hierarchy. Applied to Australian tourism demand data disaggregated by purpose of travel and geographical region, the method was shown to perform well compared to both the bottom-up and top-down approaches.

Hong [16] developed a hierarchical trending methodology for SLF combining bottom-up aggregation and top-down allocation. This approach was further refined by Evangelopoulos et al. [6], who proposed a probabilistic SLF framework based on the same hierarchical trending methodology, employing S-curves for fitting historical load data.

Hierarchical approaches represent a substantial advancement in SLF methodology, enabling coherent forecasts that improve overall accuracy [21]. However, trend-based hierarchical methods have limitations at finer spatial resolutions, where simulation methods tend to perform better [7].

2.3 Hybrid Approaches: Simulation and Hierarchical Methods

Section 2.1 established that trend-based and simulation-based methods offer complementary strengths, and that Willis [7] argued that an ideal framework would combine both. This section examines this literature in greater depth to assess the extent to which it has achieved such integration.

In the hierarchical component, Vasquez-Arnez et al. [18] and Morales et al. [17] used S-curve models to fit historical load data at the small-area level. Hong [16] proposed a hierarchical trending methodology combining bottom-up aggregation and top-down allocation to produce forecasts across multiple levels of the service territory. Evangelopoulos and Georgilakis [9] extended this hierarchical trending approach with S-curve fitting and Monte Carlo power flow simulation, constructing prediction intervals to assess the impact of load growth on distribution networks. More broadly, Hyndman et al. [20] developed optimal combination forecasts using GLS regression, achieving better performance than purely bottom-up or top-down approaches. Ye et al. [22] introduced a data-driven bottom-up approach that treats land plots as the basic forecasting unit, demonstrating greater applicability than benchmark methods in Chinese urban areas.

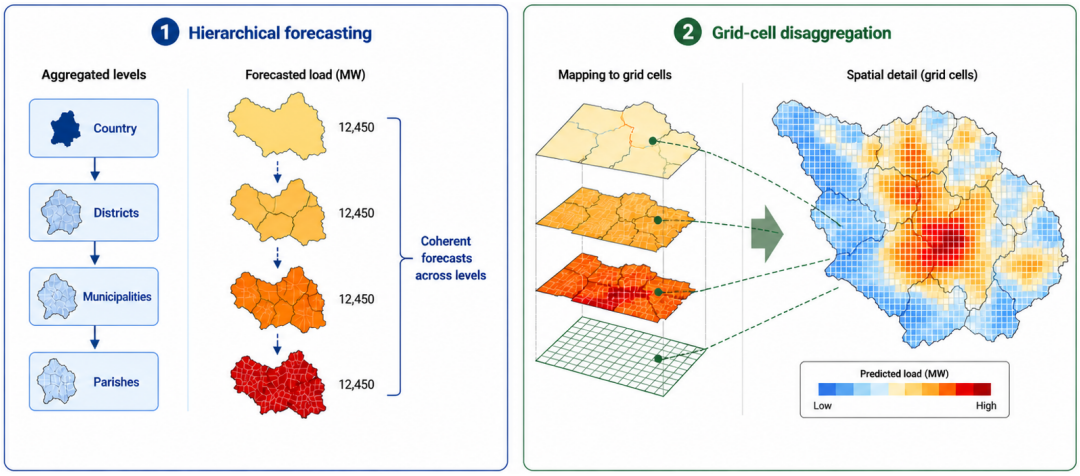


Figure 2.3: Hybrid framework linking hierarchical forecasting to grid-cell disaggregation.

On the simulation side, Carreno et al. [13] used cellular automata for the spatiotemporal allocation of new loads on a grid basis, and Melo et al. [15] applied multi-agent systems, achieving spatial errors below 6% with basic input information such as historical energy consumption data and a consumer inventory. Zambrano-Asanza et al. [23] combined artificial neural network with cellular automata model to define land-use transition rules, demonstrating that adequate normalization and the use of temporality in spatial factors improve spatial-temporal forecasting accuracy, and noting that their framework can be complemented with hierarchical load forecasting.

Full integration of hierarchical forecasting with simulation-based spatial disaggregation, as illustrated in Figure 2.3, remains an open challenge in the literature and represents a natural direction for future work.

2.4 Probabilistic Approaches, Uncertainty Quantification and Model Validation

The hierarchical and hybrid approaches reviewed above are predominantly deterministic, with the exception of the probabilistic framework introduced briefly in Section 2.2. This section examines probabilistic forecasting methods in greater detail, including that framework, as well as approaches for quantifying spatial error and the metrics used to assess spatial accuracy more broadly.

Deterministic SLF methods produce single-point forecasts that do not convey information about forecast uncertainty. In practice, however, decision-makers require not only a prediction of future load but also an assessment of the associated uncertainty, particularly for risk management and scenario planning in distribution network expansion [24], [9]. Proper quantification of spatial error is also essential for building confidence in SLF methods, as it identifies not only the magnitude of forecast errors but also their geographic distribution across the service territory [23].

One critical gap in some SLF methods is the absence of proper spatial error quantification

[23]. The purpose of calculating a spatial error metric is to determine whether forecast errors contribute to poor planning of distribution network expansion [7]. Melo et al. [25] investigated the determination of an appropriate spatial resolution for SLF using a grid-based model applied to a medium-sized city in Brazil, concluding that the proposed model represents the equilibrium between spatial error and computational effort, enabling a resolution suitable for distribution system planning.

Evangelopoulos et al. [6] proposed a probabilistic framework for SLF based on a hierarchical trending methodology, introducing the construction of prediction intervals (PIs) for probabilistic SLF (P-SLF) by employing the spatial root-mean-squared error (RMSE) across the service territory. Additional metrics employed include mean absolute error (MAE), prediction interval coverage probability (PICP), and probabilistic RMSE (p-RMSE), enabling evaluation of both the deterministic and probabilistic components of the methodology. The framework serves as a decision-making tool for DNOs and city planners, allowing them to choose between ordinary and more conservative forecasting scenarios. Results demonstrated improved accuracy and reduced spatial error compared to a conventional deterministic approach.

In a follow-up study, Evangelopoulos and Georgilakis [9] investigated the use of P-SLF results with Monte Carlo simulation (MCS) to assess the impact of load growth in primary distribution networks. Applied to a real-world distribution network in Greece, the results demonstrated that the proposed probabilistic SLF methodology can lead DNOs to more reliable decisions in comparison to single-point SLF approaches, by identifying extreme voltage violations and serving as a decision-making tool for regional planners.

Zambrano-Asanza et al. [23] proposed a complementary approach to spatial error assessment, validating their model using metrics such as overall accuracy, F1-score, kappa index, and Figure-of-Merit (FoM), measured during both training and testing phases. The FoM metric is particularly suited to SLF validation as it excludes persistent areas without changes, avoiding the overestimation of model performance that commonly used metrics such as overall accuracy and kappa index may produce.

Probabilistic approaches and rigorous spatial error assessment represent crucial advancements in SLF methodology. However, the construction of reliable prediction intervals requires sufficient historical data to characterize spatial error patterns, and the integration of Monte Carlo simulation with probabilistic SLF increases computational requirements [9].

2.5 Self-Consumption and Distributed Energy Resource Adoption

The methods reviewed so far address the forecasting and validation of spatial load in general terms. The uneven adoption of distributed energy resources is both a source of the spatial heterogeneity that these methods must account for and the specific focus of the self-consumption analysis in this dissertation.

The growth of residential self-consumption raises distinct challenges for distribution network planning, related both to predicting where new installations will occur and to assessing their impact on the network. Bernards et al. [1] developed statistical models based on logistic regression to estimate the geographical spread of photovoltaic, electric vehicle, and heat pump adoption, showing that socio-economic factors such as income, age, and household composition, together with peer effects for photovoltaics, significantly influence adoption probability at fine spatial resolution, motivated by the observation that local penetration levels of these technologies are highly spatially heterogeneous rather than evenly distributed. Avramidis et al. [2] used Monte Carlo simulation on real low-voltage feeders to evaluate demand-side flexibility schemes for accommodating high penetrations of residential distributed energy resources, showing that such schemes can increase the maximum problem-free penetration level by 10% to 40% without requiring excessive additional infrastructure investment. Heymann et al. [3] showed that the spatially heterogeneous diffusion of electric vehicle and photovoltaic adoption causes conventional, homogeneous allocation approaches in distribution network planning to systematically misestimate distributed energy resource impact, resulting in underinvestment during early adoption and overinvestment in later diffusion stages.

2.6 Research Gaps and Positioning of Current Work

Based on the literature reviewed in this chapter, several research gaps have been identified, as summarized in Table 2.3. These gaps are not equally relevant to the problem addressed in this dissertation. The temporal limitations of simulation-based models, listed in Table 2.3, and the systematic geographic quantification of spatial error, discussed in Section 2.4, concern the internal validity of individual forecasting methods and are largely independent of the specific case study considered. Forecasting in areas without historical data is similarly a general methodological limitation rather than one specific to Portugal or to the growth of self-consumption. By contrast, pairing hierarchical forecasting with self-consumption adoption analysis responds directly to the motivation set out in Chapter 1: distribution network planning increasingly needs to account for demand reshaped unevenly across a territory by distributed generation, while existing hierarchical forecasting studies do not incorporate this dimension. This is the gap addressed by this dissertation.

The choice of a trend-based hierarchical approach in this dissertation accepts the trade-off documented in Section 2.1. Specifically, the approach does not resolve the fine-scale, sub-parish spatial allocation for which simulation-based methods are better suited, a limitation discussed in Section 2.3. It does, however, avoid the temporal drift inherent in simulation methods calibrated once at a base year, as shown in Table 2.3, because it is re-fitted directly to the most recent historical data each time it is applied.

This dissertation develops and validates hierarchical forecasts at the finest administrative granularity reported in the reviewed literature, namely the parish level, across an entire national territory. It combines these forecasts with a spatially explicit analysis of self-consumption growth,

revealing penetration patterns that are not visible at coarser levels of aggregation. It also provides evidence on how the choice of reconciliation method interacts with the hierarchy and voltage scope under analysis, a question not directly examined in the reconciliation literature reviewed. The remaining gaps in Table 2.3 are not addressed here and remain open research directions, discussed further in Section 5.3.

Table 2.3: Research gaps identified in the SLF literature.

Gap	Description	Ref.	Addressed
Temporality in simulation models	CA models are typically calibrated once in the base year and evolved in discrete time steps, limiting their ability to capture dynamic spatial-temporal changes over time	[23]	No
Forecasting in areas without historical data	Trend methods rely on historical load data and cannot produce reliable forecasts for vacant or newly developing areas where no consumption history exists	[7], [22]	No
Spatial error quantification	Spatial error has not been systematically quantified in many SLF methods; overall accuracy metrics are commonly reported but the geographic distribution of errors is often overlooked	[23]	No
Self-consumption adoption and hierarchical forecasting studied independently	Statistical and diffusion-based models predict where distributed energy resource adoption will occur, and grid planning studies assess its network impact, but these analyses are rarely paired with a coherent hierarchical forecasting framework spanning multiple administrative levels	[1], [2], [3]	**Yes**

2.7 Summary

This chapter reviewed the literature underpinning this dissertation. Section 2.1 introduced spatial load forecasting, distinguishing simulation-based methods, which offer fine spatial resolution at the cost of extensive data requirements, from trend-based methods, which are simpler but less reliable at small spatial scales. Section 2.2 reviewed hierarchical forecasting approaches and the bottom-up, top-down, and middle-out reconciliation strategies compared in this dissertation. Section 2.3 examined attempts to combine simulation and hierarchical methods, concluding that their full integration remains an open challenge in the literature. Section 2.4 reviewed probabilistic approaches to spatial load forecasting and methods for validating their spatial accuracy. Section 2.5 reviewed the literature on self-consumption and distributed energy resource adoption, including its

spatial heterogeneity and its implications for distribution network planning. Finally, Section 2.6 identified the gaps that motivate and position this dissertation, most directly the absence of studies pairing hierarchical forecasting at fine spatial granularity with self-consumption analysis.

Chapter 3

Methodology

This chapter presents the methodological framework adopted in this dissertation for hierarchical time series forecasting. Section 3.1 introduces the base forecasting models, covering the ARIMA and Seasonal ARIMA models and their automatic selection through `AutoARIMA`, and the `Naive` benchmark. Section 3.2 describes the MSTL model used for time series decomposition. Section 3.3 presents the forecast reconciliation step, formalizing the hierarchical structure through the summing matrix S and describing three single-level reconciliation approaches: bottom-up, top-down, and middle-out. Section 3.4 defines the evaluation metrics used to assess forecast accuracy. Finally, Section 3.5 introduces the correlation coefficients used to quantify association between quantitative variables.

3.1 Base Forecasting Models

In time series forecasting, base forecasting models are fitted separately to each series in order to capture its underlying temporal dynamics. The resulting forecasts are then adjusted in a subsequent step to ensure consistency across the hierarchy, as described in Section 3.3.

ARIMA models are a widely used class of forecasting models that describe the current value of a series as a function of its past values and past errors, after differencing has been applied to achieve stationarity. These models are well suited to capturing trends and autocorrelation structures commonly observed in energy consumption data. When the data also exhibit seasonality, namely systematic fluctuations that recur at regular intervals such as annual heating and cooling cycles, the seasonal extension SARIMA is more appropriate because it captures these periodic patterns.

Two models are considered: `AutoARIMA`, which serves as the primary forecasting model and automatically selects the most appropriate ARIMA or SARIMA specification for each series, and `Naive`, which is used as a benchmark for comparison.

3.1.1 ARIMA and Seasonal ARIMA

The AutoRegressive Integrated Moving Average (ARIMA) model is one of the most widely used frameworks for time series forecasting. An ARIMA(p,d,q) model combines three components, each capturing a different aspect of the series behaviour: the autoregressive component of order p, which models the current value as a linear function of its p most recent past values; the integration component of order d, which applies d rounds of differencing to remove non-stationarity; and the moving average component of order q, which models the current value as a linear function of q past forecast errors. To express these components compactly, the backshift operator B is used, defined as $By_t = y_{t-1}$, so that $B^k y_t = y_{t-k}$ [26].

The autoregressive (AR) component of order p in backshift notation is written as

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y_t = c + \varepsilon_t \tag{3.1}$$

where $\phi_1, \ldots, \phi_p$ are the autoregressive coefficients, c is a constant, and ε_t is white noise.
The integration (I) component of order d is written as $(1-B)^d y_t$, where for $d = 1$

$$(1 - B) y_t = y_t - y_{t-1} \tag{3.2}$$

The moving average (MA) component of order q in backshift notation is given by

$$y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q) \varepsilon_t \tag{3.3}$$

where $\theta_1, \ldots, \theta_q$ are the moving average coefficients.
Combining these three components, the ARIMA(p,d,q) model is written as [26].

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1-B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q) \varepsilon_t \tag{3.4}$$

When the data exhibit seasonality, the seasonal extension SARIMA$(p,d,q)(P,D,Q)_s$ is more appropriate, as it forms part of the model space considered in Section 3.1.2. This model introduces additional seasonal autoregressive and moving average operators at lag s, where s denotes the seasonal period [26], [27]:

$$\phi(B)\,\Phi_P(B^s)\,(1-B)^d (1-B^s)^D\, y_t = \theta(B)\,\Theta_Q(B^s)\,\varepsilon_t \tag{3.5}$$

where:

- $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\Phi_P(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps}$ are the non-seasonal and seasonal autoregressive (AR) polynomials;
- $(1-B)^d$ and $(1-B_s)^D$ are the non-seasonal and seasonal integration (I) components, with d and D the respective differencing orders;
- $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ and $\Theta_Q(B^s) = 1 + \Theta_1 B^s + \cdots + \Theta_Q B^{Qs}$ are the non-seasonal and seasonal moving average (MA) polynomials.

Fitting an ARIMA or SARIMA model requires the manual selection of orders (p,d,q) and (P,D,Q), which is addressed automatically by `AutoARIMA`, as described in Section 3.1.2.

3.1.2 AutoARIMA

`AutoARIMA` handles order selection in ARIMA and SARIMA models by automatically identifying the specification that best balances goodness of fit and model complexity, avoiding overfitting without manual intervention. In practice, this selection is non-trivial because the number of candidate models grows rapidly as the maximum allowed orders increase, and evaluating each specification individually is computationally expensive. The algorithm used is a variation of the Hyndman-Khandakar algorithm [28], available in Python through the `statsforecast` library [29], and operates in two stages.

In the first stage, the differencing orders d and D are determined using root tests, which assess whether the series is stationary. If the original series is non-stationary, differencing is applied and the test is repeated until stationarity is achieved, up to a specified maximum order.

In the second stage, the remaining orders p, q, P, and Q are selected by minimizing the corrected Akaike Information Criterion (AICc) [28], a measure that penalizes model complexity relative to the goodness of fit. The correction term strengthens this penalty as the number of estimated parameters increases relative to the sample size, reducing the risk of overfitting. Rather than evaluating all possible combinations of orders, a stepwise search algorithm considers neighbouring models (those differing by one unit in a single order) and moves to the neighbour with the lowest AICc, continuing until no further improvement is found.

The result is an automatically selected ARIMA or SARIMA model that accommodates different seasonal patterns and degrees of non-stationarity without manual intervention.

3.1.3 Naive

The `Naive` model serves as a baseline against which the performance of `AutoARIMA` is evaluated. It produces forecasts by setting all future values equal to the last observed value of each series [28]:

$$\hat{y}_{T+h|T} = y_T \tag{3.6}$$

where T is the last time period for which observations are available, $h \geq 1$ is the forecast horizon, and y_T is the last observed value. The forecast is therefore identical across all horizons, with the prediction for one month ahead being the same as that for twelve months ahead, and it does not account for trend or seasonal structure in the data.

Despite its simplicity, the `Naive` model is a useful benchmark: if `AutoARIMA` does not outperform it, then the additional complexity of the ARIMA model is not justified for that series.

3.2 Time Series Decomposition

Time series decomposition is used to characterize the temporal structure of the data prior to forecasting. The Multiple Seasonal-Trend decomposition using Loess (MSTL) model extends the traditional STL framework to time series with multiple seasonalities, where the observed series is decomposed into seasonal, trend, and remainder components [30].

As defined in [30], MSTL decomposes a time series y_t as

$$y_t = \hat{S}_t^1 + \hat{S}_t^2 + \cdots + \hat{S}_t^n + \hat{T}_t + \hat{R}_t \tag{3.7}$$

where $\hat{S}_t^i$ denotes the i-th seasonal component, $\hat{T}_t$ the trend component, $\hat{R}_t$ the remainder component, and n the number of seasonal cycles present in y_t. For time series with a single seasonal pattern, the decomposition reduces to a seasonal term, trend, and remainder. For non-seasonal time series, MSTL returns only the trend and remainder components [30].

The MSTL algorithm begins by ranking the seasonal frequencies in ascending order, which helps prevent patterns at one seasonal scale from being absorbed by another. STL is then applied iteratively to each frequency, with an outer loop used to refine the estimated seasonal components. Once the outer loop is complete, the trend is extracted from the final STL iteration, and the remainder is obtained by subtracting the trend from the seasonally adjusted series [30].

Because Equation 3.7 is additive, each seasonal component $\hat{St}^i$ is expressed as a signed deviation from the trend: positive deviations indicate observations above the underlying trend, while negative deviations indicate observations below it. The trend $\hat{T}_t$ captures the long-term evolution of the series, whereas the remainder $\hat{R}_t$ captures the variation not explained by trend or seasonality. A small remainder relative to the trend therefore suggests that the decomposition captures the structure of the series well.

3.3 Forecast Reconciliation

Forecast reconciliation is applied as a post-processing step to ensure that independently generated forecasts satisfy the aggregation constraints of a hierarchy. Without reconciliation, forecasts produced separately at different levels are internally inconsistent: the sum of a node's children forecasts does not necessarily equal the forecast of that node, and analogous discrepancies arise throughout the hierarchy. The reconciliation step adjusts the forecasts jointly so that they are coherent across the full hierarchy, from the bottom level to the top-level aggregate.

3.3.1 Hierarchical Structure of the Model

The forecasting model is organized around a hierarchy with multiple levels of aggregation, from a single top-level total down to a set of bottom-level series, which form the finest level of granularity in the hierarchy. Each series above the bottom level is obtained by summing a subset of the bottom-level series, following the hierarchy's branching structure.

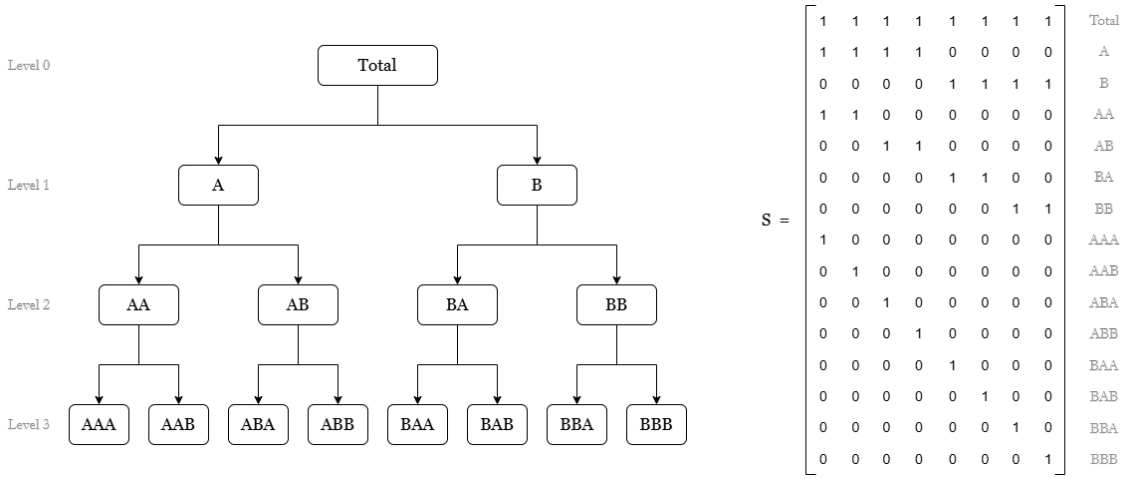


Figure 3.1: Illustrative example of a four-level hierarchical structure and its summing matrix S.

The aggregation relationships are encoded in a summing matrix S of dimension $n \times m$, where n denotes the total number of series across all levels m the number of bottom-level series. Let y_t denote the vector of all observations at time t and b_t the vector of bottom-level observations. The hierarchical structure is then written as [26]

$$y_t = Sb_t \tag{3.8}$$

Each bottom-level series contributes a value of 1 to every ancestor node in its path through the hierarchy, and the bottom block of S corresponds to an identity matrix I_m. Figure 3.1 presents this structure for a small illustrative hierarchy with two Level-1 nodes, four Level-2 nodes, and eight bottom-level nodes, alongside the corresponding summing matrix S.

3.3.2 Single-level Reconciliation Approaches

Any set of coherent forecasts can be expressed as

$$\tilde{y}_h = SG\hat{y}_h \tag{3.9}$$

where $\hat{y}_h$ denotes the base forecasts at horizon h and G is a mapping matrix that determines how information is extracted from the base forecasts before being re-aggregated through S, [26].

The three single-level approaches implemented in this dissertation correspond to different specifications of G, as illustrated in Figure 3.2. All three approaches are available through the open-source `hierarchicalforecast` Python library [31], which provides a unified framework for hierarchical forecast reconciliation and has been adopted in recent forecasting research [21].

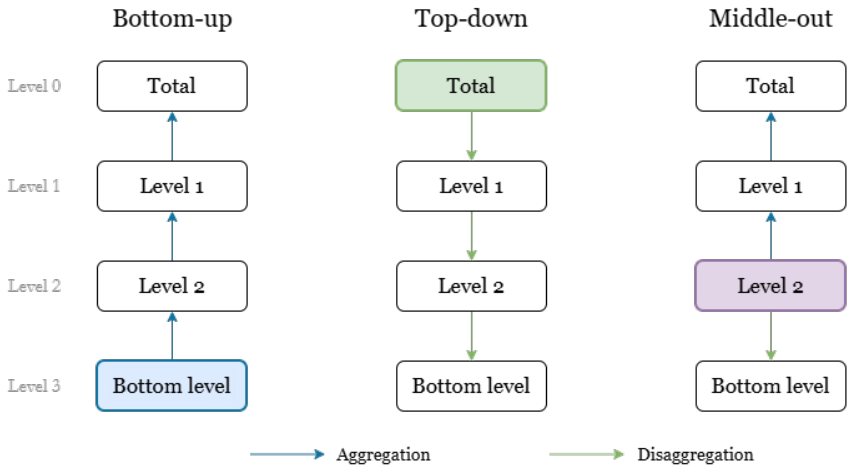


Figure 3.2: The three single-level reconciliation approaches.

3.3.2.1 Bottom-up

In the bottom-up approach, only the bottom-level base forecasts are used, and all upper levels are obtained by direct aggregation. This corresponds to

$$G = \left[0_{m\times(n-m)} \mid I_m\right] \tag{3.10}$$

so that G extracts only the bottom-level rows of $\hat{y}_h$, discarding the base forecasts produced independently for all upper levels. The summing matrix, S, then reconstructs every higher-level series by simple addition, so coherence is guaranteed by construction.

The main drawback follows directly from Equation 3.10 since any systematic error in a bottom-level base forecast propagates upward through S without correction, so the quality of the reconciled forecasts at the upper levels is bounded by the quality of the weakest bottom-level models.

This property is particularly appealing when patterns are expected to vary substantially across bottom-level series, since the full hierarchy directly inherits the dynamics captured by the bottom-level models and no information about local dynamics is discarded.

3.3.2.2 Top-down

In the top-down approach, G retains only the forecast for the top-level aggregate and allocates it across the m bottom-level series through a vector of proportions $p = [p_1, \ldots, p_m]^T$, where p_i denotes the share assigned to bottom-level series i. These proportions must satisfy $\sum_{i=1}^{m} p_i = 1$. In matrix form, G is given by

$$G = \left[p \mid 0_{m\times(n-1)}\right] \tag{3.11}$$

The `forecast_proportions` method computes the allocation proportions from the base forecasts themselves rather than from historical averages. For a hierarchy with K levels of disaggregation, these proportions are obtained recursively from the top to the bottom of the hierarchy,

with each step computing the share of a node's base forecast relative to the combined base forecast of all nodes that share the same parent:

$$p_j = \prod_{\ell=0}^{K-1} \frac{\hat{y}_{j,h}^{(\ell)}}{\hat{S}_{j,h}^{(\ell+1)}} \tag{3.12}$$

where $\hat{y}_{j,h}^{(\ell)}$ is the base forecast of the node located ℓ levels above bottom-level series j, and $\hat{S}_{j,h}^{(\ell+1)}$ is the aggregate base forecast of all children of the node located $\ell+1$ levels above bottom-level series j. This procedure results in coherent forecasts by disaggregating the top-level base forecast across the full hierarchy.

The main limitation of the top-down approach is that the allocation proportions are derived from the base forecasts, which may not adequately capture local dynamics at more disaggregated levels. Local variables can substantially alter the proportions across bottom-level series in ways that the base forecasts fail to reflect, potentially compromising forecast accuracy at the bottom level.

3.3.2.3 Middle-out

The middle-out approach combines the two mappings above around a pivot level, [26]. In our example case, the pivot level is set to Level 2, as it represents an intermediate level of aggregation that balances the reliability of higher-level forecasts with the granularity of lower-level series. For the levels above the pivot, forecasts are obtained by aggregating the Level-2 base forecasts upward, exactly as in Equation 3.10, while treating Level 2 as the new bottom level. For the bottom level below the pivot, forecasts are obtained by disaggregating each Level-2 base forecast into its constituent bottom-level series through allocation proportions, as in Equation 3.12 with $K = 1$.

Combining these two directions, G takes the block form:

$$G = \left[0_{m\times(1+n_1)} \mid P \mid 0_{m\times m}\right] \tag{3.13}$$

where n_1 is the number of Level-1 nodes, and P is an $m \times n_2$ allocation matrix that maps Level-2 forecasts to bottom-level forecasts, where n_2 is the number of Level-2 nodes. Each row j of P contains a single nonzero entry, equal to p_j, in the column corresponding to the Level-2 node containing bottom-level series j. In this way, G extracts only the Level-2 forecasts and allocates each one to its constituent bottom-level series, and the summing matrix S then reconstructs all upper levels by aggregating the resulting bottom-level forecasts upwards.

As a result, middle-out combines the strengths of both directions: above the pivot level, forecasts are obtained by aggregation and therefore preserve the accuracy of the Level-2 base forecasts; below the pivot level, forecasts are allocated through shares, which may introduce bias in the bottom-level forecasts.

3.4 Evaluation Metrics

The performance of the forecasting models and reconciliation methods is evaluated using three point forecast accuracy metrics: the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), and the Mean Absolute Scaled Error (MASE). MAE and RMSE are widely used scale-dependent measures of forecast accuracy, while MASE is a scale-independent metric that allows comparison across series with different scales [26].

Following [26], let e_t denote the forecast error at time t, defined as the difference between the observed value y_t and the corresponding forecast $\hat{y}_t$:

$$e_t = y_t - \hat{y}_t \tag{3.14}$$

The MAE measures the average magnitude of forecast errors:

$$\text{MAE} = \text{mean}(|e_t|) \tag{3.15}$$

The RMSE penalizes larger errors more heavily, as deviations are squared before averaging:

$$\text{RMSE} = \sqrt{\text{mean}(e_t^2)} \tag{3.16}$$

Both MAE and RMSE are scale-dependent metrics, meaning their values depend on the scale of the series being forecast. In a hierarchical setting, where series at different levels of aggregation have very different scales, these metrics should be compared within each level rather than across levels [26].

The MASE addresses this limitation by scaling the forecast errors relative to the in-sample errors of a Seasonal Naive benchmark, defined as a model that sets each future value equal to the last observed value from the same season [26]. This benchmark is computed once, from the training data alone, and used as a normalizing constant in the denominator of Equation 3.18. The scaled error is defined as:

$$\text{MASE} = \text{mean}(|q_j|) \tag{3.17}$$

where

$$q_j = \frac{e_j}{\frac{1}{T-s}\sum_{t=s+1}^{T}|y_t - y_{t-s}|} \tag{3.18}$$

and T is the length of the training series and s is the seasonal period [26]. A MASE below 1 indicates that the model outperforms the Seasonal Naive benchmark, whereas values above this threshold indicate poorer performance. As a scale-independent metric, MASE allows comparison of forecast accuracy across series at different levels of the hierarchy [26].

3.5 Correlation Analysis

Pearson's correlation coefficient and Spearman's rank correlation coefficient quantify the association between pairs of quantitative variables. Pearson's coefficient characterizes linear association, whereas Spearman's coefficient provides a complementary rank-based measure of monotonic association [32].

The Pearson correlation coefficient, denoted by r, summarizes the linear relationship between two variables, x and y [26]:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{3.19}$$

where x_i and y_i denote paired observations, $\bar{x}$ and $\bar{y}$ are the corresponding sample means, and n is the number of paired observations. The coefficient lies between -1 and 1: its sign indicates whether the linear relationship is negative or positive, while its magnitude indicates its strength. A value close to zero indicates little or no linear relationship, but does not exclude the presence of a nonlinear relationship.

For a simple linear regression model fitted by ordinary least squares with an intercept, the coefficient of determination is equal to the squared Pearson correlation coefficient, $R^2 = r^2$. In this setting, R^2 represents the proportion of sample variation in the response variable accounted for by the fitted linear relationship. Since squaring removes the sign of r, R^2 does not indicate the direction of the association. It is interpreted as a descriptive measure of the variation accounted for by the linear relationship and does not provide evidence of causality [32].

Spearman's rank correlation coefficient, denoted by ρ, is calculated by replacing the original observations with their ranks and then evaluating the Pearson correlation of the ranked values:

$$\rho = \frac{\sum_{i=1}^{n}(R(x_i) - \overline{R(x)})(R(y_i) - \overline{R(y)})}{\sqrt{\sum_{i=1}^{n}(R(x_i) - \overline{R(x)})^2}\sqrt{\sum_{i=1}^{n}(R(y_i) - \overline{R(y)})^2}}, \tag{3.20}$$

where $R(x_i)$ and $R(y_i)$ denote the ranks assigned to x_i and y_i, respectively. In the presence of ties, observations are assigned the mean of the ranks that they would otherwise occupy. As with Pearson's coefficient, ρ ranges from -1 to 1: its sign indicates the direction and its magnitude the strength of the monotonic association. Unlike Pearson's coefficient, this association need not be linear. Spearman's coefficient is applicable to ordinal variables and provides a complementary rank-based assessment for continuous variables when the relationship may be monotonic but nonlinear or when extreme values may substantially affect Pearson's coefficient. Extreme observations should nevertheless be examined rather than removed automatically [32].

Pearson's and Spearman's coefficients can be interpreted together with scatter plots to assess whether an observed association is adequately represented by a linear relationship or may instead be monotonic, nonlinear, or sensitive to influential observations. Correlation coefficients do not adequately characterize all possible relationships: in particular, a coefficient close to zero does not exclude a nonlinear or non-monotonic relationship [32].

The magnitude of r and ρ is commonly interpreted using their absolute values. The conventional guidelines summarized in Table 3.1 provide a descriptive classification, but their boundaries are arbitrary and should be interpreted in the context of the scientific question rather than as strict decision rules [32]. The sign of each coefficient describes the direction of the association.

Table 3.1: Conventional interpretation of correlation coefficient magnitude, following [32].

Absolute Magnitude	**Interpretation**
0.00–0.10	Negligible correlation
0.10–0.39	Weak correlation
0.40–0.69	Moderate correlation
0.70–0.89	Strong correlation
0.90–1.00	Very strong correlation

Correlation coefficients quantify association and do not, by themselves, provide evidence of causality. Two variables may be correlated without either causing the other, for example when both are influenced by a third variable not included in the analysis [26]. Statistical significance, where assessed, is evaluated using a two-sided test of the null hypothesis of zero correlation at $\alpha = 0.05$. A statistically significant result does not establish a causal relationship or practical importance, whereas a non-significant result does not prove that no association exists. Interpretation should consider the coefficient magnitude, sample size, confidence interval, and graphical inspection of the paired observations [32].

3.6 Summary

This chapter presented the methodological framework adopted in this dissertation. In the base forecasting stage, `AutoARIMA` was selected as the primary model for its ability to automatically identify the most appropriate ARIMA or SARIMA specification for each series, while `Naive` serves as a benchmark for comparison. The MSTL model was introduced as a tool for time series decomposition, enabling the characterization of trend and seasonal components prior to forecasting.

In the reconciliation stage, the hierarchical structure was formalized through the summing matrix S, and three single-level approaches were presented. Bottom-up aggregates bottom-level forecasts upward, preserving local dynamics but propagating bottom-level errors through the hierarchy. Top-down disaggregates the top-level forecast through allocation proportions, preserving the aggregate-level forecast but remaining blind to local heterogeneity at more disaggregated levels. Middle-out combines both directions around a pivot level, preserving the accuracy of the pivot-level forecasts above and allocating them downward through proportions below.

Three evaluation metrics were then defined: MAE and RMSE, which are scale-dependent and should be compared within each hierarchical level, and MASE, which is scale-independent and allows comparison across levels.

Finally, Pearson's and Spearman's correlation coefficients were introduced as complementary measures of linear and monotonic association between quantitative variables, together with the coefficient of determination and conventional guidelines for interpreting the magnitude of these coefficients.

The specific configuration and implementation of these methods for the case study are described in Chapter 4.

Chapter 4

Case Study, Results and Discussion

This chapter applies the hierarchical forecasting framework described in Chapter 3 to mainland Portugal as a case study. Section 4.1 describes the study area and the E-REDES consumption data used, including its coverage and known limitations. Section 4.2 presents an exploratory analysis of the data, including a time series decomposition of the national consumption series and an assessment of self-consumption growth in mainland Portugal. Section 4.3 details the implementation of the forecasting pipeline and the configuration of the base models. Sections 4.4, 4.5, and 4.6 report the forecasting results for the low voltage, medium, high, and very high voltage, and all voltage analyses, respectively. Finally, Section 4.7 compares the three analyses, examines the January 2026 forecast anomaly, analyses parish-level spatial patterns in forecast accuracy and their relationship with self-consumption density and hierarchical error propagation, discusses the implications of the results for self-consumption and distribution network planning, and presents forecasts beyond the evaluation period.

4.1 Case Study and Data Sources

This section describes the geographical scope of the case study, the datasets used, and the preprocessing steps applied prior to modelling.

4.1.1 Study Area

The case study covers mainland Portugal, whose territory is organized around a four-level administrative hierarchy: the national aggregate (Level 0), 18 districts (Level 1), 278 municipalities (Level 2), and parishes (Level 3), as illustrated in Figure 4.1. This administrative structure directly defines the hierarchical structure of the forecasting framework described in Chapter 3.

4.1.2 Data Sources and Preprocessing

The data used in this dissertation were obtained from the E-REDES Open Data Portal [33], which publishes electricity distribution data for mainland Portugal. E-REDES is the main electricity

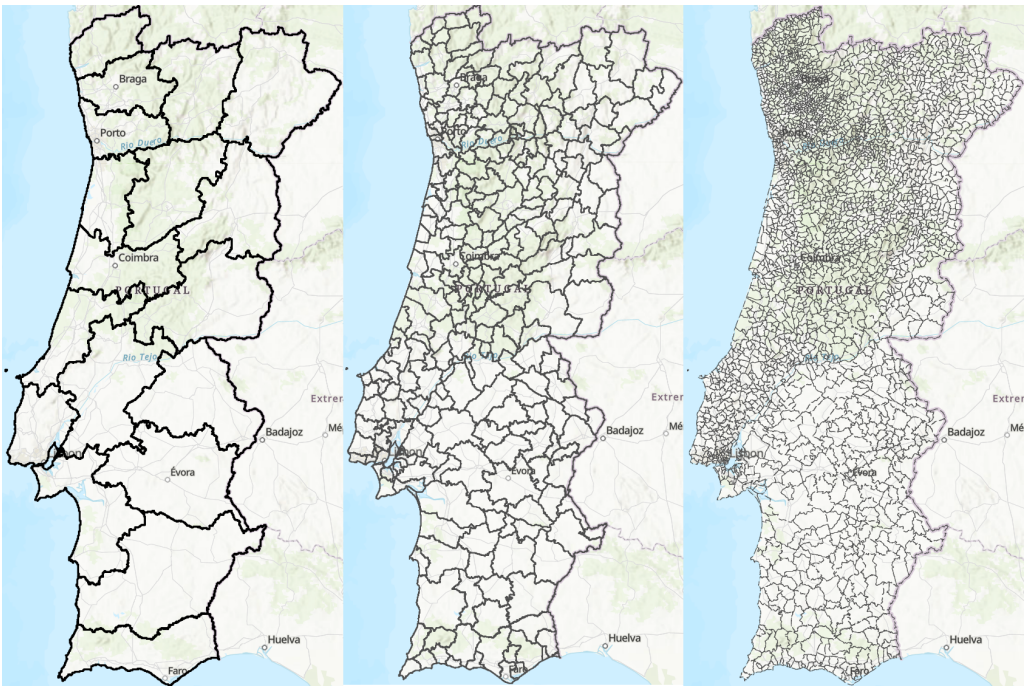


Figure 4.1: Administrative hierarchy of mainland Portugal: districts (left), municipalities (center), and parishes (right).

distribution network operator in mainland Portugal, responsible for the operation and maintenance of the very high, high, medium, and low voltage distribution grids. The portal was the sole data source for this work, providing billed consumption and self-consumption data openly available at the parish level, which is the spatial granularity required by the hierarchical framework.

This study relies on two datasets. The primary dataset, *Monthly consumption by municipality* [34], reports monthly active energy consumption billed since November 2020, disaggregated by district, municipality, parish, and voltage level. The year and month in each record refer to the consumption period rather than the billing date. For data protection reasons under the General Data Protection Regulation (GDPR), consumption that cannot be disclosed at the parish level is aggregated by E-REDES into a residual category identified as *OUTROS* followed by the district name.

The secondary dataset, *Total production Units for Self-Consumption - Monthly* [35], reports the monthly number of certified self-consumption units (UPAC) connected to the E-REDES network with an active contract on the last day of each month, together with their total installed capacity (kW). Records are disaggregated by district, municipality, parish, and voltage level, and technology type, and cover the period from January 2023 to June 2026.

Each series in the hierarchy is identified by a six-digit code following the DICOFRE standard [36], in which the first two digits identify the district, the next two the municipality, and the final two the parish. Higher-level series are obtained by truncating this code: district level series use the first two digits and municipality-level series the first four. The national aggregate is assigned the identifier `PT`. This encoding ensures that the hierarchical relationships between series are directly

recoverable from their identifiers, and that the consumption and self-consumption datasets can be matched at every level of aggregation.

4.1.2.1 Voltage levels

The consumption dataset covers the period from November 2020 to April 2026, yielding 66 monthly observations per series. It distinguishes two voltage categories: low voltage (*Baixa Tensão*, BT) and very high, high, and medium voltage (*Muito Alta, Alta e Média Tensões*, MAT, AT, MT). Because the latter aggregates three voltage levels into a single class, the dataset does not allow medium, high, and very high voltage consumption to be analyzed separately, which constitutes a limitation of the available data. Three analyses were therefore conducted: one restricted to low voltage, one restricted to medium, high, and very high voltage, and another including all voltage levels combined. The share of low voltage in total consumption is seasonal, peaking in December and January, at around 54.3% on average and reaching 57.9% in the winter of 2025, and falling below 46.5% on average between May and September, with troughs recurring each year in late spring or early autumn. This pattern reflects the greater weight of residential demand during the colder season.

4.1.2.2 Records aggregated for data protection

The *OUTROS* records are identified by placeholder codes in which the municipality and parish digits are masked, so that they can only be recovered at the district level and cannot be attributed to any municipality or parish. There is exactly one such identifier per district, giving 18 in total; each identifier appears once per month and voltage category for which E-REDES has aggregated consumption to report. Together, these records account for 1.2% of the total energy in the dataset, corresponding to 3.05 TWh out of 254.91 TWh, and are concentrated almost entirely in medium, high, and very high voltage, with a low-voltage contribution below 0.001 TWh. Because the hierarchical framework requires every bottom-level series to be assignable to a single parish, these records were excluded from both analyses. Their exclusion has a negligible effect on the low-voltage analysis and reduces aggregate consumption by approximately 1.2% in the all-voltage analysis.

4.1.2.3 Coverage of the distribution network

E-REDES is the dominant, but not the sole, distribution network operator in mainland Portugal. Low-voltage distribution is a municipal responsibility that may be concessioned or operated directly, and eleven distribution network operators are active in mainland Portugal: E-REDES, which holds 99.49% of low-voltage contracts and is present in every district and municipality, and ten smaller operators, mostly electricity cooperatives, that act exclusively at low voltage and together account for the remaining 0.51% [37]. These ten operators serve customers in eight municipalities, with more than one operator active in the same municipality in some cases, for example Paredes, served by LORD and A CELER, and Vila Nova de Famalicão, served by CEVE and CESSN.

Table 4.1: Parishes absent from the low-voltage analysis and corresponding cooperative operators.

DICOFRE	Parish	Municipality	Cooperative operator(s)
011304	Loureiro	Oliveira de Azeméis	CEL
030831	Moreira de Cónegos	Guimarães	A Eléctrica Moreira de Cónegos
031223	Louro	V. N. de Famalicão	CEVE, CESSN
031227	Nine	V. N. de Famalicão	CEVE, CESSN
031255	UF Gondifelos, Cavalões e Outiz	V. N. de Famalicão	CEVE, CESSN
031256	UF Lemenhe, Mouquim e Jesufrei	V. N. de Famalicão	CEVE, CESSN
031257	UF Ruivães e Novais	V. N. de Famalicão	CEVE, CESSN
131419	Roriz	Santo Tirso	CoopRoriz, CE de Vilarinho
131432	Vilarinho	Santo Tirso	CoopRoriz, CE de Vilarinho

The consequence for the dataset is that, among the 2882 parishes present when all voltage levels are considered, excluding the aggregated *OUTROS* records, 2873 have low-voltage records and nine do not. In these nine parishes, the low-voltage network is operated entirely by a cooperative, so no E-REDES billing data are available at low voltage; the remaining consumption recorded in these parishes at medium, high, and very high voltage remains under E-REDES operation. As shown in Table 4.1, the nine parishes are located in four of the eight municipalities where a cooperative operator is present: Vila Nova de Famalicão, with 5 of 34 parishes; Santo Tirso, with 2 of 14; Guimarães, with 1 of 48; and Oliveira de Azeméis, with 1 of 12. In the other four cooperative municipalities, Barcelos, Paredes, Águeda, and Covilhã, E-REDES low-voltage coverage remains complete at the parish level because the cooperative operator serves only part of one or more parishes, while E-REDES continues to serve the remaining customers in the same parishes.

The low-voltage hierarchy therefore covers the E-REDES concession area rather than the entirety of mainland Portugal, and the results should be interpreted accordingly.

A related case is the parish of Rebordosa (DICOFRE code 131018, Paredes, Porto district), whose low-voltage records span only from May 2021 to October 2025, yielding 48 months of training data instead of 54. Rebordosa is located in the municipality of Paredes, where two cooperative operators hold low-voltage concessions, one of which is A CELER - Cooperativa Electrificação de Rebordosa [37], consistent with the partial coverage observed. The parish was retained in the analysis, as 48 months still span four complete annual cycles; however, it is the only series in the hierarchy with an incomplete training window, and its test window is likewise limited to six of the twelve held-out months.

A similar, and more pronounced, pattern of incomplete coverage affects the medium, high, and very high voltage segment. Of the parishes for which any consumption record is available, only 1626 have a medium, high or very high voltage record during the training period. Of these, only 1531 have a complete 54-month training history. The remaining 95 parishes were excluded from the medium, high, and very high voltage hierarchy used in Section 4.5, since many medium, high, and very high voltage connections, often large and intermittent industrial or commercial

loads, do not report consumption continuously throughout the period. Coverage in the test period is imperfect even among these 1,531 parishes. One parish has no test-period records at all, and twelve others have between one and eleven of the twelve test months. This sparser and more fragmented coverage means that the parishes retained for the medium, high, and very high voltage analysis are not a representative sample of the segment as a whole, but rather those with the most continuous consumption histories.

Finally, the dataset reflects the administrative division of mainland Portugal in force before the reform introduced by Law 25-A/2025 of 13 March [38]. Although the reform took effect with the 2025 local elections, the E-REDES dataset continues to use the pre-2025 codification throughout its full time span, so all parish-level analyses in this dissertation are based on the pre-2025 administrative boundaries.

4.2 Exploratory Data Analysis

Prior to forecasting, an exploratory analysis was conducted to characterize the temporal structure of the consumption series and the spatial distribution of self-consumption installations.

4.2.1 Time Series Decomposition

The MSTL model was applied to the national aggregate of the training set in each analysis, decomposing the observed series into trend, seasonal, and remainder components, as described in Section 3.2.

Figure 4.2 presents the decomposition of the low-voltage consumption series. The trend component rose from 1.75 TWh/month in November 2020 to 1.96 TWh/month in April 2025, a growth of 11.8% over the training period. The seasonal component follows a clear annual pattern, peaking at 0.39 TWh above trend in January and reaching a minimum of 0.18 TWh below trend in June, consistent with higher residential demand in winter. Although the remainder shows visible fluctuations, it remains relatively small compared with the observed series, with a standard deviation of 0.04 TWh, suggesting that the trend and seasonal components account for most of the variation in the series.

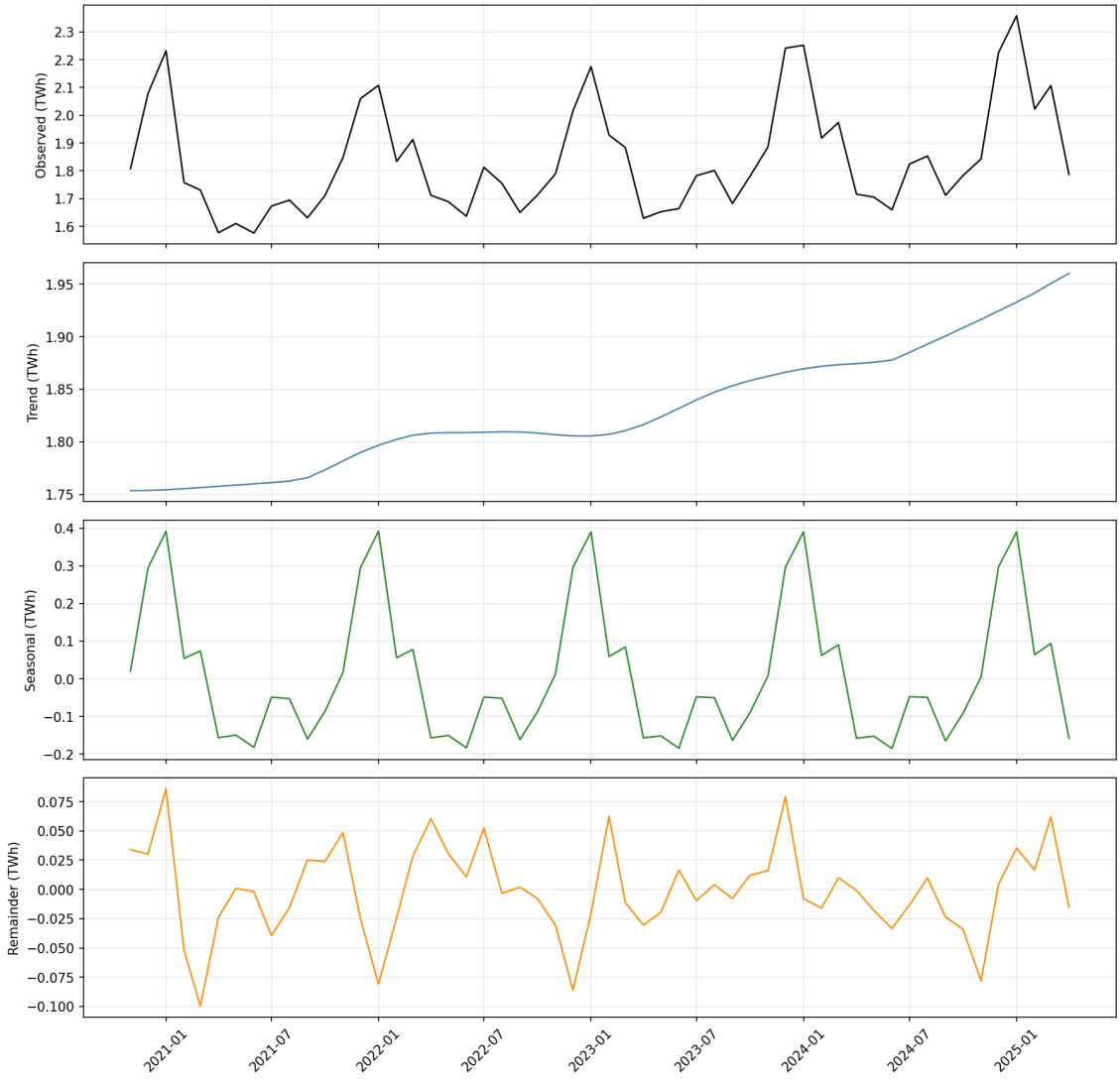


Figure 4.2: MSTL decomposition of the national low-voltage consumption series (November 2020 - April 2025).

Figure 4.3 presents the same decomposition for the medium, high, and very high voltage segment, computed on the filtered subset of 1,531 parishes with complete training histories described in Section 4.1. The trend grew from 1.89 TWh per month in November 2020 to 1.96 TWh per month in April 2025, corresponding to a growth of 3.4% over the training period, which is substantially below the 11.8% observed at low voltage. Unlike the low-voltage series, whose seasonal component peaks in January, the medium, high, and very high voltage seasonal component peaks in July, at 0.13 TWh above trend, and reaches its minimum in February, at 0.14 TWh below trend. This pattern is consistent with a summer-weighted consumption profile, more typical of industrial cooling and seasonal commercial activity than of residential heating. The remainder standard deviation is 0.03 TWh, or 1.4% of the mean trend level. In relative terms, this is smaller than for the low-voltage series (2.1%), although this may in part reflect the selection of more continuously recorded parishes, rather than a property of the segment as a whole.

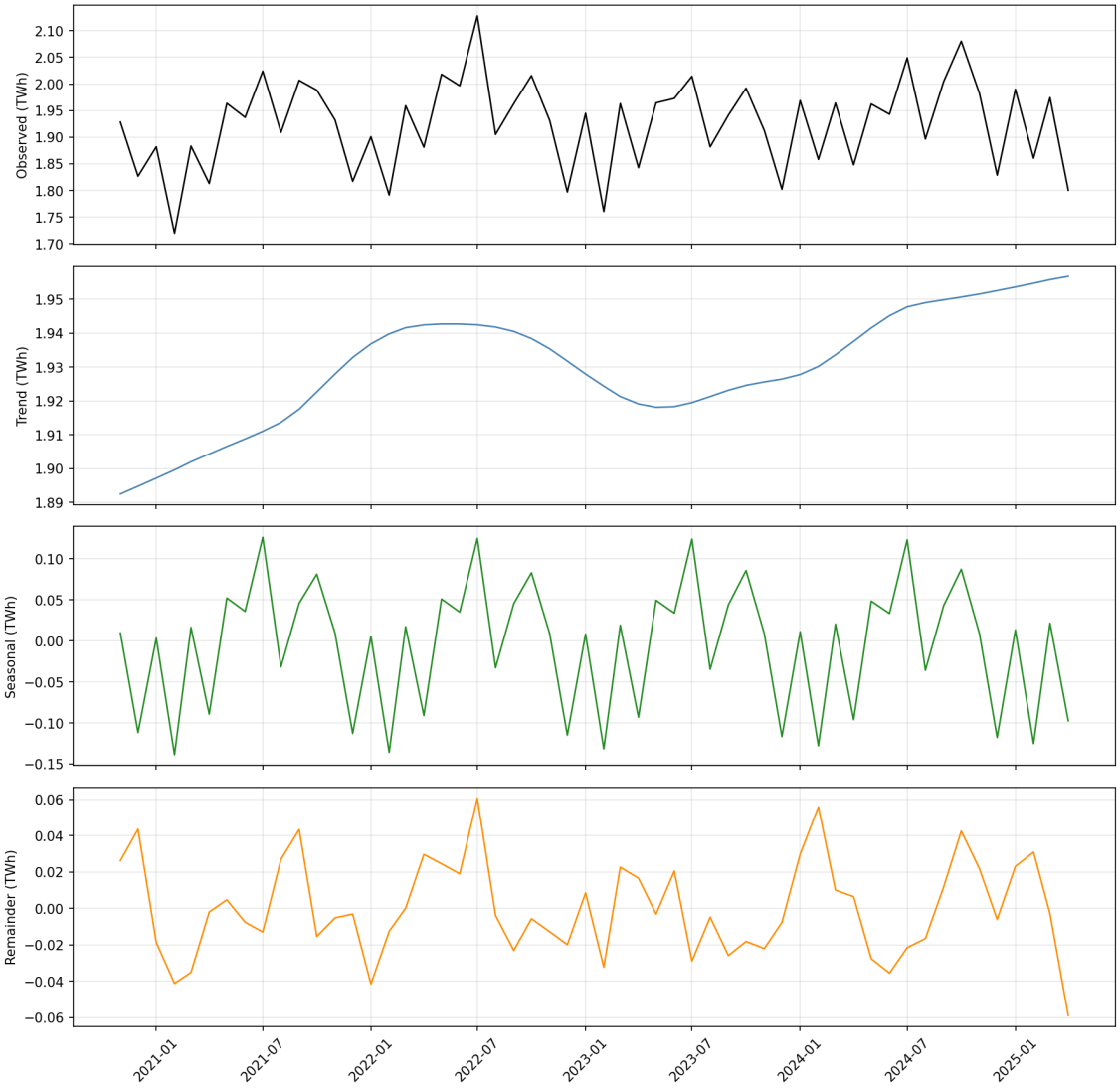


Figure 4.3: MSTL decomposition of the national medium, high, and very high voltage consumption series (November 2020 - April 2025).

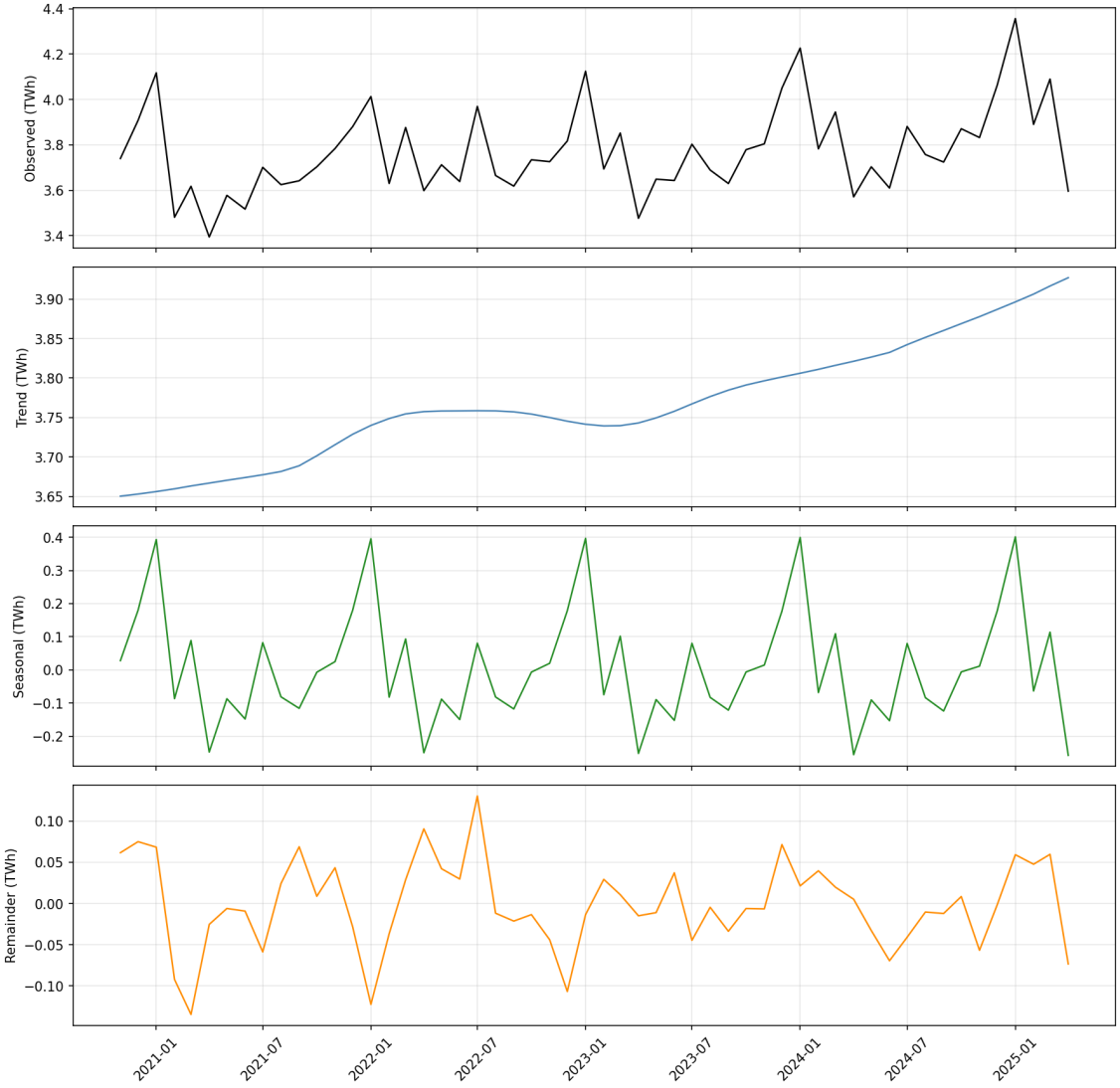


Figure 4.4: MSTL decomposition of the national all-voltage consumption series (November 2020 - April 2025).

Figure 4.4 presents the MSTL decomposition of the all-voltage series, which aggregates low-voltage and medium- and high-voltage consumption. Over the sample period, the trend increased from 3.65 TWh/month to 3.93 TWh/month, corresponding to growth of 7.6%. Low-voltage consumption accounted for approximately half of the all-voltage trend throughout the period, increasing from 48.0% of the trend level in November 2020 to 49.9% in April 2025. The medium, high, and very high voltage segment grew by 3.4% over the same period (Figure 4.3), substantially below the 11.8% observed at low voltage. This pattern may be related to the larger share of non-residential consumption at medium, high, and very high voltage, which is less exposed to residential electrification processes, such as heat pump and electric vehicle adoption, typically associated with rising low-voltage demand over this period. The seasonal component peaks at 0.40 TWh above trend in January and reaches a minimum of 0.26 TWh below trend in April. This minimum falls between the June minimum observed at low voltage and the February minimum

observed at medium, high, and very high voltage, consistent with the all-voltage seasonal pattern reflecting a blend of the two underlying, and markedly different, seasonal profiles. Although the remainder exhibits visible fluctuations, it remains relatively small compared with the observed series, with a standard deviation of 0.05 TWh. In relative terms, the ratio of the remainder standard deviation to the trend level is lower in the all-voltage case (1.4%) than in the low-voltage case (2.1%). This is consistent with the medium-, high-, and very-high-voltage segment itself exhibiting a lower relative remainder, also 1.4%, so that aggregating the two segments does not increase, and may reduce, the relative variability of the combined series.

4.2.2 Self-Consumption Growth Analysis

The UPAC dataset covers several generation technologies, but as of April 2026 solar photovoltaic accounts for 99.2% of installed capacity and 99.99% of the number of self-consumption units in the dataset, with the remainder split among wind, biogas, hydro, biomass, and non-renewable cogeneration. The analysis in this section is therefore restricted to solar PV installations, which reveal sustained growth in self-consumption across mainland Portugal. Between January 2023 and April 2026, the national installed capacity of low-voltage solar PV units grew from 301 MW to 934 MW, corresponding to an increase of 211%, distributed across 272 922 installations. Considering all voltage levels, installed solar PV capacity reaches 2346 MW across 280 503 installations. The medium- and high-voltage segment therefore accounts for only 2.7% of installations but 60.2% of installed capacity, reflecting much larger average installation sizes at medium, high, and very high voltage (186 kW, against 3.4 kW at low voltage). By April 2026, all 278 municipalities had at least one active low-voltage installation, as did 2864 of the 2872 parishes reporting low-voltage consumption in that month (Rebordosa is excluded from this count, since its records end in October 2025, as discussed in Section 4.1). The eight remaining parishes without any installation are all rural, spread across six municipalities in the Barroso, Beira Interior, and Alto Minho regions.[1] Average installed capacity per municipality had increased by 246% over the period.

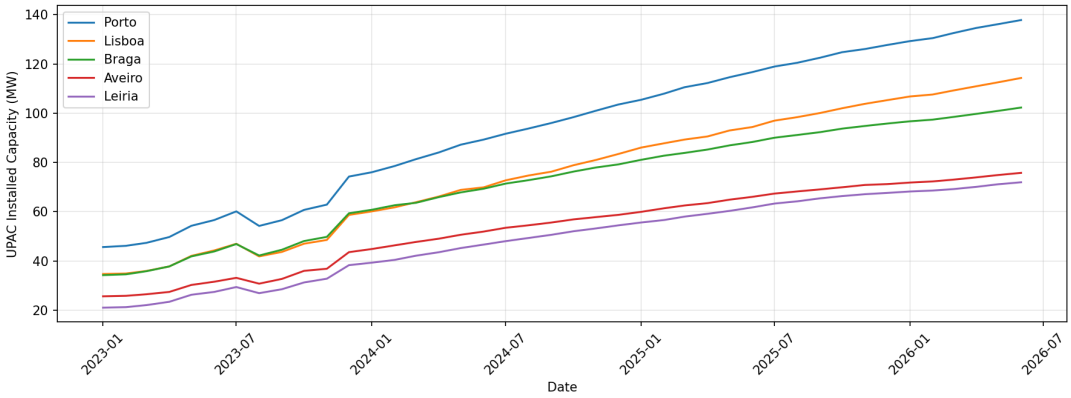


Figure 4.5: Evolution of installed low-voltage UPAC solar PV capacity for the five districts with the highest total installed capacity (January 2023 - June 2026).

[1] Gralhas, Sarraquinhos, and Tourém (Montalegre); Teixeira (Seia); Gavieira (Arcos de Valdevez); Fiães (Melgaço); União das Freguesias de Constantim e Cicouro (Miranda do Douro); and Cortiçada (Aguiar da Beira).

Figure 4.5 shows that Porto leads in total installed capacity (134.6 MW in April 2026), followed by Lisboa (110.9 MW), Braga (99.7 MW), Aveiro (73.9 MW), and Leiria (70.1 MW). At the municipal level, the ten largest low-voltage consumers in April 2026 are listed in Table 4.2. The pattern is dominated by the two metropolitan areas: five of the ten municipalities belong to the district of Lisboa and three to the district of Porto, with Braga and Faro completing the list.

Table 4.2: Ten low-voltage consumers with the highest monthly consumption in April 2026 and their UPAC capacity.

DICO	**District**	**Consumption (GWh/month)**	**UPAC (MW)**	**Number of Installations**	**Density (kW/GWh)**
1106	Lisboa	134.08	9.36	1 549	69.8
1312	Porto	63.67	7.81	1 852	122.6
1111	Lisboa	56.46	19.77	5 158	350.1
1317	Porto	51.26	17.76	4 985	346.5
1105	Lisboa	42.63	16.74	5 176	392.6
0303	Braga	32.47	19.13	5 719	589.2
1308	Porto	32.07	8.68	2 295	270.5
1110	Lisboa	30.10	8.90	2 047	295.8
0808	Faro	29.96	14.91	2 765	497.8
1107	Lisboa	29.78	9.12	2 517	306.3

Within this table, the density column, defined as the ratio of installed UPAC capacity to monthly consumption, reveals substantial heterogeneity in self-consumption penetration among municipalities with comparable consumption levels. Lisboa (DICO 1106) records the highest absolute consumption, 134.08 GWh/month, but only 69.8 kW/GWh, whereas Braga (DICO 0303) records consumption of 32.47 GWh/month, roughly one quarter of Lisboa, and has more than twice the installed UPAC capacity in absolute terms, 19.13 MW versus 9.36 MW, yielding a density of 589.2 kW/GWh, more than eight times higher.

When installed capacity is normalized by monthly consumption across the whole country, a broader spatial pattern emerges. Beja records the highest district-level density, at 841.8 kW/GWh, followed by Leiria (788.6 kW/GWh) and Santarém (785.4 kW/GWh), whereas Lisboa (272.3 kW/GWh) and Porto (424.2 kW/GWh) rank lowest among the 18 districts, with full district-level figures provided in Appendix A. Although these two metropolitan districts have the largest absolute installed capacity, their much larger consumption base means that even substantial capacity growth translates into comparatively low density. The contrast is even sharper at the municipal level: the highest densities occur in small interior municipalities, Vouzela (DICO 1824, Viseu district, 1542.8 kW/GWh) and Aguiar da Beira (DICO 0901, Guarda district, 1455.0 kW/GWh), each with monthly consumption below 2 GWh. This pattern suggests that self-consumption penetration is comparatively higher where absolute consumption is lower, though the available data do

not allow the specific drivers, such as solar resource, land or roof availability, or local economic activity, to be disentangled.

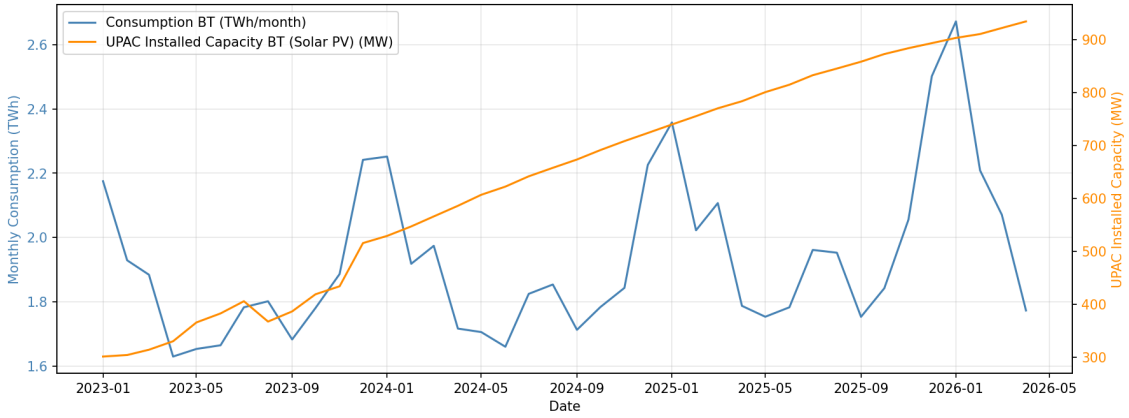


Figure 4.6: National low-voltage monthly consumption and low-voltage installed PV UPAC capacity, January 2023 - April 2026.

Figure 4.6 presents the evolution of low-voltage consumption and installed UPAC capacity at the national level from January 2023. Consumption exhibits a pronounced seasonal cycle without a sustained decline, whereas installed capacity increases steadily over time, reflecting the cumulative nature of new installations.

It should be noted that billed consumption reflects the net balance between energy drawn from the grid and energy produced locally by self-consumption installations, consistent with the standard metering and billing mechanism applied to all self-consumption installations in mainland Portugal.

The trend component identified by the MSTL decomposition in Section 4.2 (Figure 4.2), computed over the training period from November 2020 to April 2025, indicates that this net consumption trend increased rather than decreased over that interval, despite installed self-consumption capacity more than tripling between January 2023 and April 2026. This pattern is consistent with the electrification of end uses offsetting, or even outweighing, the demand-reducing effect of distributed self-consumption. Because billed consumption reflects net electricity demand at the grid connection point, after accounting for locally self-consumed generation, the two effects cannot be separated with the available data. An increase in net demand is therefore compatible with either a relatively modest reduction in grid purchases due to self-consumption being outweighed by electrification, or a substantial reduction being outweighed by even faster growth in gross electricity demand from other sources. The explanation offered here should therefore be interpreted as plausible rather than demonstrated. This limitation is particularly relevant when comparing areas with different levels of PV adoption, since their billed-consumption series may reflect different combinations of gross demand growth and locally self-consumed PV generation.

4.3 Implementation

The framework described in Chapter 3 was implemented in Python using the Nixtla ecosystem: the `statsforecast` library [29] for the base forecasting models and the MSTL decomposition, the `hierarchicalforecast` library [31] for the reconciliation methods, and the `utilsforecast` library for the evaluation metrics. The pipeline was executed three times, one for each of the low voltage, medium, high, and very high voltage, and all voltage analyses, producing an independent set of results for each. The correlation analysis presented in Section 4.7.3 was computed separately, using the `SciPy` library [39].

4.3.1 Pipeline

The pipeline consists of six sequential stages, illustrated in Figure 4.7 and described below.

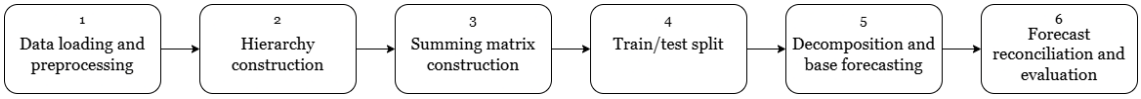


Figure 4.7: Overview of the forecasting pipeline implemented.

The raw consumption dataset is first loaded and preprocessed. Records aggregated under the *OUTROS* category are removed, as discussed in Section 4.1, since they cannot be assigned to individual parishes and would otherwise disrupt the hierarchical structure. For the low-voltage analysis, records at medium, high, and very high voltage are also filtered out at this stage. For the medium, high, and very high voltage analysis, low-voltage records are filtered out instead. The remaining observations are aggregated by parish and month, and each parish is assigned its six-digit DICOFRE code as a unique identifier.

The parish-level series are then aggregated upward to construct the municipality, district, and national series. Since the DICOFRE code is hierarchical by construction, the municipality identifier is obtained by truncating the parish code to its first four digits and the district identifier by truncating it to the first two, so that each level is constructed through a simple group-and-sum operation over the relevant prefix.

Once the hierarchy is defined, the summing matrix S is built from the parish codes following the definition in Section 3.3. This matrix has 3170 rows and 2873 columns for the low-voltage analysis, 1825 rows and 1531 columns for the medium, high, and very high voltage analysis, and 3179 rows and 2882 columns for the all-voltage analysis. Each row identifies the bottom-level parishes that contribute to a given series: the row corresponding to Portugal contains ones in every column; each district row contains ones in the columns of the parishes belonging to that district; each municipality row is populated analogously; and the lower block of the matrix is the identity matrix, so that each parish series coincides with itself. A hierarchy tag dictionary, which associates each level of the hierarchy with the corresponding list of identifiers, is built together with S and passed to the reconciliation routines.

A fixed cutoff of April 2025 is then used to split every series into a training set (November 2020 - April 2025, 54 months) and a test set (May 2025 - April 2026, 12 months). Using a fixed cutoff, rather than a fixed number of trailing observations, ensures that all series in the hierarchy are aligned in time and that the held-out period is identical across parishes, municipalities, districts, and the national aggregate. For the medium, high, and very high voltage analysis, parishes without a complete 54-month training history are further excluded, as discussed in Section 4.1.

The exploratory decomposition and base forecasting models are fitted on the training set. The MSTL decomposition is applied only to the national series, since its purpose in this dissertation is to characterize the trend and seasonal structure of the aggregate rather than to forecast, and its outputs are saved to disk. Separately, `AutoARIMA` is fitted and `Naive` and `Seasonal Naive` forecasts are computed for every series in the hierarchy. Twelve-month-ahead forecasts are then generated for all three models.

Finally, the base forecasts are passed to the reconciliation routines. The three reconciliation methods discussed in Section 3.3, namely bottom-up, top-down using forecast proportions, and middle-out pivoting at the municipality level and also using forecast proportions, are applied independently. The outcome is a set of reconciled forecasts for every combination of base model and reconciliation method. The reconciled forecasts are then compared with the observed values in the test set using the metrics defined in Section 3.4, and the results are reported by hierarchical level in Sections 4.4, 4.5 and 4.6.

4.3.2 Model Configuration

`AutoARIMA` was configured with a seasonal period of 12 months and maximum orders of seasonal and non-seasonal differencing set to one ($d_{\max} = 1$, $D_{\max} = 1$), so that the search space included seasonal ARIMA specifications while avoiding over-differencing on short series. The stepwise selection procedure of [28] was used instead of an exhaustive grid search in order to keep the computational cost manageable across up to 3179 series in the hierarchy. The approximation option of `statsforecast`, which by default accelerates model selection by evaluating candidate models on a subset of the data, was disabled so that every observation in the training series was used when comparing candidate models. The `Naive` benchmark, which projects the last observed value forward without seasonal adjustment, was fitted with its default configuration.

The MSTL decomposition was configured with the same annual period ($m = 12$). The trend forecaster inside MSTL was set to `AutoARIMA` with a seasonal period of one, as required by the `statsforecast` implementation of MSTL, since the seasonal component is handled by the MSTL algorithm itself rather than by the trend model. The reconciliation methods were configured as follows: bottom-up with its default settings; top-down using forecast proportions as the disaggregation rule; and middle-out pivoting at the municipality level, also using forecast proportions below the pivot. The forecast horizon was set to $h = 12$ months to match the length of the test set.

4.4 Low Voltage Results

This section presents the forecasting results for the low-voltage analysis. Section 4.4.1 reports the accuracy metrics by hierarchical level, and Section 4.4.2 illustrates the reconciled forecasts at the national level. The corresponding metrics for the all-voltage analysis are reported in Section 4.6, and a comparison between the two analyses is provided in Section 4.7.

4.4.1 Accuracy Metrics

Table 4.3 reports MAE, RMSE, and MASE for `AutoARIMA`, `Naive`, and `Seasonal Naive`, together with the results of the three reconciliation methods for `AutoARIMA`, aggregated by hierarchical level. Because MAE and RMSE are scale-dependent, they are reported in level-specific units, namely TWh for Portugal, GWh for districts and municipalities, and MWh for parishes, and should be compared only within each level. By contrast, MASE is scale-independent and can be compared across levels.

Table 4.3: Forecast accuracy for the low-voltage analysis by hierarchical level.[2]

	AutoARIMA			**Naive**	**Seasonal Naive**
Level / Metric	Bottom-Up	Top-Down	Middle-Out		
Portugal (TWh)					
MAE	0.092	0.095	0.093	0.255	0.129
RMSE	0.126	0.132	0.127	0.375	0.160
MASE	1.386	1.431	1.397	3.825	1.936
Districts (GWh)					
MAE	5.24	5.43	5.28	14.87	7.22
RMSE	7.19	7.51	7.25	21.57	9.07
MASE	1.418	1.472	1.430	4.021	1.857
Municipalities (GWh)					
MAE	0.356	0.374	0.364	0.995	0.478
RMSE	0.483	0.505	0.490	1.420	0.599
MASE	1.448	1.529	1.488	3.925	1.802
Parishes (MWh)					
MAE	38.15	39.82	38.85	99.69	48.05
RMSE	50.19	52.29	50.89	140.53	60.01
MASE	1.430	1.484	1.455	3.211	1.609

[2]Naive forecasts are constant across reconciliation methods by construction. Seasonal Naive forecasts produce numerically identical results across reconciliation methods up to floating-point precision. Both are therefore reported as a single column.

Two main observations emerge from the table. First, `AutoARIMA` outperforms the `Naive` benchmark by a substantial margin at every hierarchical level and under every reconciliation method. Across the hierarchy, the MASE of `AutoARIMA` ranges between 1.39 and 1.53, whereas the corresponding values for `Naive` range between 3.21 and 4.02. In practical terms, the error of the `Naive` benchmark is between two and three times larger than that of `AutoARIMA`, which suggests that the additional complexity of automatically selecting an ARIMA specification is justified for these series.

Second, among the three reconciliation methods, bottom-up consistently yields the lowest errors across all four hierarchical levels and all three metrics, followed by middle-out and then top-down. The differences are small at the national level, where bottom-up attains a MASE of 1.386, compared with 1.397 for middle-out and 1.431 for top-down, and remain modest throughout the hierarchy. At the parish level, for example, the MAE of bottom-up is 1.8% lower than that of middle-out and 4.2% lower than that of top-down. Since the `Naive` base forecasts are identical across reconciliation methods at every level, the corresponding metrics coincide; consequently, the choice of reconciliation method affects only the `AutoARIMA` results.

It is also noteworthy that all AutoARIMA MASE values exceed one. This does not indicate that AutoARIMA underperforms the Seasonal Naive model over the test period, since MASE is normalized using the in-sample error of the training-period Seasonal Naive benchmark. Table 4.3 also reports the performance of a Seasonal Naive forecast over the test period, obtained by repeating the value observed twelve months earlier. AutoARIMA outperforms this forecast at every hierarchical level, with MASE values ranging from 1.386 to 1.529, compared with 1.609 to 1.936 for Seasonal Naive. The relatively high MASE values are partly attributable to the atypical conditions during the test period. National low-voltage consumption reached 2.672 TWh in January 2026, compared with 2.358 TWh in January 2025, corresponding to a 13.3% increase, and exceeded every monthly value observed during the training period. This issue is discussed further in Section 4.7.

4.4.2 National Forecast Plots

Figure 4.8 compares the three reconciled `AutoARIMA` forecasts at the national level with the observed test values. The three curves remain similar throughout the twelve test months: the difference between the highest and lowest forecast never exceeds 0.67% of the observed consumption in any month, with the largest spread occurring in March 2026. This is consistent with the small differences reported in Table 4.3 at the Portugal level. All three methods reproduce the seasonal pattern of low-voltage consumption, but systematically underestimate consumption in the second half of the test period, with the largest error occurring in January 2026, where the forecast error ranges from 10.9% to 11.3% depending on the reconciliation method. This is also the only month in the test set that exceeds the maximum value observed during training.

For reference, Figure 4.9 presents the same national series alongside the Naive and Seasonal Naive baselines. Because Naive projects the last observed value forward, the forecast is flat at 1.79 TWh and does not capture the seasonal cycle. Seasonal Naive, which repeats the value observed

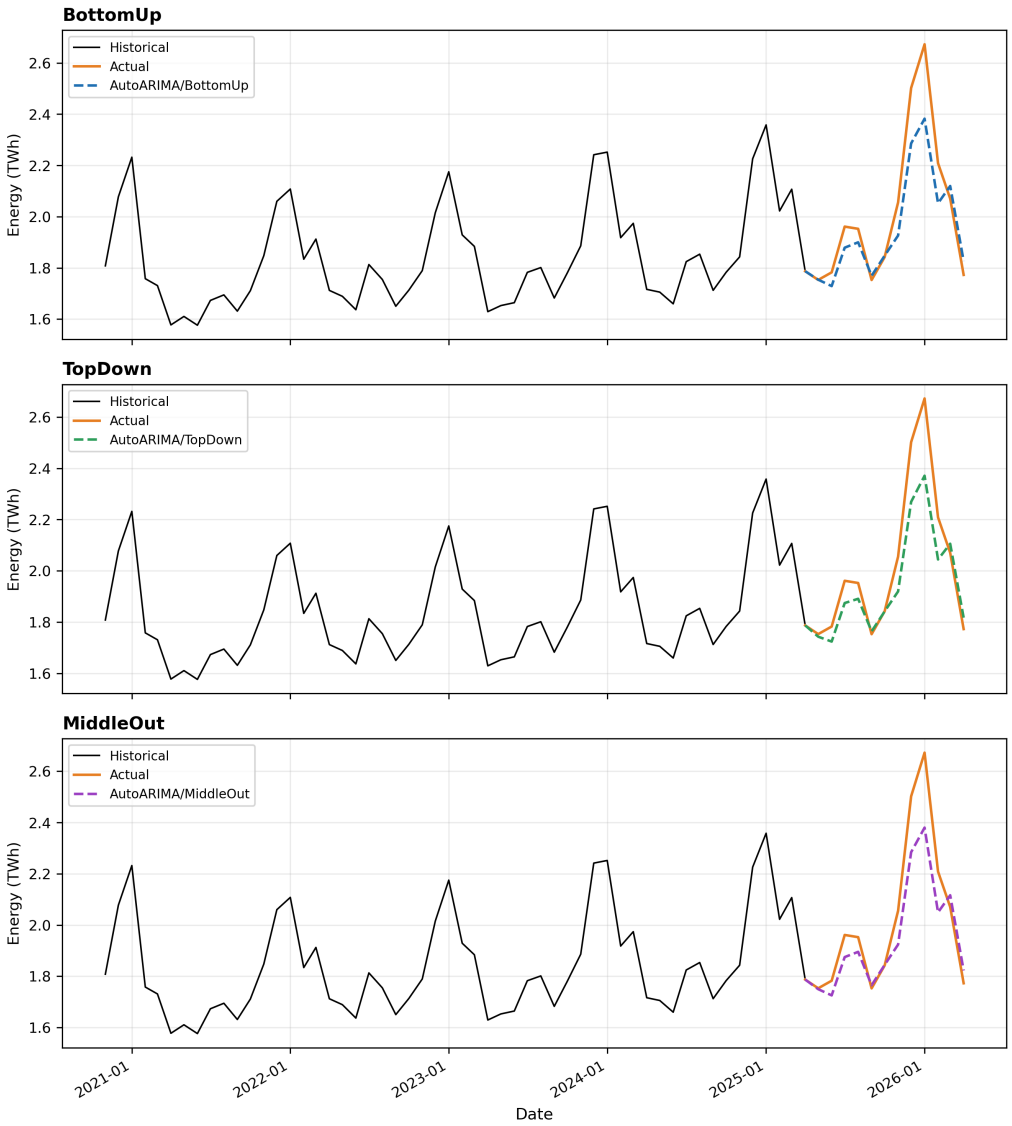


Figure 4.8: National low-voltage total: reconciled `AutoARIMA` forecasts versus observed values, May 2025 - April 2026.

twelve months earlier, tracks the seasonal pattern closely but still underestimates the January 2026 peak, consistent with the MASE values reported in Table 4.3.

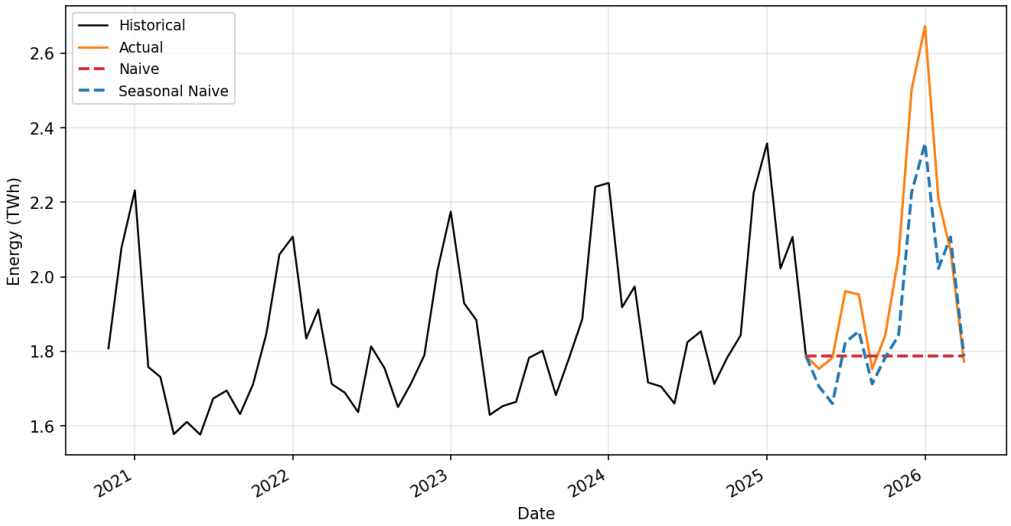


Figure 4.9: National low-voltage total: Naive and Seasonal Naive forecasts versus observed values, May 2025 - April 2026.

4.5 Medium, High, and Very High Voltage Results

This section presents the forecasting results for the medium, high, and very high voltage. Because this segment could be forecast only for a filtered subset of parishes (Section 4.1), the comparison with the low-voltage results of Section 4.4 should be regarded as indicative rather than as a controlled test, and is interpreted accordingly throughout.

4.5.1 Accuracy Metrics

Table 4.4 reports MAE, RMSE, and MASE for the medium, high, and very high voltage analysis.

Two observations distinguish this analysis from the low voltage results presented in Section 4.4. First, `AutoARIMA` is markedly more accurate for this filtered medium, high, and very high voltage subset. MASE ranges from 0.974 to 1.251 across the hierarchy, compared with 1.386 to 1.529 for low voltage. At the district level, the bottom-up AutoARIMA forecast achieves a MASE of 0.974, below one. When evaluated directly over the test period, however, Seasonal Naive performs better at the district and national levels, achieving MASE values of 0.903 and 0.951, respectively, compared with 0.974 and 1.041 for AutoARIMA. AutoARIMA performs better at the municipality and parish levels, with MASE values of 1.056 and 1.117, respectively, compared with 1.066 and 1.142 for Seasonal Naive.

Second, bottom-up is again the best-performing reconciliation method in most cases, namely in nine of the twelve level-metric combinations. This represents a less complete dominance than the twelve-of-twelve result for low voltage (Section 4.4), but reflects the same qualitative pattern, with bottom-up outperforming the other two methods overall. At the parish level, the bottom-up MASE (1.117) is 10.7% lower than top-down (1.251), which is the largest such margin against top-down observed at any level in this analysis, and 9.0% lower than middle-out (1.227). At

Table 4.4: Forecast accuracy for the medium, high and very high voltage analysis by hierarchical level.

	AutoARIMA			**Naive**	**Seasonal Naive**
Level / Metric	Bottom-Up	Top-Down	Middle-Out		
Portugal (TWh)					
MAE	0.037	0.038	0.043	0.162	0.034
RMSE	0.043	0.045	0.049	0.185	0.041
MASE	1.041	1.071	1.212	4.542	0.951
Districts (GWh)					
MAE	5.059	4.917	5.071	10.869	4.837
RMSE	5.832	5.708	5.866	12.742	5.633
MASE	0.974	1.029	1.022	2.560	0.903
Municipalities (GWh)					
MAE	0.544	0.556	0.555	0.959	0.522
RMSE	0.637	0.656	0.648	1.147	0.636
MASE	1.056	1.100	1.053	2.073	1.066
Parishes (MWh)					
MAE	139.24	148.01	146.73	209.05	135.00
RMSE	163.00	182.15	180.79	251.99	163.72
MASE	1.117	1.251	1.227	1.823	1.142

the national level, by contrast, the bottom-up margin against middle-out is even larger, at 14.1%, although its margin against top-down there is comparatively small, at 2.8%. At the municipality level, middle-out rather than bottom-up attains the lowest MASE.

The lower MASE values reported in Table 4.4 are partly attributable to the test period itself. The maximum value observed during training is 2.128 TWh, reached in July 2022, which is consistent with the summer peak identified in the MSTL decomposition below. Neither January 2026 (1.985 TWh) nor the peak of the test period, October 2025 (2.114 TWh), exceeds this training maximum. This differs from the low voltage case, where January 2026 exceeds any value observed during training (Section 4.4).

4.5.2 National Forecast Plots

Figure 4.10 compares the three reconciled `AutoARIMA` forecasts against observed test values. The three methods track the observed series closely. In the October 2025 peak month, errors range from -3.6% (top-down) to -4.4% (middle-out), which is substantially smaller than the errors observed at the equivalent peak month for low voltage (10.9% to 11.3%, Section 4.4).

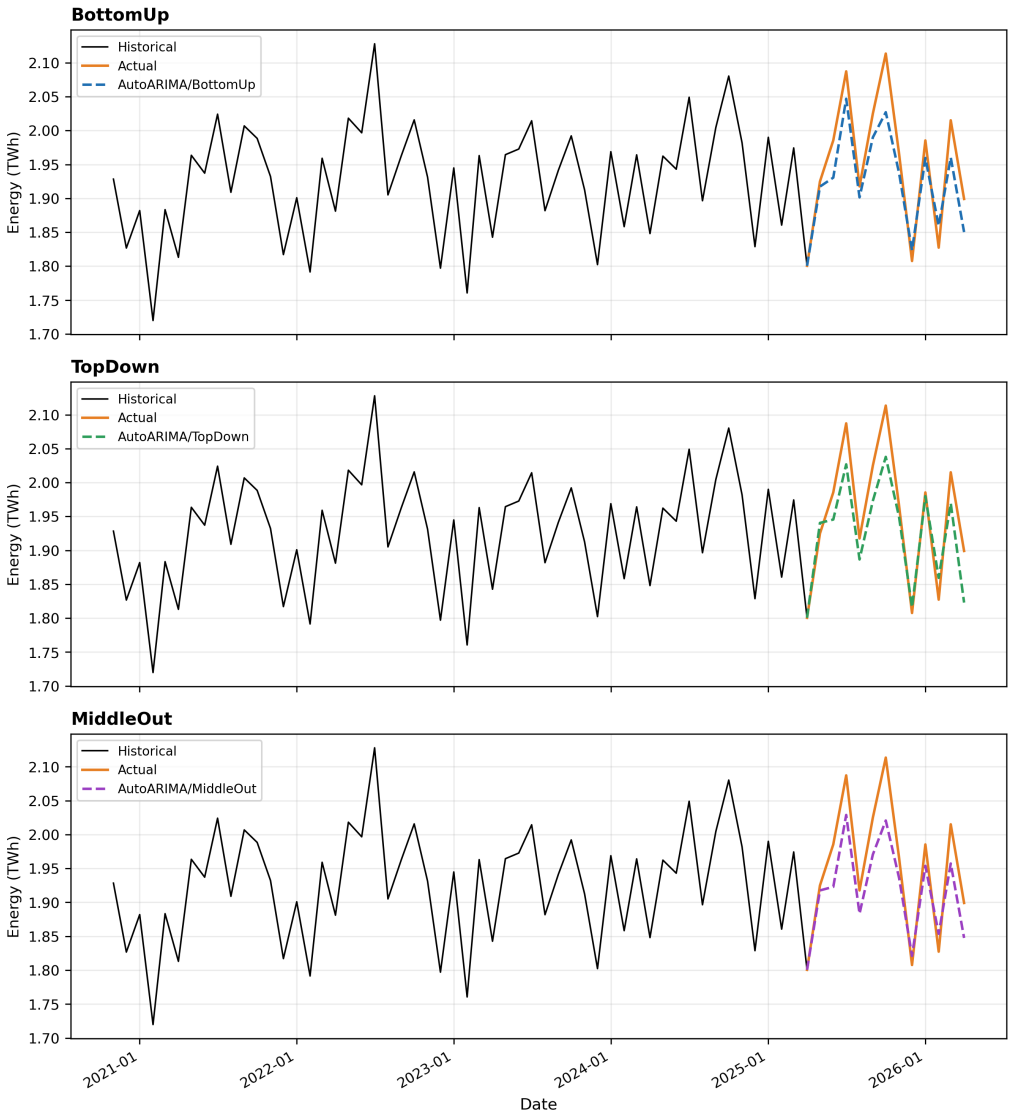


Figure 4.10: National medium, high, and very high voltage total: reconciled `AutoARIMA` forecasts versus observed values, May 2025 - April 2026.

For reference, Figure 4.11 presents the same national series alongside the Naive and Seasonal Naive baselines. As in the low-voltage case, Naive remains flat throughout the test period. Seasonal Naive, by contrast, tracks the observed series closely, and even outperforms AutoARIMA at the national and district levels, as reported in Table 4.4.

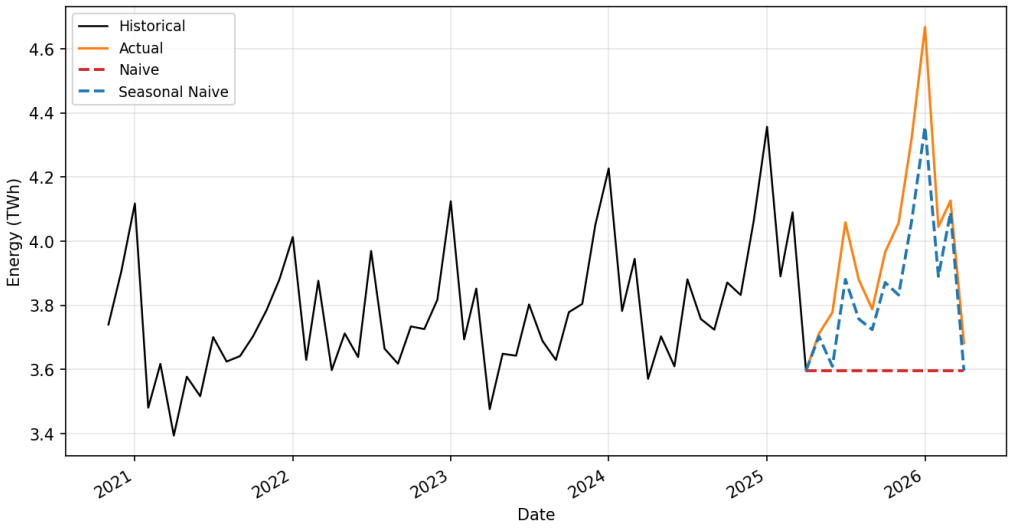


Figure 4.11: National medium, high, and very high voltage total: Naive and Seasonal Naive forecasts versus observed values, May 2025 - April 2026.

4.6 All Voltage Results

This section presents the forecasting results for the all-voltage analysis and follows the same structure as Section 4.4, with accuracy metrics by hierarchical level in Section 4.6.1 and national forecast plots in Section 4.6.2.

4.6.1 Accuracy Metrics

Table 4.5 reports MAE, RMSE, and MASE for the all-voltage analysis.

As in the low-voltage analysis, `AutoARIMA` outperforms the `Naive` benchmark at every level and under every reconciliation method, with MASE ranging from 1.12 to 1.47 compared with 2.85 to 4.63 for `Naive`. The gap between the two models is wider than in the low-voltage case, particularly at the national level, where the `Naive` MASE of 4.63 is about four times that of `AutoARIMA` (1.12 to 1.19).

Unlike the low-voltage case, no single reconciliation method dominates across the hierarchy. Top-down is the best performer at the national and district levels, where it achieves the lowest MAE, RMSE, and MASE. As the hierarchy becomes more disaggregated, this advantage diminishes and eventually reverses. At the municipality level, top-down still yields the lowest MAE and RMSE, but bottom-up already has the lowest MASE, with the two methods being nearly tied (1.263 versus 1.264). At the parish level, the reversal is clear: bottom-up is the best method by a substantial margin, with a MASE of 1.365, which is 6.0% lower than top-down (1.451) and 7.2% lower than middle-out (1.471). Since parishes account for more than 90% of the 3179 series in the hierarchy, a single error measure computed over all series together, rather than separately by level as in Table 4.5, would be dominated by parish-level performance. Accordingly, the reversal

Table 4.5: Forecast accuracy for the all-voltage analysis by hierarchical level.

	AutoARIMA			Naive	Seasonal Naive
Level / Metric	Bottom-Up	Top-Down	Middle-Out		
Portugal (TWh)					
MAE	0.101	0.099	0.106	0.411	0.142
RMSE	0.136	0.126	0.140	0.491	0.167
MASE	1.144	1.121	1.192	4.630	1.604
Districts (GWh)					
MAE	8.41	7.76	8.27	24.01	10.09
RMSE	10.59	9.81	10.70	30.13	12.12
MASE	1.208	1.134	1.179	3.497	1.413
Municipalities (GWh)					
MAE	0.787	0.778	0.797	1.798	0.892
RMSE	0.988	0.962	0.997	2.286	1.114
MASE	1.263	1.264	1.281	3.244	1.474
Parishes (MWh)					
MAE	102.8	106.0	106.8	198.0	111.31
RMSE	125.9	132.4	134.6	253.9	137.16
MASE	1.365	1.451	1.471	2.850	1.494

observed at the parish level would also determine the ranking of the three reconciliation methods under such a global measure. Overall, these results suggest that the preferred reconciliation method depends on the level of the hierarchy being forecast and, near the point of reversal, may also depend on the accuracy metric used. This contrasts with the low-voltage analysis, where bottom-up was uniformly superior at every level and under every metric.

As in Section 4.4.1, all AutoARIMA MASE values exceed one. As reported in Table 4.5, AutoARIMA again outperforms Seasonal Naive evaluated as a competing model at every level, ranging from 1.121 to 1.471 against 1.413 to 1.604. The test period again includes an atypical month: national all-voltage consumption reached 4.668 TWh in January 2026, compared with 4.357 TWh in January 2025, a 7.1% increase, and was higher than in any month observed in the training set.

4.6.2 National Forecast Plots

Figure 4.12 compares the three reconciled `AutoARIMA` forecasts at the national level with the observed test values.The three curves remain relatively close, although to a lesser extent than in the low-voltage case: across the twelve test months, the difference between the highest and lowest forecast never exceeds 1.15% of the observed consumption in any month, with the largest spread occurring in May 2025, compared with 0.67% for low voltage. In January 2026, all three methods underestimate the observed peak, with errors ranging from -6.2% for top-down to -7.2% for middle-out.

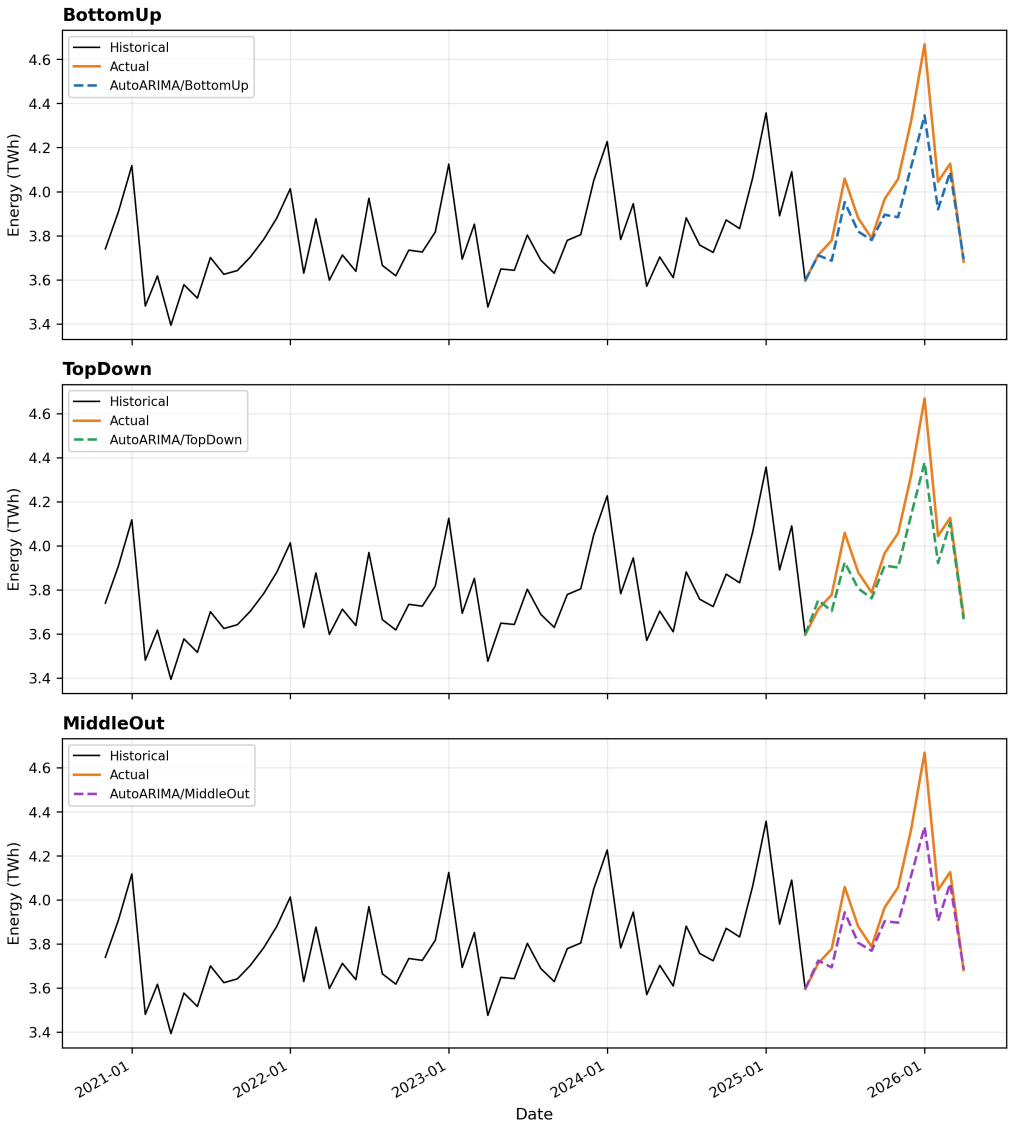


Figure 4.12: National all-voltage total: reconciled `AutoARIMA` forecasts versus observed values, May 2025 - April 2026.

Figure 4.13 presents the corresponding comparison for the Naive and Seasonal Naive baselines. As in the two previous analyses, Naive yields a flat forecast, while Seasonal Naive tracks the seasonal pattern but underestimates the January 2026 peak, consistent with the MASE values reported in Table 4.5.

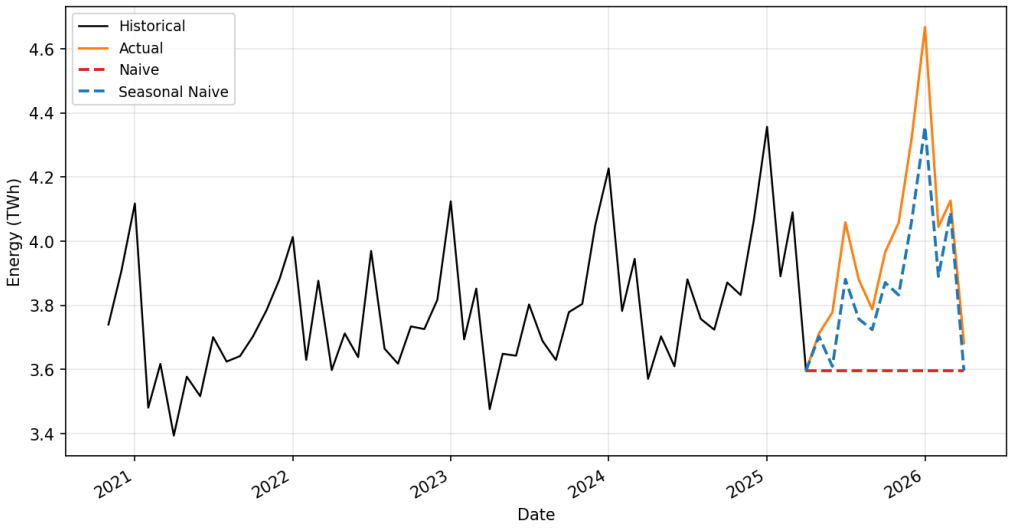


Figure 4.13: National all-voltage total: Naive and Seasonal Naive forecasts versus observed values, May 2025 - April 2026.

4.7 Discussion

This section integrates the results of the low-voltage, medium, high, and very high voltage, and all-voltage analyses presented in Sections 4.4, 4.5, and 4.6 and discusses their implications.

4.7.1 Cross-Voltage Comparison

The three analyses differ markedly in terms of relative forecastability. AutoARIMA MASE ranges from 1.386 to 1.529 for low voltage, from 0.974 to 1.251 for the filtered medium, high, and very high voltage subset examined in Section 4.5, and from 1.121 to 1.471 for all voltage. The medium, high, and very high voltage range is the lowest of the three, and the only one to include a value below one. Since MASE is scale-independent by construction (Section 3.4), this ordering reflects a genuine difference in relative forecastability rather than an artefact of scale. For medium, high, and very high voltage, this conclusion must however be read together with the sample caveat noted in Section 4.1: the parishes retained there are specifically those with the most continuous consumption histories, and do not constitute a representative sample of the segment. In line with this pattern, the MSTL remainder scales more slowly than the trend when moving from low voltage, where it represents 2.1% of the trend level, to medium, high, and very high voltage (1.4%, subject to the sample caveat noted above) and to all voltage (also 1.4%), as shown in Section 4.2.

The choice of the best reconciliation method also varies across the three analyses, although not in a way that cleanly separates low voltage from medium, high, and very high voltage. Bottom-up achieves the lowest error in all 12 level-metric combinations for low voltage, and in 9 of 12 for medium, high, and very high voltage, which represents a less complete but still clear dominance. For all voltage, by contrast, bottom-up wins only 4 of 12 combinations, with top-down instead dominating at the national and district levels. This is a non-trivial pattern. Since all voltage is the

combination of the other two segments, and bottom-up dominates in both of them individually, it is not obvious why the combined series should favor top-down at the upper levels of the hierarchy. A plausible explanation is that the differing seasonal profiles of the two segments, low voltage peaking in January and medium, high, and very high voltage peaking in July (Section 4.2), interact differently once aggregated. In this case, forecasting the combined national series directly may capture structure that is lost when the many bottom-level series of the two segments are forecast individually and summed. This explanation is consistent with the observed pattern but was not tested directly. The sensitivity of the best reconciliation method to the specific hierarchy and voltage scope being forecast is, in itself, a relevant finding. It indicates that the reconciliation method should not be fixed *a priori*, but selected empirically, as was done in this dissertation.

When evaluated as a competing model rather than used only as a normalizing benchmark, Seasonal Naive is outperformed by AutoARIMA in ten of the twelve level-analysis combinations examined in this chapter (Tables 4.3, 4.4, and 4.5). The two exceptions occur at the national and district levels in the medium-, high-, and very-high-voltage analysis, where Seasonal Naive achieves a lower MASE than AutoARIMA. As discussed below, this is also the only segment without an anomalous test month. This suggests that AutoARIMA's advantage over simple seasonal repetition is concentrated in series and periods involving genuine deviations from the previous year's pattern, rather than being uniform across all forecasting conditions.

4.7.2 The January 2026 Anomaly and Forecastability

The fact that every MASE value for low voltage and all voltage exceeds one, unlike medium, high, and very high voltage, where the district level falls below one, is directly attributable to a single month. Both the low voltage and all voltage test periods include an unusually large increase in January 2026 relative to January 2025. Low voltage consumption rose from 2.358 TWh to 2.672 TWh, a 13.3% increase, and all-voltage consumption rose from 4.357 TWh to 4.668 TWh, a 7.1% increase. In both cases, the January 2026 values exceed any month observed during training. The medium, high, and very high voltage test period contains no such anomaly. January 2026 (1.985 TWh) remains below the training maximum of 2.128 TWh, reached in July 2022 (Section 4.5). This asymmetry is noteworthy. The same calendar month was anomalous for low voltage demand but not for medium-, high-, and very-high-voltage demand, suggesting that the two segments are driven, at least in part, by different factors. This remains indicative rather than conclusive, given that it is based on a single twelve-month test period (May 2025 to April 2026) and, for medium, high, and very high voltage, on the filtered subset of parishes described in Section 4.1. All reconciled `AutoARIMA` forecasts underestimate the January 2026 peak for low voltage (10.9% to 11.3%) and all voltage (6.2% to 7.2%). The smaller October 2025 peak in the medium-, high-, and very-high-voltage test period is underestimated by a comparable but smaller margin (3.6% to 4.4%, Section 4.5). Because MASE benchmarks against the Seasonal Naive error observed in the training period, a single atypical month in the twelve-month test window is sufficient to push the ratio above one, even when the reconciled forecasts track the seasonal pattern correctly for the remaining months, as Figures 4.8, 4.10, and 4.12 illustrate.

This explanation can be assessed directly by recalculating the MASE after excluding January 2026 from the test period, while keeping the models, reconciliation procedures, and training-period benchmark unchanged. The resulting values are reported in Table 4.6. For the low-voltage analysis, MASE decreases at every hierarchical level, from approximately 10.8% to 19.7% at the national level, although it remains above one throughout. For the medium-, high-, and very-high-voltage analysis, the recalculated values remain close to the original ones at all levels, consistent with the absence of a comparable anomaly in these segments. For the all-voltage analysis, the effect is most pronounced at the national level: after excluding January 2026, all three reconciliation methods produce MASE values below one there, while the reduction at the district, municipality, and parish levels is consistently smaller. These results confirm that the MASE values above one observed throughout the low-voltage and all-voltage analyses are driven substantially, although not exclusively, by a single atypical month, rather than indicating a general inability to outperform the Seasonal Naive benchmark.

Table 4.6: AutoARIMA MASE with and without January 2026 in the test period, by hierarchical level.[3]

		With January 2026			**Without January 2026**		
	Level	Bottom-Up	Top-Down	Middle-Out	Bottom-Up	Top-Down	Middle-Out
LV	Portugal	1.386	1.431	1.397	1.115	1.150	1.125
	Districts	1.418	1.472	1.430	1.177	1.218	1.185
	Municipalities	1.448	1.529	1.488	1.230	1.306	1.274
	Parishes	1.430	1.484	1.455	1.272	1.320	1.297
MHV	Portugal	1.041	1.071	1.212	1.070	1.154	1.242
	Districts	0.974	1.029	1.022	0.955	1.029	1.007
	Municipalities	1.056	1.101	1.053	1.030	1.090	1.032
	Parishes	1.117	1.251	1.227	1.097	1.191	1.159
AV	Portugal	1.144	1.121	1.192	**0.917**	**0.926**	**0.956**
	Districts	1.208	1.134	1.179	1.125	1.086	1.051
	Municipalities	1.263	1.265	1.281	1.125	1.138	1.135
	Parishes	1.365	1.452	1.471	1.245	1.287	1.282

4.7.3 Parish-Level Error Patterns and Hierarchical Propagation

The aggregate parish-level metrics reported in Tables 4.3, 4.4, and 4.5 provide a concise comparison of forecasting models and reconciliation methods. However, they summarize the performance of all parishes within each voltage scope through a single average value and therefore do not describe the dispersion or spatial distribution of forecast errors. This distinction is relevant for distribution-network planning, where the geographical location of forecast errors may matter in addition to their average magnitude.

[3] LV: Low Voltage; MHV: Medium, High, and Very High Voltage; AV: All Voltage.

To examine this aspect, parish-level MASE values obtained with AutoARIMA and Bottom-Up reconciliation were retained separately for the low-voltage, medium-, high- and very-high-voltage, and all-voltage analyses. MASE was selected as the primary spatial error metric because it is scale-independent and therefore allows comparisons between parishes with different consumption levels.

Figure 4.14 presents the spatial distribution of parish-level MASE across mainland Portugal, with enlarged, individually-scaled versions of each map provided in Appendix C. The maps reveal substantial spatial heterogeneity in forecast accuracy that is not visible in the average values reported in Tables 4.3, 4.4, and 4.5. In the low-voltage and all-voltage analyses, MASE values vary across the territory, with most parishes showing intermediate values and a smaller number of parishes displaying substantially higher errors. The maps do not suggest a single uniform geographical gradient. Instead, the observed variation is consistent with local differences in the temporal structure of billed consumption. These differences may arise from consumption scale, volatility, seasonal behavior, changes in the active customer base, and other local factors not explicitly modelled in the forecasting framework.

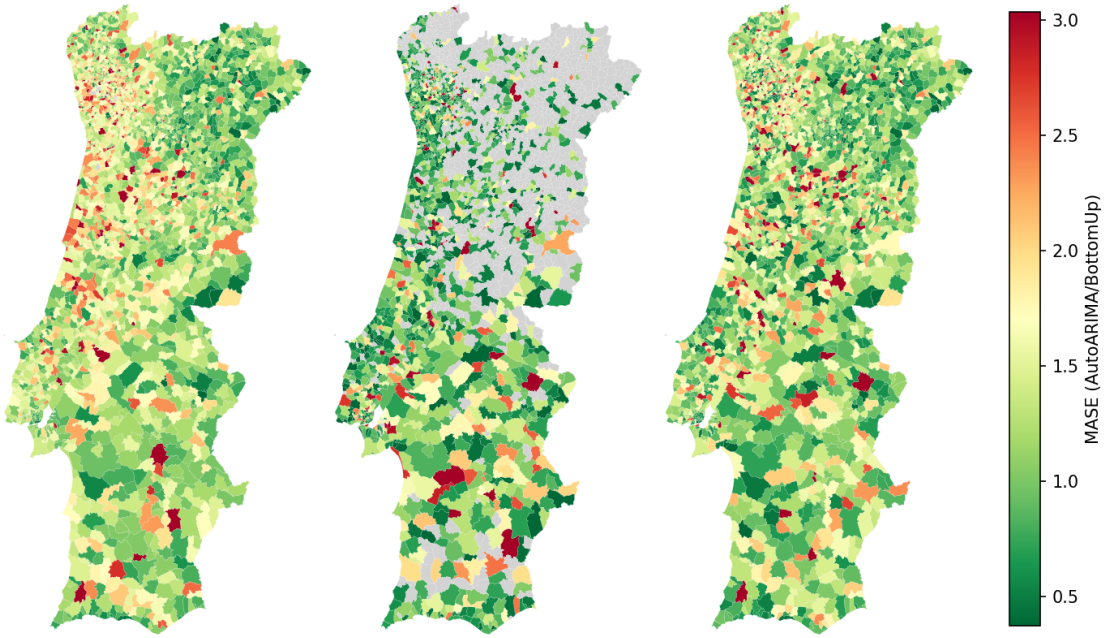


Figure 4.14: Parish-level MASE under AutoARIMA and Bottom-Up reconciliation for low voltage (left), medium, high, very-high voltage (center), and all voltage levels (right).

The medium-, high-, and very-high-voltage map requires a more cautious interpretation. As described in Section 4.1, this hierarchy contains 1531 parishes with complete training histories. However, one of these parishes has no observations during the test period and is therefore excluded from the spatial error map, leaving 1530 parishes with computable MASE values. By contrast, the low-voltage and all-voltage analyses cover 2873 and 2882 parishes, respectively.

The grey areas visible in the medium-, high-, and very-high-voltage panel of Figure 4.14 represent parishes excluded from that analysis, rather than areas with zero forecasting error. The map should therefore be interpreted with this restriction in mind: it shows forecast accuracy for the 1530 parishes retained in the medium-, high-, and very-high-voltage analysis, rather than for the entire population of parishes with connections at these voltage levels.

To explore whether the spatial heterogeneity of forecast error is associated with self-consumption adoption, Figure 4.15 compares parish-level MASE with local installed UPAC-capacity density for the three voltage scopes. UPAC density is calculated as installed solar PV capacity divided by billed consumption within the same parish and voltage scope, and is expressed in kW/GWh. The horizontal axis is shown on a logarithmic scale because UPAC density spans several orders of magnitude across parishes.

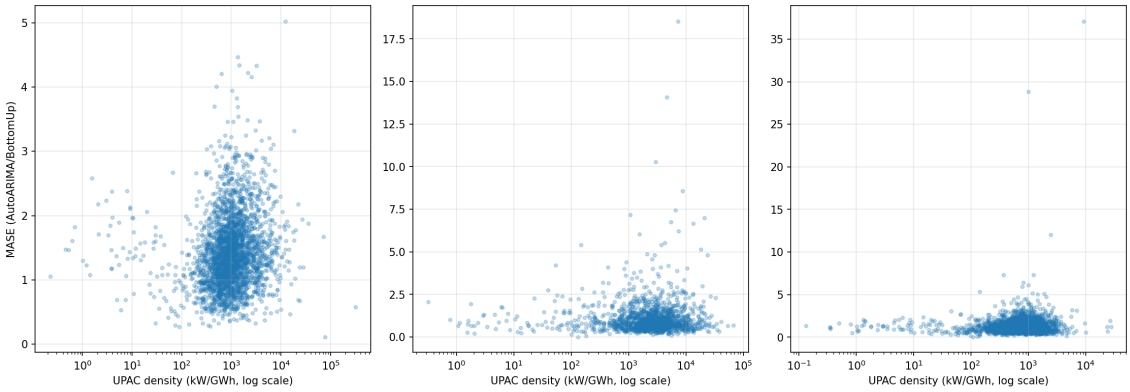


Figure 4.15: Parish-level MASE versus UPAC installed capacity density for low voltage (left), medium, high, very-high voltage (center), and all voltage levels (right).

Table 4.7: Correlation between parish-level MASE and UPAC installed-capacity density, for low voltage (LV), medium, high, and very high voltage (MHV), and all voltage (AV).

	Pearson			**Spearman**		
	r	95% CI	R^2	ρ	95% CI	n
LV	0.001	(–0.036, 0.038)	0.000	0.194^{***}	(0.158, 0.229)	2800
MHV	0.039	(–0.013, 0.089)	0.001	0.009	(–0.042, 0.060)	1473
AV	0.096^{***}	(0.059, 0.132)	0.009	0.070^{***}	(0.033, 0.106)	2807

$^{***}p < 0.001$

Table 4.7 quantifies the association between parish-level MASE and UPAC capacity density using Pearson's and Spearman's coefficients. The results indicate that the association is negligible in most cases. Following the conventional thresholds introduced in Section 3.5, Pearson's r falls within the negligible range for all three voltage scopes, while the corresponding coefficient of determination does not exceed $R^2 = 0.009$. Thus, the linear association with UPAC density accounts for less than 1% of the sample variation in parish-level MASE, including in cases where

the correlation is statistically significant. Spearman's ρ is also negligible for the medium-, high-, and very-high-voltage and all-voltage analyses, and falls within the weak range for low voltage ($\rho = 0.194$).

The statistically significant associations, namely low-voltage Spearman and all-voltage Pearson and Spearman correlations, should therefore be interpreted in light of their small effect sizes and the large sample sizes. Even the upper limits of their 95% confidence intervals remain within the negligible-to-weak range.

The results do not provide evidence that higher behind-the-meter PV density is systematically associated with poorer forecast performance.

This comparison should be interpreted as exploratory rather than causal. The billed-consumption series represent net electricity drawn from the grid and may reflect different combinations of gross demand growth and locally self-consumed PV generation, as discussed in Section 4.2.2. Forecast error may also be influenced by parish size, consumption scale, volatility, annual seasonal structure, missing observations, and the mix of residential, commercial and industrial consumers. Consequently, the scatter plots do not isolate the effect of PV adoption on forecast performance. They assess whether areas with greater installed PV capacity relative to billed consumption are systematically associated with different forecasting errors. Within the available data and the selected test period, UPAC density alone does not provide a sufficient explanation for the observed variation in parish-level MASE.

Beyond their association with self-consumption density, the parish-level error patterns examined above also have implications for the propagation of local forecast errors under Bottom-Up reconciliation. Under this approach, forecasts are generated independently for each parish and then aggregated to the municipality, district, and national levels. Bottom-Up therefore preserves local temporal dynamics, but it also propagates parish-level forecast errors to every higher level of the hierarchy.

The spatial maps indicate that the bottom level is heterogeneous. Some parishes exhibit relatively stable consumption with clear seasonal structure, whereas others show greater volatility, weaker seasonal repetition, or isolated deviations during the test period.

This distinction is important for interpreting the performance of Bottom-Up reconciliation. In parishes with stable annual patterns, bottom-level forecasts can capture local dynamics effectively, and their aggregation can improve coherence without substantially increasing forecast error. In more volatile or structurally changing parishes, however, the base forecasts may fail to anticipate abrupt changes, and these errors propagate uncorrected to the parent municipality, district, and national aggregate.

The effect of a parish-level error on aggregate accuracy also depends on the consumption scale of the parish. Unlike MASE, which is normalized using a scale-specific benchmark for each parish, MAE and RMSE at the municipality, district and national levels are expressed in absolute energy units and are therefore more strongly influenced by errors in larger-consumption parishes. A small parish may exhibit a high MASE because its forecast error is large relative to its own seasonal benchmark, while its absolute error has a limited effect on aggregate MAE or RMSE.

However, the contribution of each parish also depends on the sign and timing of its errors, since errors from different parishes may partly offset one another during aggregation. For example, in the low-voltage analysis, Vila Nova da Rainha (Azambuja) has the highest parish-level MASE among the 2,873 parishes examined, with a value of 5.02, yet its average monthly consumption is approximately 266 MWh. By contrast, the central Porto parish union [4] has the highest average monthly consumption among the same parishes, approximately 18.86 GWh, roughly 70.9 times that of Vila Nova da Rainha, yet a MASE of only 0.57. This example illustrates that parish-level MASE and aggregate-level MAE or RMSE capture different aspects of Bottom-Up performance.

This behavior highlights the main trade-off associated with Bottom-Up reconciliation. The method preserves local heterogeneity and avoids imposing a common allocation pattern across parishes, but it remains sensitive to poorly predictable bottom-level series. The resulting higher-level forecasts depend not only on the average accuracy of the parish models, but also on the location, volatility, and absolute consumption scale of the errors that are propagated through the hierarchy.

4.7.4 Implications for Self-Consumption and Grid Planning

The results in this chapter should be interpreted in conjunction with the self-consumption growth documented in Section 4.2 and the parish-level patterns examined in Section 4.7.3. As discussed in Section 4.2.2, installed low-voltage solar PV capacity more than tripled between January 2023 and April 2026. However, the underlying low-voltage consumption trend also continued to grow over the same period, and the available data do not allow this association to be disentangled into its underlying drivers.

For distribution network planning, this has two practical implications. First, the hierarchical forecasts produced in this dissertation, disaggregated to the level of individual parishes and municipalities, provide a spatially explicit basis for identifying where load growth and self-consumption penetration occur concurrently, even though Section 4.7.3 found no systematic association between local PV density and forecast accuracy. Second, the sensitivity of the best-performing reconciliation method to the analytical setting indicates that forecasting pipelines intended for operational use should not adopt a single reconciliation method without validating it for the specific hierarchy and voltage scope being forecast.

4.7.5 Forecasting Beyond the Evaluation Period

AutoARIMA was refitted to the full available series, from November 2020 to April 2026, extending the framework beyond the evaluation period used in the preceding sections. Forecasts were generated for the following twelve months, from May 2026 to April 2027, and reconciled using the Bottom-Up approach. This combination achieved the lowest parish-level MASE in each of the three voltage analyses, as reported in Tables 4.3, 4.4, and 4.5.

[4] Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau, and Vitória

Figure 4.16 presents the resulting national forecasts. The low-voltage projection retains a pronounced seasonal pattern, whereas the projection for medium, high, and very high voltage varies within a narrower range. The all-voltage projection reflects the contribution of both segments.

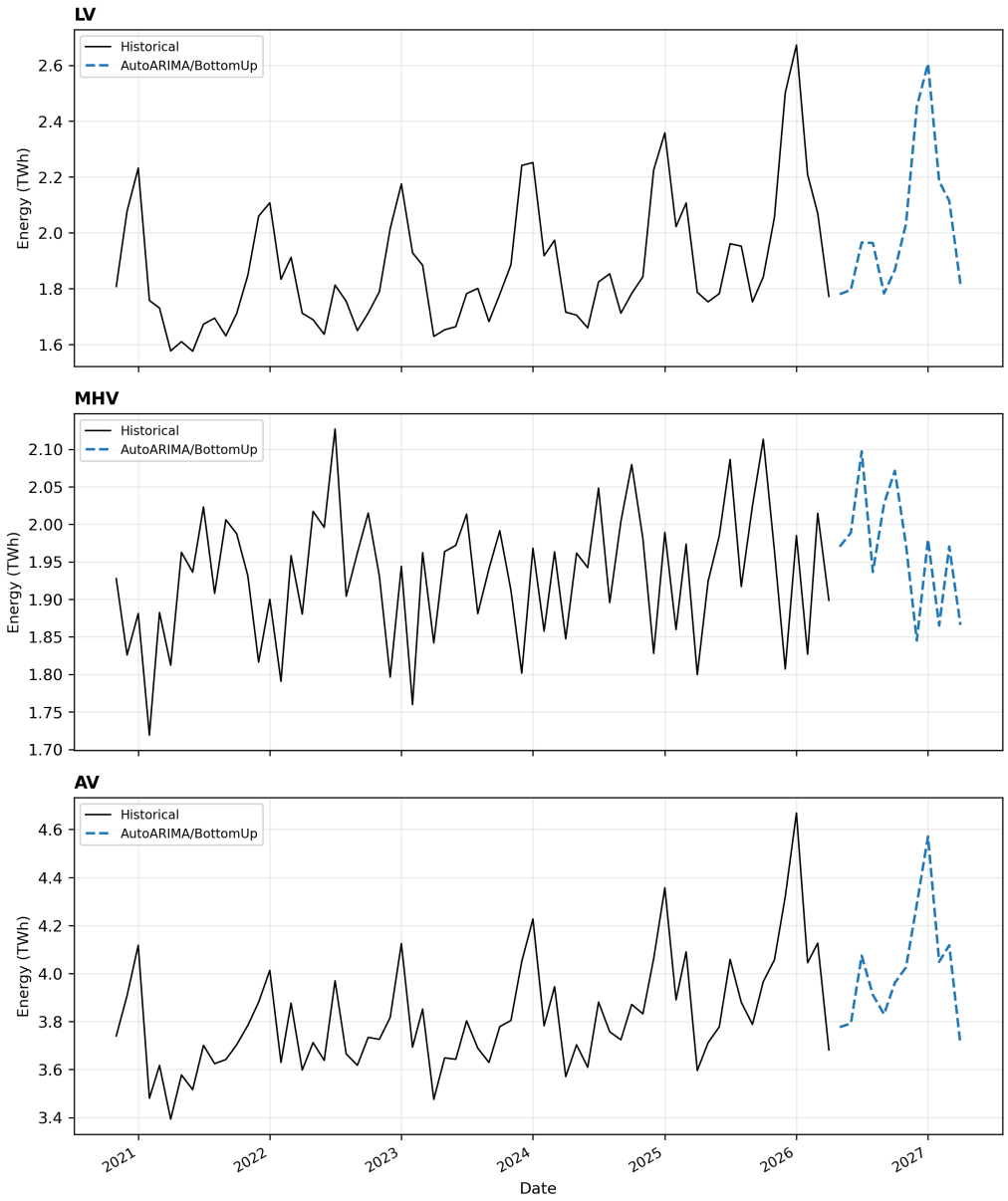


Figure 4.16: Historical national consumption and forecasts for low voltage, medium, high, and very high voltage, and all voltage levels, from top to bottom.

To examine the spatial distribution of the projections, Figure 4.17 compares parish-level consumption forecast for May 2026 to April 2027 with consumption observed from May 2025 to April 2026. The maps show that projected percentage changes vary across parishes and between voltage scopes. Grey parishes are not included in the corresponding analysis. Enlarged, individually scaled versions of the three maps are provided in Appendix D.

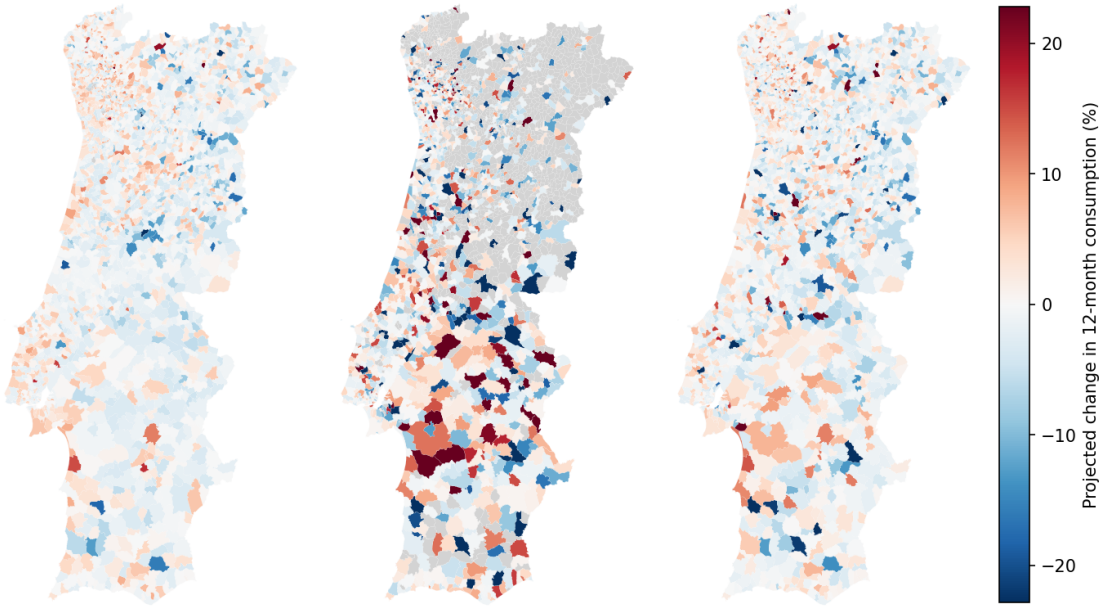


Figure 4.17: Projected percentage change in parish-level consumption for low voltage, medium, high, and very high voltage, and all voltage levels, from left to right.

These forecasts constitute an application of the framework beyond the evaluation period. Because the forecast period extends to April 2027, observed values for the full period are not yet available, and forecast accuracy cannot be assessed.

Chapter 5

Conclusion and Future Work

This dissertation developed and evaluated a hierarchical electric load forecasting framework across three voltage scopes and examined the growth of self-consumption in mainland Portugal. This closing chapter presents its conclusions. Section 5.1 brings together its main findings, Section 5.2 discusses data, methodological, and scope limitations, and Section 5.3 outlines directions for future research.

5.1 Conclusion

This dissertation developed and applied a hierarchical forecasting framework for electricity consumption in mainland Portugal, alongside an analysis of self-consumption growth, showing that electricity demand and distributed self-consumption are evolving concurrently across multiple levels of spatial aggregation. Using publicly available E-REDES data, the framework was structured around the four-level administrative hierarchy of mainland Portugal, namely national, district, municipality, and parish, and applied to three hierarchies comprising 3170, 1825, and 3179 series, covering low-voltage, medium, high, and very high voltage, and all-voltage consumption, respectively. For each hierarchy, `AutoARIMA` and a `Naive` benchmark were fitted to every series and reconciled using bottom-up, top-down, and middle-out approaches.

`AutoARIMA` substantially outperformed the `Naive` benchmark at every level and under every reconciliation method, with MASE values between two and four times lower, and the medium and high voltage forecasts were the most accurate in relative terms, followed by all-voltage and then low-voltage. `AutoARIMA` also outperformed `Seasonal Naive` in ten of the twelve level-voltage combinations examined. In this comparison, `Seasonal Naive` was evaluated as a competing model, rather than used only as the normalizing benchmark in MASE. The two exceptions occurred at the national and district levels within the medium-, high-, and very-high-voltage group, whose test period contained no atypical month. However, `AutoARIMA` retained its advantage at the municipality and parish levels for the same group.

The best reconciliation method, however, was not the same across the three analyses: bottom-up dominated the entire low-voltage hierarchy completely and the medium, high, and very high

voltage hierarchy to a lesser extent, whereas in the all-voltage case top-down performed best at the national and district levels and bottom-up regained the advantage at the parish level, indicating that the reconciliation strategy should be selected empirically rather than fixed *a priori*. Every model, at every level, produced a MASE above one for the low voltage and all-voltage analyses. This result is traced to an atypical surge in consumption in January 2026 that exceeded any month observed during training. No such anomaly was observed for medium, high, and very high voltage, offering an indication that the two underlying segments respond, at least in part, to different drivers.

The exploratory analysis of self-consumption growth also yielded results relevant to distribution network planning. Installed low-voltage solar PV capacity more than tripled between January 2023 and April 2026, reaching 934 MW across 272 922 installations in all 278 municipalities and the great majority of parishes served by E-REDES. Over the same period, the underlying billed consumption trend continued to rise rather than decline, despite this trend being understood to be net of self-consumption. This is consistent with the electrification of end uses offsetting, or even outweighing, the demand-reducing effect of distributed self-consumption, although it remains an observed association rather than a causal claim. The spatial distribution of this growth also differed from that of consumption. The Lisboa and Porto metropolitan areas, which install the most capacity in absolute terms, show the lowest penetration once normalised by their consumption base. By contrast, interior and southern municipalities exhibit substantially higher penetration despite modest absolute capacity. This pattern is only visible at the fine spatial granularity adopted in this work.

Parish-level forecast accuracy was found to be spatially heterogeneous in a way that aggregate metrics conceal: parish-level MASE ranged from close to zero to several times the corresponding hierarchy-wide average, with no evidence of a single, uniform geographical gradient. No systematic association was found between this heterogeneity and local self-consumption density, suggesting that, within the available data and test period, higher behind-the-meter PV penetration does not, by itself, degrade forecastability. Under bottom-up reconciliation, parish-level errors propagate directly to higher levels. However, their influence on aggregate accuracy depends on both the absolute consumption associated with each parish and the extent to which errors reinforce or offset one another. Consequently, a small number of large-consumption parishes may have a disproportionate influence on national and district accuracy, even when forecast accuracy varies substantially elsewhere in the hierarchy.

Finally, the case study revealed two structural characteristics of the E-REDES dataset that are relevant beyond the specific results reported here. First, it reflects the administrative geography that predates the 2025 parish reform, consistent with the recency of that reform. Second, its low-voltage coverage does not extend to parishes served by other distribution operators, such as local electricity cooperatives, which lie outside E-REDES's concession area. Both characteristics were identified, quantified, and consistently accounted for throughout the analysis. The parish codification is stable across the entire study period, so it does not affect the internal validity of the hierarchical framework or its results, but it does limit direct comparability with datasets that

already adopt the post-2025 administrative division.

5.2 Limitations

There are also limitations relating to the data used, the methodology adopted, and the scope of the analysis, and these should be taken into account when interpreting the results.

5.2.1 Data Limitations

The E-REDES consumption dataset aggregates very high, high, and medium voltage consumption into a single category, so only the low-voltage level could be analysed in isolation. Coverage of the low-voltage network is also incomplete: nine parishes, served by local electricity cooperatives rather than E-REDES, have no low-voltage records and were excluded from that hierarchy, and 1.2% of total billed energy could not be attributed to any parish and was excluded from all analyses. One parish, Rebordosa, has an incomplete series, with 48 training months and 6 test months instead of 54 and 12, consistent with the presence of a cooperative low-voltage operator in its municipality, as discussed in Section 4.1.

Finally, since billed consumption is net of self-consumption, the data do not allow gross consumption to be recovered from the billed series alone.

5.2.2 Methodological Limitations

The evaluation covers a single twelve-month test period, which happened to include an atypical surge in January 2026 that exceeded any month observed during training for the low voltage and all voltage analyses. The reported accuracy results may therefore not generalize to other periods.

The all-voltage series combines two segments, low voltage and medium/high voltage, whose forecast volatility may differ. This was proposed to explain the reversal in the best reconciliation method between the low-voltage and all-voltage analyses, and was partly examined through an isolated analysis of the medium, high, and very high voltage segment (Section 4.5). That analysis showed no anomalous test month for medium, high, and very high voltage, and a bottom-up-dominated reconciliation pattern resembling low voltage rather than all voltage, both consistent with the two segments responding to different underlying drivers. However, the specific role of seasonal misalignment between the segments was not tested statistically, and the isolated analysis itself relies on a filtered, non-representative subset of parishes with complete training histories (Section 4.1). The reversal therefore remains only partially explained.

Only `AutoARIMA`, `Naive` and `Seasonal Naive` were considered as competing forecasting models, so other approaches were not evaluated.

Finally, the framework produces point forecasts only, without quantifying forecast uncertainty, which would be directly relevant to network planning under demand and self-consumption growth.

5.2.3 Scope Limitations

One extension was identified but was not undertaken within this dissertation. The *Energy Injected into Self-Consumption Production Units* dataset [40], which reports actual energy injected into the grid rather than installed capacity, was not incorporated, as it is only available at the municipality level and is therefore incompatible with the parish-level granularity adopted throughout this work.

5.3 Future Work

Several avenues for future work arise naturally from the results and limitations discussed in this dissertation.

The isolated medium-, high-, and very-high-voltage analysis (Section 4.5) relies on a filtered subset of 1531 parishes with complete training histories, out of 1626 with any record. Extending it to the full segment is a challenging problem, since the excluded parishes are precisely those with the shortest and most irregular histories, and medium-, high-, and very-high-voltage connections are themselves highly heterogeneous, ranging from small commercial premises to large industrial loads. Global forecasting models, trained jointly across all parish-level series rather than independently per parish, are one possible avenue worth exploring. Their effectiveness would, however, depend on the extent to which parishes with limited history share structure with better-recorded ones, which remains an open question given this heterogeneity.

The *Energy Injected into Self-Consumption Production Units* dataset [40], which reports the surplus energy actually delivered to the grid by self-consumption units rather than their installed capacity, was identified but not incorporated in this work, since it is available only at the municipality level. A natural extension would be to construct a separate higher-level hierarchy for injected energy, covering the national, district, and municipality levels without a parish-level bottom, alongside the parish-level consumption hierarchy used in this dissertation, and to compare them at the levels they have in common. This would provide a more direct measure of the impact of self-consumption on grid flows than installed capacity alone, without requiring the two hierarchies to be reconciled together.

Only `AutoARIMA`, `Naive`, and `Seasonal Naive` were evaluated as forecasting models. Testing additional statistical and machine-learning forecasters within the same reconciliation framework would clarify how specific the results are to `AutoARIMA`. Two further extensions would improve the robustness of the evaluation independently of one another: extending the framework to produce probabilistic forecasts would quantify forecast uncertainty, while evaluating it with a rolling-origin scheme across multiple test periods would provide a more robust accuracy assessment than the single test window used in this dissertation.

Finally, this dissertation addresses only the hierarchical component of the broader integration between hierarchical forecasting and simulation-based spatial disaggregation discussed in Section

2.3. The parish-level forecasts produced here could serve as the input layer to a subsequent simulation stage, such as a cellular automaton or multi-agent model, as reviewed there, disaggregating consumption to a finer, sub-parish resolution.

References

[1] R. Bernards, J. Morren, and H. Slootweg, "Development and implementation of statistical models for estimating diversified adoption of energy transition technologies," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1540–1554, 2018. DOI: 10.1109/TSTE.2018.2794579

[2] I. I. Avramidis, V. A. Evangelopoulos, P. S. Georgilakis, and N. D. Hatziargyriou, "Demand side flexibility schemes for facilitating the high penetration of residential distributed energy resources," *IET Generation, Transmission and Distribution*, vol. 12, pp. 4079–4088, 2018, ISSN: 17518687. DOI: 10.1049/iet-gtd.2018.5415

[3] F. Heymann, J. Silva, V. Miranda, J. Melo, F. J. Soares, and A. Padilha-Feltrin, "Distribution network planning considering technology diffusion dynamics and spatial net-load behavior," *International Journal of Electrical Power Energy Systems*, vol. 106, pp. 254–265, 2019, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2018.10.006

[4] H. Willis and J. Northcote-Green, "Spatial electric load forecasting: A tutorial review," *Proceedings of the IEEE*, vol. 71, no. 2, pp. 232–253, 1983. DOI: 10.1109/PROC.1983.12562

[5] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016, ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2015.11.011

[6] V. Evangelopoulos, P. Karafotis, and P. Georgilakis, "Probabilistic spatial load forecasting based on hierarchical trending method," *Energies*, vol. 13, no. 18, Sep. 2020, ISSN: 19961073. DOI: 10.3390/en13184643

[7] H. Willis, *Spatial Electric Load Forecasting*, 2nd ed. New York, NY, USA: Marcel Dekker, 2002.

[8] S. Sasmono, N. I. Sinisuka, M. W. Atmopawiro, and D. Darwanto, "Macro demand spatial approach (mdsa) at spatial demand forecasting for transmission system planning," International Journal on Electrical Engineering and Informatics, Tech. Rep., 2015.

[9] V. A. Evangelopoulos and P. S. Georgilakis, "Probabilistic spatial load forecasting for assessing the impact of electric load growth in power distribution networks," *Electric Power Systems Research*, vol. 207, Jun. 2022, ISSN: 03787796. DOI: 10.1016/j.epsr.2022.107847

[10] H.-C. Wu and C.-N. Lu, "A data mining approach for spatial modeling in small area load forecast," *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 516–521, 2002. DOI: 10.1109/TPWRS.2002.1007927

[11] V. Miranda and C. Monteiro, "Fuzzy inference in spatial load forecasting," in *2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.00CH37077)*, vol. 2, 2000, pp. 1063–1068. DOI: 10.1109/PESW.2000.850087

[12] E. M. Carreno and A. Padilha-Feltrin, "Evolutionary heuristic to determine future land use," in *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–6. DOI: 10.1109/PES.2008.4596675

[13] E. M. Carreno, R. M. Rocha, and A. Padilha-Feltrin, "A cellular automaton approach to spatial electric load forecasting," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 532–540, 2011. DOI: 10.1109/TPWRS.2010.2061877

[14] Y. He, J. Zhang, Y. Xu, Y. Gao, T. Xia, and H. He, "Forecasting the urban power load in china based on the risk analysis of land-use change and load density," *International Journal of Electrical Power Energy Systems*, vol. 73, pp. 71–79, 2015, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2015.03.018

[15] J. D. Melo, E. M. Carreno, and A. Padilha-Feltrin, "Multi-agent simulation of urban social dynamics for spatial load forecasting," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1870–1878, 2012. DOI: 10.1109/TPWRS.2012.2190109

[16] T. Hong, S. M. Hsiang, and L. Xu, "Human-machine co-construct intelligence on horizon year load in long term spatial load forecasting," in *2009 IEEE Power Energy Society General Meeting*, 2009, pp. 1–6. DOI: 10.1109/PES.2009.5275308

[17] D. X. Morales, Y. Besanger, S. Moscoso, and P. Pesantez, "Development of a spatial load-forecasting module for optimizing planning of electricity supply," in *2017 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, 2017, pp. 1–6. DOI: 10.1109/ISGT-LA.2017.8126738

[18] R. L. Vasquez-Arnez, J. A. Jardini, R. Casolari, L. C. Magrini, R. Semolini, and J. R. Pascon, "A methodology for electrical energy forecast and its spatial allocation over developing boroughs," in *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, 2008, pp. 1–6. DOI: 10.1109/TDC.2008.4517252

[19] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 357–363, 2014, ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2013.07.001

[20] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, "Optimal combination forecasts for hierarchical time series," *Computational Statistics and Data Analysis*, vol. 55, pp. 2579–2589, Sep. 2011, ISSN: 01679473. DOI: 10.1016/j.csda.2011.03.006

[21] 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. DOI: 10.1016/j.ijforecast.2023.10.010

[22] C. Ye, Y. Ding, P. Wang, and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Transactions on Power Systems*, vol. 34, pp. 1966–1979, 3 May 2019, ISSN: 15580679. DOI: 10.1109/TPWRS.2018.2889995

[23] S. Zambrano-Asanza, R. Morales, J. A. Montalvan, and J. F. Franco, "Integrating artificial neural networks and cellular automata model for spatial-temporal load forecasting," *International Journal of Electrical Power Energy Systems*, vol. 148, p. 108 906, 2023, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2022.108906

[24] A. Panagiotelis, P. Gamakumara, G. Athanasopoulos, and R. J. Hyndman, "Probabilistic forecast reconciliation: Properties, evaluation and score optimisation," *European Journal of Operational Research*, vol. 306, pp. 693–706, 2 Apr. 2023, ISSN: 03772217. DOI: 10.1016/j.ejor.2022.07.040

[25] J. D. Melo, E. M. Carreno, A. Calviño, and A. Padilha-Feltrin, "Determining spatial resolution in spatial load forecasting using a grid-based model," *Electric Power Systems Research*, vol. 111, pp. 177–184, 2014, ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2014.02.019

[26] R. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice," in 3rd ed. Australia: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/index.html

[27] X. Chang, M. Gao, Y. Wang, and X. Hou, "Seasonal autoregressive integrated moving average model for precipitation time series," *Journal of Mathematics and Statistics*, vol. 8, pp. 500–505, 4 2012, ISSN: 15493644. DOI: 10.3844/jmssp.2012.500.505

[28] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," *Journal of Statistical Software*, vol. 27, 2008. DOI: 10.18637/jss.v027.i03

[29] A. Garza, M. M. Canseco, C. Challú, and K. G. Olivares, *StatsForecast: Lightning fast forecasting with statistical and econometric models*, PyCon Salt Lake City, Utah, US 2022, 2022. [Online]. Available: https://github.com/Nixtla/statsforecast

[30] K. Bandara, R. J. Hyndman, and C. Bergmeir, *MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns*, 2021. arXiv: 2107.13462 [stat.AP]. [Online]. Available: https://arxiv.org/abs/2107.13462

[31] K. G. Olivares et al., *HierarchicalForecast: A Reference Framework for Hierarchical Forecasting in Python*, 2024. [Online]. Available: https://arxiv.org/abs/2207.03517

[32] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018. DOI: 10.1213/ANE.0000000000002864

[33] E-REDES, *E-REDES Open Data Portal*, https://e-redes.opendatasoft.com/pages/homepage/.

[34] E-REDES, *Monthly consumption by municipality*, https://e-redes.opendatasoft.com/explore/dataset/3-consumos-faturados-por-municipio-ultimos-10-anos/.

[35] E-REDES, *Total production Units for Self-Consumption - Monthly*, https://e-redes.opendatasoft.com/explore/dataset/8-total-upac-mensal/.

[36] Direção-Geral do Território, "Catálogo de Entidades da CAOP2019," Direção-Geral do Território, Tech. Rep., 2019.

[37] ERSE, "Relatório de Ponto de Situação sobre as Redes Inteligentes de Energia Elétrica em Portugal Continental," Entidade Reguladora dos Serviços Energéticos, Lisboa, Portugal, Tech. Rep., 2025, Data referenced to the first semester of 2025. [Online]. Available: https://www.erse.pt/media/o0yaryjc/relat%C3%B3rio-redes-inteligentes-2025.pdf

[38] Diário da República, *Lei n.º 25-a/2025, de 13 de março*, Diário da República, 2025. [Online]. Available: https://diariodarepublica.pt/dr/detalhe/lei/25-a-2025-910933580

[39] P. Virtanen et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2

[40] E-REDES, *Energy Injected into Self-Consumption Production Units*, https://e-redes.opendatasoft.com/explore/dataset/energia_injectada_upac/.

Appendix A

District-Level Self-Consumption Figures

This appendix presents the solar photovoltaic self-consumption capacity, monthly consumption, number of installations, and resulting density for all 18 districts of mainland Portugal in April 2026, referenced in Section 4.2. Density is defined as installed UPAC capacity divided by monthly consumption, expressed in kW per GWh; districts are ordered from highest to lowest density.

Table A.1: Solar PV self-consumption capacity, consumption, and density by district (April 2026).

District	Consumption (GWh)	UPAC (MW)	Installations	Density (kW/GWh)
Beja	26.69	22.47	4,484	841.8
Leiria	88.87	70.08	18,706	788.6
Santarém	74.50	58.50	18,329	785.4
Viseu	57.38	44.61	13,860	777.5
Évora	30.02	22.87	5,272	761.8
Braga	134.61	99.72	32,168	740.8
Bragança	22.74	16.62	4,528	731.0
Portalegre	19.72	14.33	3,200	726.9
Vila Real	30.73	20.52	7,303	667.7
Guarda	24.62	16.14	4,329	655.5
Castelo Branco	35.38	22.96	6,633	649.0
Aveiro	114.17	73.93	21,695	647.5
Coimbra	69.23	43.75	14,827	631.9
Viana do Castelo	41.73	24.56	8,022	588.5
Faro	139.15	69.12	15,254	496.7
Setúbal	138.29	68.64	25,011	496.3
Porto	317.38	134.62	38,233	424.2
Lisboa	407.39	110.94	31,068	272.3

Appendix B

Pipeline Pseudocode

This appendix presents pseudocode for the two core procedures of the forecasting pipeline described in Section 4.3: construction of the hierarchy and summing matrix from parish-level codes (Algorithm 1), and the end-to-end pipeline that ties data loading, decomposition, base forecasting, reconciliation, and evaluation together (Algorithm 2).

Algorithm 1 Build Hierarchy and Summing Matrix

Input: Y_{bottom}, a table of (parish_code, date, consumption) records, where each parish_code is a six-digit DICOFRE string

Output: Y, the consumption table extended with national, district, and municipality series; S, the summing matrix; tags, a dictionary mapping each hierarchical level to its list of series identifiers; bottom_ids, municipality_ids, district_ids, the sorted identifier lists for each level

function BUILDHIERARCHY(Y_{bottom})
  $Y_{\text{municipality}} \leftarrow$ copy of Y_{bottom} with each parish_code truncated to its first 4 characters
  $Y_{\text{municipality}} \leftarrow$ group $Y_{\text{municipality}}$ by (municipality_code, date), summing consumption
  $Y_{\text{district}} \leftarrow$ copy of Y_{bottom} with each parish_code truncated to its first 2 characters
  $Y_{\text{district}} \leftarrow$ group Y_{district} by (district_code, date), summing consumption
  $Y_{\text{national}} \leftarrow$ copy of Y_{bottom} with every code replaced by `"PT"`
  $Y_{\text{national}} \leftarrow$ group Y_{national} by date, summing consumption
  $Y \leftarrow$ concatenate $Y_{\text{national}}, Y_{\text{district}}, Y_{\text{municipality}}, Y_{\text{bottom}}$
  bottom_ids $\leftarrow$ sorted list of unique parish codes in Y_{bottom}
  municipality_ids $\leftarrow$ sorted list of unique municipality codes in $Y_{\text{municipality}}$
  district_ids $\leftarrow$ sorted list of unique district codes in Y_{district}
  all_ids $\leftarrow$ ["PT"]+ district_ids + municipality_ids + bottom_ids
  tags $\leftarrow$ {"Portugal" : ["PT"], "Districts" : district_ids, "Municipalities" : municipality_ids, "Parishes" : bottom_ids}
  Initialise S as a matrix of zeros, with one row per identifier in all_ids and one column per identifier in bottom_ids
  Set row `"PT"` of S to 1 in every column
  for each parish code b in bottom_ids **do**
    Set $S[b[0{:}2], b] \leftarrow 1$ ▷ district row
    Set $S[b[0{:}4], b] \leftarrow 1$ ▷ municipality row
    Set $S[b, b] \leftarrow 1$ ▷ parish row (identity)
  end for
  return Y, S, tags, bottom_ids, municipality_ids, district_ids
end function

Algorithm 2 Forecasting Pipeline

Input: voltage_filter, one of {`"Baixa Tensão"`, `"Muito Alta, Alta e Média Tensões"`, or `None` for all voltage levels}; cutoff, the train/test split date; h, the forecast horizon

Output: Y_{rec}, the reconciled forecasts for every series and every combination of base model and reconciliation method; eval, the accuracy metrics by hierarchical level

function RUNPIPELINE(voltage_filter, cutoff, h)
    $D \leftarrow$ load the raw consumption dataset
    Remove records belonging to the `OUTROS` category, and records with a missing parish name, from D
    if voltage_filter is not `None` **then**
        Keep only records in D matching voltage_filter
    end if
    $Y_{\text{bottom}} \leftarrow$ aggregate D by (parish_code, date), summing consumption
    $Y, S, \text{tags}, \text{bottom_ids}, \text{municipality_ids}, \text{district_ids} \leftarrow$ BUILDHIERARCHY(Y_{bottom})
    $Y_{\text{train}} \leftarrow$ rows of Y with date $\leq$ cutoff
    $Y_{\text{test}} \leftarrow$ rows of Y with date $>$ cutoff
    $Y_{\text{national,train}} \leftarrow$ rows of Y_{train} with identifier `"PT"` only
    Decompose $Y_{\text{national,train}}$ into trend, seasonal, and remainder components using MSTL (season_length $= 12$)
    Save the MSTL components to disk
    Fit `AutoARIMA` (season_length $= 12$, $d_{\max} = 1$, $D_{\max} = 1$, stepwise, no approximation) and `Naive` jointly to every series in Y_{train}, producing h-month-ahead forecasts $\hat{Y}$ for both models in a single call
    Reconcile $\hat{Y}$ jointly under BottomUp, TopDown (forecast proportions), and MiddleOut (pivoting at the municipality level, forecast proportions), using S, tags, and Y_{train}, in a single call, producing Y_{rec}
    $Y_{\text{eval_input}} \leftarrow$ join Y_{rec} with Y_{test} on (identifier, date), attaching the observed value
    eval $\leftarrow$ compute MAE, RMSE, and MASE on $Y_{\text{eval_input}}$, broken down by the hierarchical levels in tags and by every combination of model and reconciliation method, using Y_{train} as the reference period for MASE, in a single call
    Save {voltage_filter, cutoff, h, |bottom_ids|, |municipality_ids|, |district_ids|, |all_ids|} to disk as experiment metadata
    return Y_{rec}, eval
end function

Appendix C

Spatial Error Maps

This appendix presents enlarged versions of the parish-level spatial error maps summarized in Figure 4.14, Section 4.7.3, for closer inspection of individual parishes and regions.

Unlike Figure 4.14, which uses a colour scale shared across the three analyses for direct visual comparison, each map in this appendix uses its own scale, calculated from that analysis alone, to maximize the visibility of spatial variation within each voltage scope.

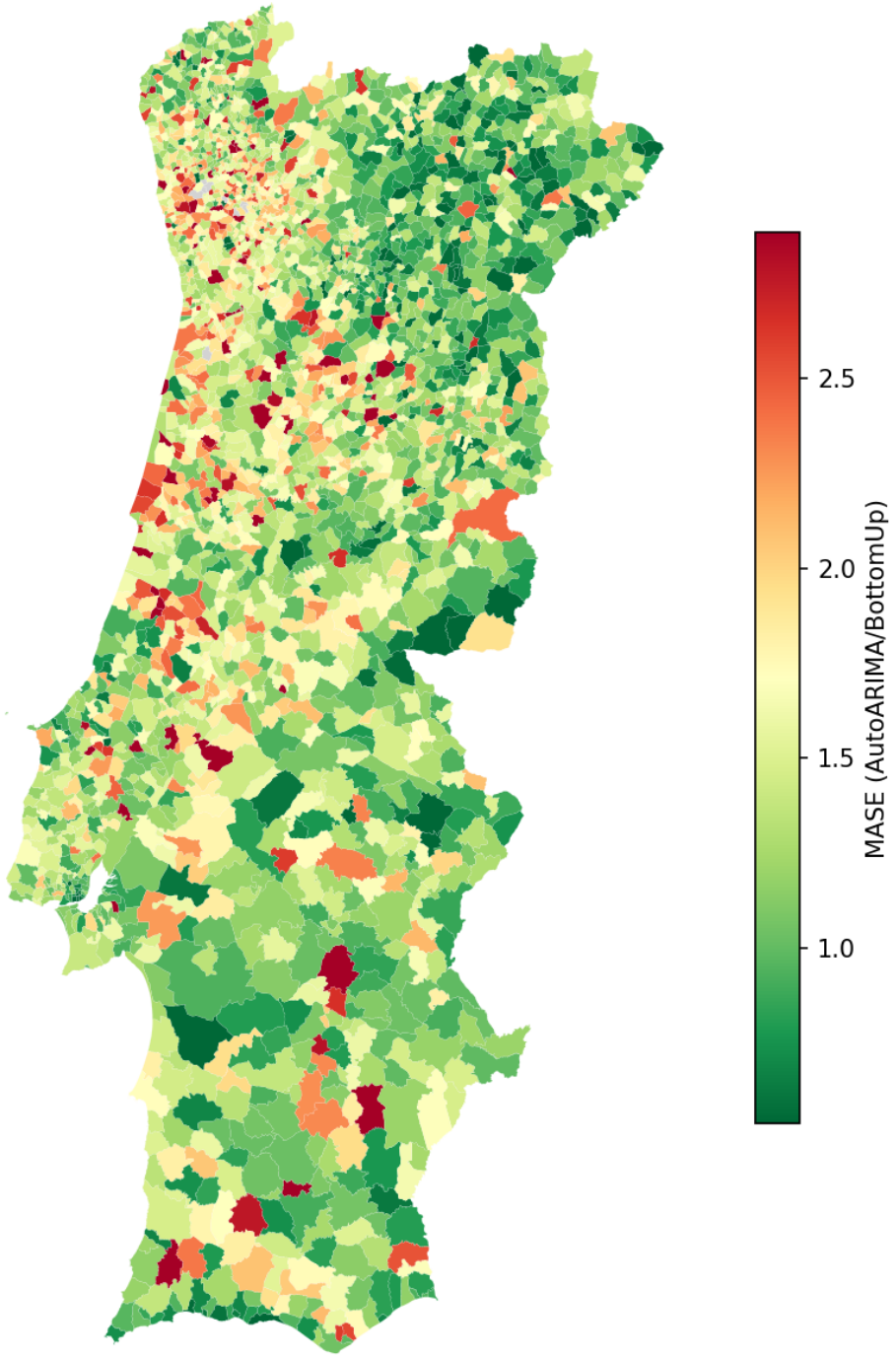


Figure C.1: Parish-level MASE obtained with AutoARIMA and Bottom-Up reconciliation, low voltage.

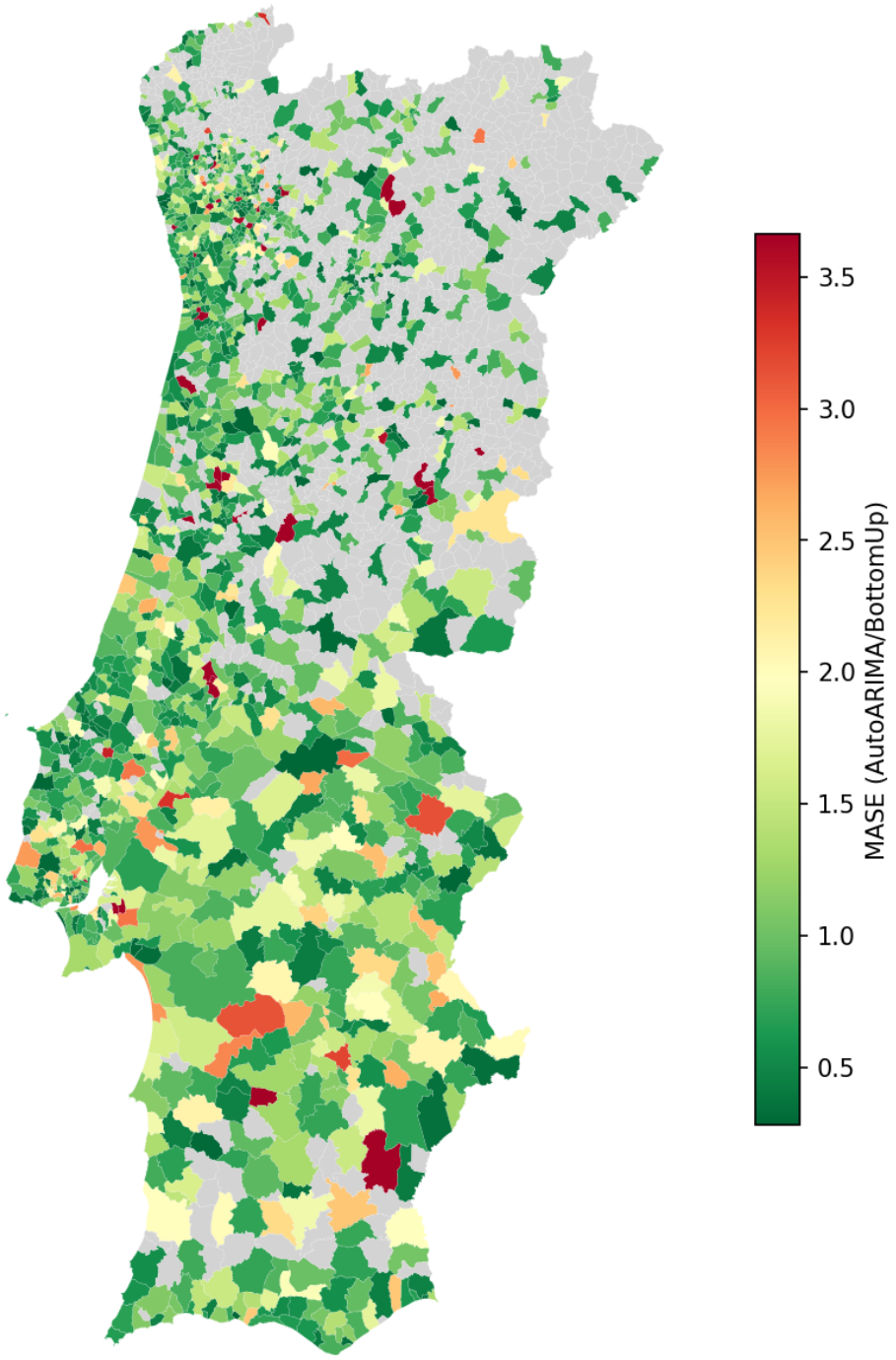


Figure C.2: Parish-level MASE obtained with AutoARIMA and Bottom-Up reconciliation, medium, high, and very high voltage.

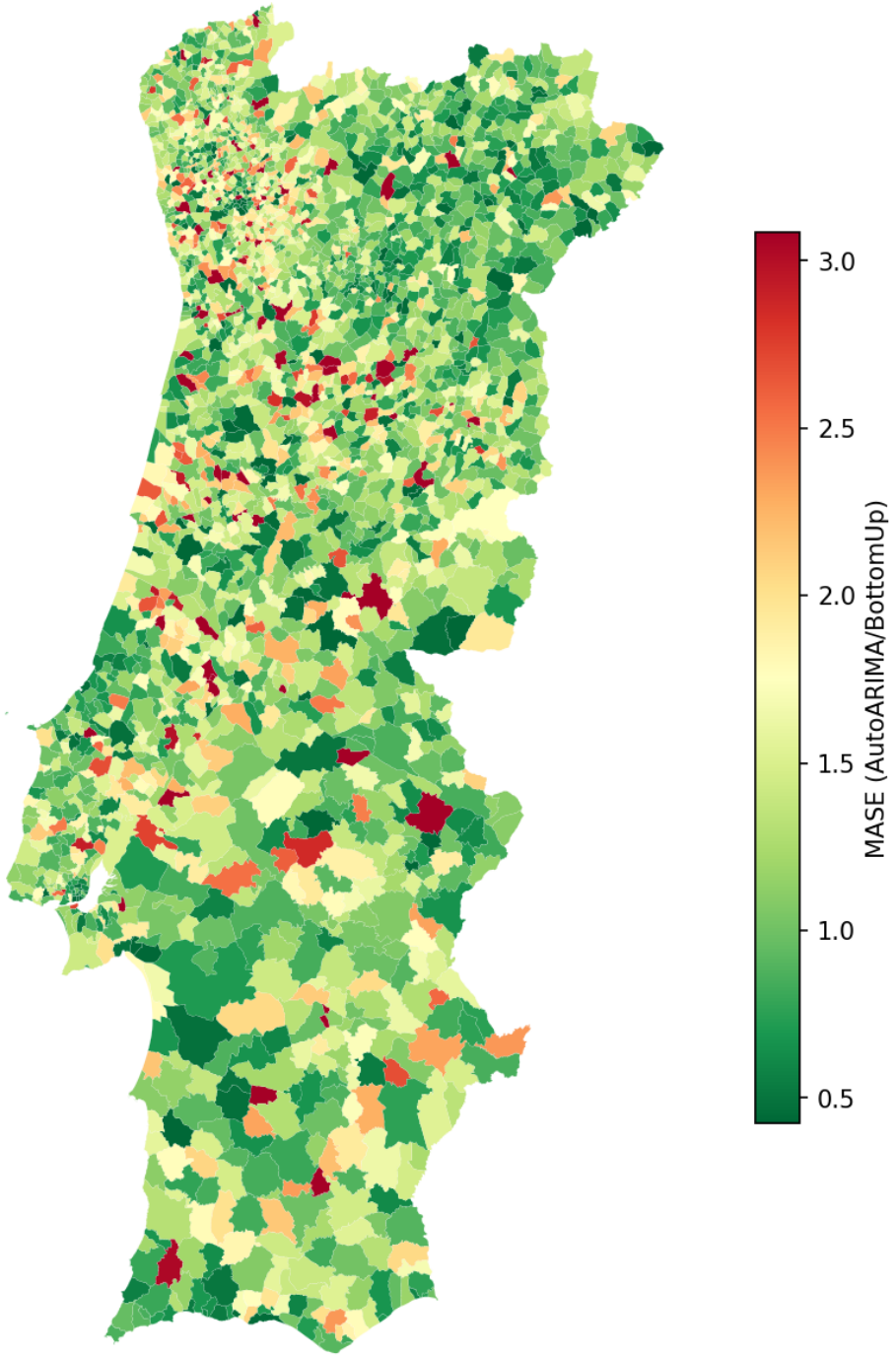


Figure C.3: Parish-level MASE obtained with AutoARIMA and Bottom-Up reconciliation, all voltage.

Appendix D

Projected Changes in Parish-Level Consumption

This appendix presents enlarged versions of the three parish-level maps shown in Figure 4.17, allowing closer inspection of the projected changes within each voltage scope.

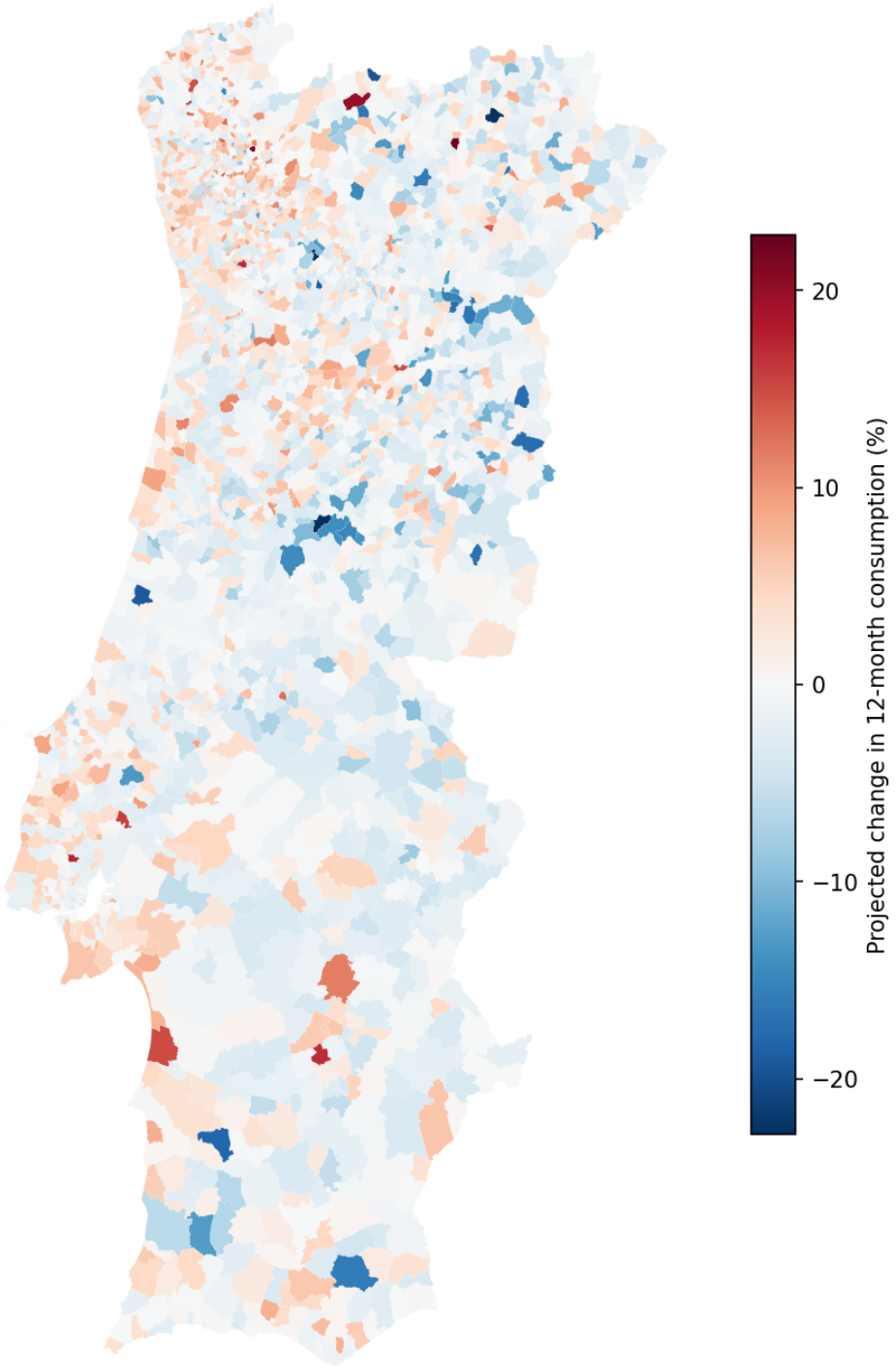


Figure D.1: Projected percentage change in parish-level consumption, low voltage.

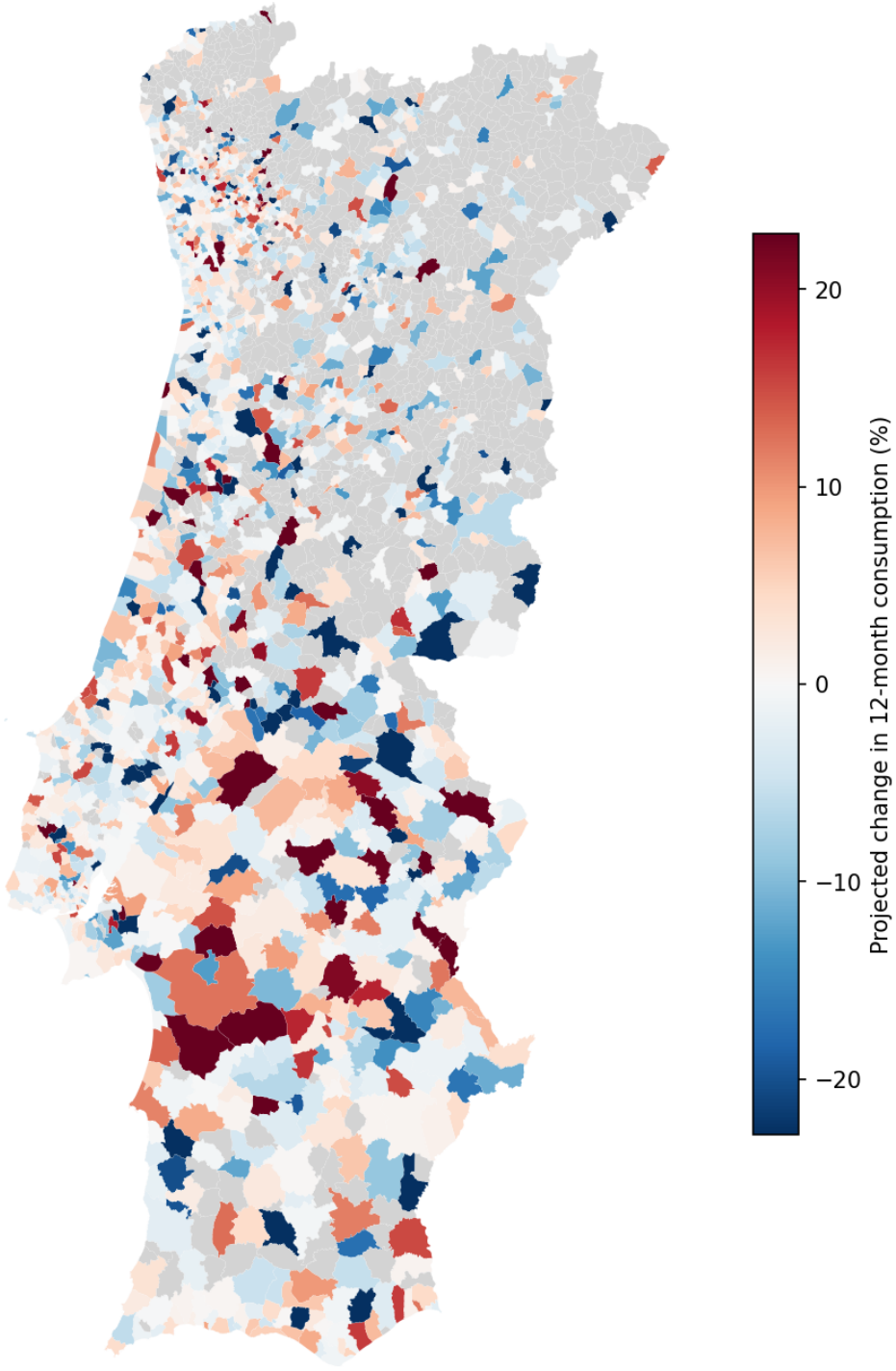


Figure D.2: Projected percentage change in parish-level consumption, medium, high, and very high voltage.

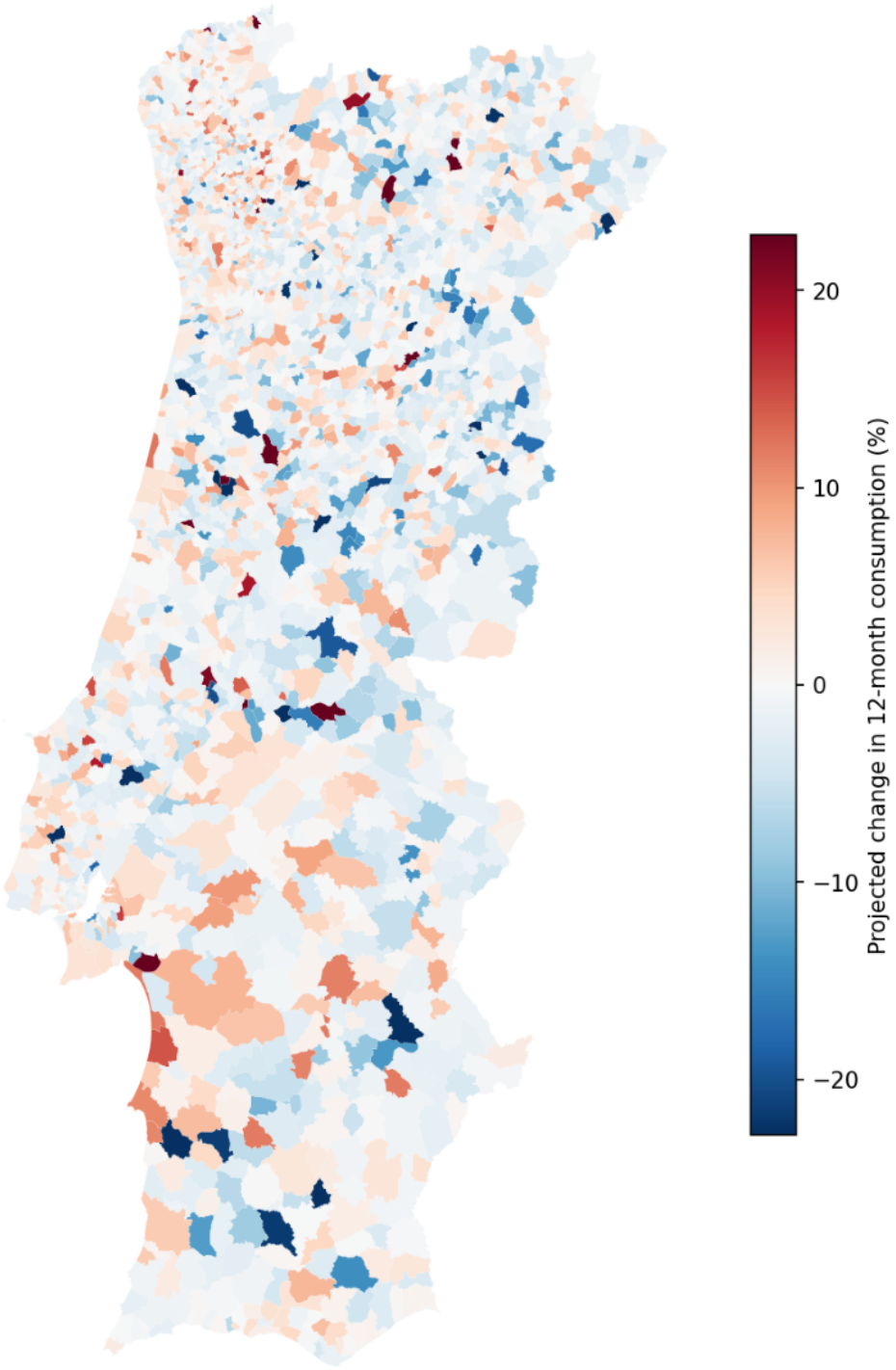


Figure D.3: Projected percentage change in parish-level consumption, all voltage.