



**POLITECNICO
MILANO 1863**

DOCTORAL PROGRAM IN DATA ANALYTICS AND DECISION SCIENCES
DEPARTMENT OF ELECTRONICS, INFORMATION AND BIOENGINEERING

DATA-DRIVEN MODELS FOR ELECTRICITY SPOT MARKETS

MONITORING, FORECASTING AND STRATEGIC BIDDING

Supervisor and Chair of the Doctoral Program:

Prof. Piercesare Secchi

Co-Supervisor:

Dr. Filippo Bovera

Tutor:

Prof. Simone Vantini

Funding:

Energy Team S.p.A.

Doctoral Dissertation of:
Guillaume Koechlin

Year 2026 – Cycle XXXVIII

Abstract

ELECTRICITY markets are becoming increasingly complex due to the transition towards decarbonized power systems, characterized by high penetration of variable renewable energy. In this context, this PhD thesis develops data-driven electricity spot market models, applied to the Italian case. It addresses three critical steps of a market actor's decision-making pipeline: monitoring market fundamentals, predicting participant behaviors to forecast market results, and defining optimal bidding strategies to maximize revenues and foster competition. The proposed approaches rely on statistical and machine learning methods for binary, time series and functional data, with a particular emphasis on uncertainty quantification and model interpretability. Specifically, the first contribution is a blind source separation framework based on a constrained variant of non-negative matrix factorization, which enables the high-frequency monitoring of unobserved sectoral power demand from aggregate grid load measurements. The second contribution introduces a general functional data analysis framework for predicting supply and demand merit-order curves in auction-based markets, from which both point and probabilistic electricity price forecasts can be derived. The third contribution proposes a probabilistic classification pipeline to model pay-as-bid power reserve markets, allowing participants to estimate price-dependent bid acceptance probabilities and optimally price their flexibility.

Keywords: electricity spot markets; functional data analysis; strategic bidding; load disaggregation; probabilistic forecasting

Sommario

I MERCATI elettrici stanno diventando sempre più complessi a causa della transizione verso sistemi energetici decarbonizzati, caratterizzati da un'alta penetrazione di fonti rinnovabili variabili. In questo contesto, questa tesi di dottorato sviluppa modelli data-driven per i mercati elettrici spot, applicati al caso italiano. La ricerca affronta tre fasi critiche della decision-making pipeline di un attore di mercato: il monitoraggio dei fondamentali di mercato, la previsione dei comportamenti dei partecipanti per stimare i risultati di mercato, e la definizione di strategie di offerta (bidding) ottimali per massimizzare i ricavi e favorire la concorrenza. Gli approcci proposti si basano su metodi statistici e di machine learning per dati binari, serie storiche e dati funzionali, con un particolare accento sulla quantificazione dell'incertezza e sull'interpretabilità dei modelli. Nello specifico, il primo contributo consiste in un framework di blind source separation basato su una variante vincolata della non-negative matrix factorization, che consente il monitoraggio ad alta frequenza della domanda elettrica settoriale non osservata a partire dalle misure aggregate del carico di rete. Il secondo contributo introduce un framework generale di functional data analysis per la previsione delle curve di merit-order di domanda e offerta nei mercati basati su aste, da cui è possibile derivare previsioni di prezzo dell'elettricità sia puntuali che probabilistiche. Il terzo contributo propone una pipeline di classificazione probabilistica per modellare i mercati della riserva di tipo pay-as-bid, permettendo ai partecipanti di stimare le probabilità di accettazione delle offerte in funzione del prezzo e di prezzare in modo ottimale la propria flessibilità.

Parole chiave: mercati elettrici spot; analisi dei dati funzionali; bidding strategico; disaggregazione del carico; previsione probabilistica

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Piercesare Secchi and Filippo Bovera, for their continuous support and guidance throughout this research. Piercesare's expertise on statistics and Filippo's deep knowledge of power system economics, combined with their critical feedback, were essential to the completion of this work.

I am also grateful to Energy Team S.p.A. for providing the funding and resources necessary for this project. In particular, I thank Riccardo Barilli and Giulio Caprara for their support in creating the dataset used for the analysis in Chapter 2, and for their contribution to the deep and engaging discussions on the various topics developed in this thesis.

I must also acknowledge the Statistics Group of Terna S.p.A., and specifically Elena Degli Innocenti, for the collaboration that led to the framework in Chapter 1. Their trust in this project provided a vital foundation for the work that followed.

Finally, my warmest thanks go to Alessandra and my family for their constant encouragement. I am also grateful to my colleagues at the MOX laboratory of the Department of Mathematics and those at the Department of Energy for their friendship and making these years of doctoral study so stimulating.

Contents

Acronyms	XI
Scientific production	XIII
Introduction	1
Outline of the thesis	4
1 Blind source separation for sectoral demand monitoring	7
1.1 Introduction	7
1.1.1 Motivations	7
1.1.2 Objectives of the Work	9
1.2 Literature review	9
1.2.1 Load profiling and classification	10
1.2.2 Load disaggregation	11
1.2.3 Constrained NMF	14
1.2.4 Research gap and main contributions	14
1.3 Methodology	15
1.3.1 Data	15
1.3.2 Qualitative description	16
1.3.3 Model	19
1.3.4 Solution method	22
1.4 Results	23
1.4.1 LCNMF solutions analysis and interpretation	23
1.4.2 Disaggregated consumption	29
1.5 Discussion and conclusion	31

2	Forecasting merit-order curves in day-ahead markets	33
2.1	Introduction	33
2.1.1	Context	33
2.1.2	Objectives of the work	34
2.2	Literature review	35
2.2.1	A thematic review of prior work	35
2.2.2	Knowledge gap and contributions	40
2.3	Methodology	42
2.3.1	Curves definition	42
2.3.2	Curves representation	43
2.3.3	Forecasting the curves' vector representation	45
2.3.4	Clearing price point and probabilistic forecasting	47
2.3.5	Benchmark curves and price forecasting models	48
2.3.6	Forecasts evaluation	50
2.3.7	Daily recalibration	52
2.4	Data	53
2.4.1	Supply and demand curves	53
2.4.2	Exogenous predictors	54
2.5	Results	55
2.5.1	Functional principal component analysis	55
2.5.2	Curves forecasting	56
2.5.3	Clearing price forecast	57
2.5.4	Price probabilistic forecasting	59
2.5.5	Computational efficiency	61
2.6	Discussion and Conclusion	62
3	Strategic bidding in pay-as-bid reserve markets	67
3.1	Introduction	67
3.1.1	Context	67
3.1.2	The ancillary services market design	68
3.1.3	Related works and research novelty	70
3.2	Data	73
3.2.1	Target variable	73
3.2.2	Structural data	73
3.2.3	Exogenous data	74
3.2.4	Calendar data	75
3.2.5	Market data	75
3.3	Methodology	78
3.3.1	Bid acceptance prediction	78
3.3.2	Optimal bidding strategy	83
3.4	Results	85
3.4.1	Comparison with benchmark methods	85
3.4.2	On the importance of frequent model updating	86

3.4.3	Generalisation error assessment	87
3.4.4	Model explainability	89
3.4.5	Optimal bidding strategy	93
3.5	Discussion	96
3.6	Conclusion	97
Conclusions		99
	Summary of contributions	99
	Limitations and future research	100
Bibliography		105
A Supplementary material to Chapter 1		119
A.1	Modified multiplicative updates for the resolution of LCNMF . .	119
B Supplementary material to Chapter 2		121
B.1	Example of curves and clearing price prediction	121
B.2	Reliability and resolution of probabilistic forecasts	121

Acronyms

ASD	annual sector demand
AS	ancillary services
ASM	ancillary services market
AUC-PR	area under the precision-recall curve
BESS	battery energy storage systems
BSS	blind source separation
CDF	cumulative distribution function
CDM	central dispatching model
CPP	cocktail party problem
CRPS	continuous ranked probability score
CU	consumption unit
DER	distributed energy resource
DM	Diebold-Mariano
DR	downward regulation
DSO	distribution system operator
EBGL	Electricity Balancing Guideline
EPF	electricity price forecasting
FDA	functional data analysis
FPC	functional principal component
FPCA	functional principal component analysis
GLM	generalized linear model

Acronyms

GU	generation unit
HP	historical prices
PCA	principal component analysis
PSH	pumped storage hydropower
ID	intraday
ISP	integrated scheduling process
LCNMF	linearly-constrained non-negative matrix factorization
MGP	Mercato del Giorno Prima
MSI	monthly sectors' indicators
MU	multiplicative updates
NEMO	nominated electricity market operator
NILM	non-intrusive load monitoring
NMF	non-negative matrix factorization
NNLS	non-negative least squares
NTC	net transfer capacity
PaB	pay-as-bid
PI	prediction interval
PR	precision-recall
RES	renewable energy sources
RF	random forest
SDM	self-dispatching model
SHAP	SHapley Additive exPlanations
SO	system operator
TSO	transmission system operator
UP	uniform-price
UR	upward regulation
VRE	variable renewable energy
XAI	eXplainable AI

Scientific production

This dissertation is based on the following three main manuscripts (journal papers):

Koechlin, G., Bovera, F., & Secchi, P. (2026). Day-ahead electricity price forecasting using merit-order curves time series. *Under review*. (Koechlin et al., 2026b)

Koechlin, G., Bovera, F., Degli Innocenti, E., Santini B., Venturi A., Vazio, S., & Secchi, P. (2026). A blind source separation framework to monitor sectoral power demand from grid-scale load measurements. *Sustainable Energy, Grids and Networks*. (Koechlin et al., 2026a)

Koechlin, G., Bovera, F., & Secchi, P. (2025). Strategic bidding in pay-as-bid power reserve markets: a machine learning approach. *Energy Economics*. (Koechlin et al., 2025c)

Additional published material related to this thesis can be found in the following conference proceedings:

Koechlin, G., Bovera, F., & Secchi, P. (2025). Forecasting electricity spot market merit-order curves with functional time series modeling. *21st International Conference on the European Energy Market (EEM)*, Lisbon, Portugal. (Koechlin et al., 2025b)

Koechlin, G., Bovera, F., & Secchi, P. (2025). Deconvolution of the electrical load curve using blind source separation. *Methodological and Applied Statistics and Demography IV, SIS 2024*. (Koechlin et al., 2025a)

Koechlin, G., Bovera, F., & Secchi, P. (2024). Longitudinal study of dispatchable power units trading strategies on electricity spot markets in Italy 2024 *20th International Conference on the European Energy Market (EEM)*, Istanbul, Turkiye. (Koechlin et al., 2024)

and oral contribution:

Koechlin, G., Bovera, F., & Secchi, P. (2025) Forecasting electricity spot market merit-order curves with functional time series modeling. *14th International Ruhr Energy Conference (INREC)*, Essen, Germany. (Koechlin et al., 2025b)

Code repositories:

-  **load-bss**: Python code to produce results of Chapter 1.
-  **moc-forecast**: Python code to produce results of Chapter 2.
-  **asm-forecast**: Python code to produce results of Chapter 3.

Introduction

Over the past decades, liberalized electricity wholesale markets have been progressively adopted worldwide, supporting a significant economic growth characterized by a request of cost-efficiency and an increasing power demand. Market liberalization led to the unbundling of the power sector, separating electricity generation and trading – fully exposed to market forces – from the construction and management of power grids, which remained under regulated monopolistic control (IEA, 2005). In regions with developed power systems, electricity production was historically dominated by large, centralized, and fully programmable generation units (GUs), primarily based on nuclear and fossil fuels (Joskow, 2008).

This traditional structure of power systems has been profoundly altered by the rapid diffusion of renewable energy sources (RES), pushed by decarbonization targets such as those established by the Paris Agreement (United Nations, 2015) and, in Europe, by the Green Deal (European Commission, 2019). Nowadays, almost 50% of European electricity production originates from RES with wind and solar – often referred to as variable renewable energy (VRE)¹ – accounting for about half of this share (Ember, 2025). In 2025, electricity generation from VRE exceeded fossil-based production for the first time in the European Union (Petrovitch, 2026). Unlike conventional plants, VRE technologies are characterized by high modularity and intermittent generation, driving the transition from centralized to decentralized power production, with numerous small, geographically dispersed and non-programmable GUs.

The increasing penetration of VRE has introduced a structural and conceptual shift in electricity systems, challenging the long-standing paradigm of demand-driven power supply. As an increasing fraction of generation becomes

¹Note VRE also includes some form of hydropower including run-of-the-river, tidal and wave power, but is largely dominated by solar and wind power.

intermittent and non-dispatchable², the ability of the system to adjust supply to demand is fundamentally reduced. Consequently, demand itself is progressively becoming supply-driven, required to respond to stochastic availability of generation rather than the opposite. This inversion of the traditional supply-demand relationship has elevated the role of electricity storage technologies – such as battery energy storage systems (BESS) and pumped storage hydropower (PSH) – as well as demand-side flexibility mechanisms, as essential tools for system operation and reliability. Concurrently, uncertainty surrounding fossil-fuel supply, as highlighted by recent geopolitical events, has further intensified energy price volatility.

These transformations have made electricity markets substantially more volatile and more difficult to forecast. Yet, paradoxically, they have also coincided with an unprecedented expansion in the availability of real-time market, grid, and exogenous data. In Europe, initiatives such as the ENTSO-E Transparency Platform (ENTSO-E, 2025), the development of open market data repositories by nominated electricity market operators (NEMOs), and the proliferation of high-resolution weather datasets have dramatically increased the volume and granularity of accessible information (Andersen and Anandha, 2025). Combined with advances in computational power & intelligence, this data abundance has opened new possibilities for data-driven monitoring, forecasting, and decision-making tools, which have become essential for all actors of the electric power industry.

In this context, the central objective of this PhD thesis is to develop data-driven electricity market models to better understand and predict market dynamics, define optimal short-term strategies and pave the way for more efficient long-term regulations. The thesis leverages various statistical learning techniques spanning matrix factorization, classification, time series forecasting, and functional data analysis, with particular emphasis on uncertainty quantification and probabilistic prediction. Beyond predictive accuracy, a core focus is placed on model *interpretability* (Molnar, 2020) – either by explicitly modeling underlying market mechanisms or by applying eXplainable AI (XAI) methods to interpret black-box machine-learning models. The proposed methodologies are empirically validated using data from the Italian electricity market (GME, 2026), but are broadly applicable to other European markets due to regulatory harmonization and market coupling, as well as to other liberalized power markets in the rest of the world.

Electricity wholesale markets can be broadly classified into *spot* markets and *forward* markets. Spot markets – on which this thesis focuses – concern the trading of electricity for short-term delivery and consumption, whereas forward mar-

²Dispatchable generation refers to power plants whose output can be increased, decreased, or scheduled on demand by system operators within technical constraints, in order to match expected or real-time electricity demand.

kets involve the exchange of longer-term supply and purchase obligations. While forward dynamics are largely driven by financial considerations, spot market outcomes are primarily determined by short-term power system fundamentals such as demand, generation availability, and transmission constraints (Weron, 2014).

Despite the increasing importance of supply-side uncertainty induced by VRE integration, electricity demand variations remain the dominant driver of spot market outcomes. Thorough analysis of demand patterns is therefore a cornerstone of electricity market modeling. While aggregate system load is routinely monitored and accurately forecast at high temporal resolution by transmission system operators (TSOs), its underlying composition – i.e., the contribution of different consumer categories such as residential, commercial and industrial sectors – remains largely unobserved. Moreover, beyond well-known weather and calendar effects, the main drivers of demand variability have not been extensively investigated³. This limited visibility into demand structure, resulting from the absence of sectoral granularity in grid-scale power consumption data, motivates the development of disaggregation-based load monitoring methods.

Electricity spot markets can be further divided into *commercial* markets and *technical* markets, depending on the nature of the actors involved and the objectives governing market outcomes.

Commercial markets facilitate exchanges between electricity producers, retailers, and traders, and are operated primarily according to economic efficiency criteria. They include *day-ahead* markets, where a single auction determines generation and consumption schedules one day prior to delivery, and *intraday* markets, where multiple auctions and continuous trading platforms allow market participants to adjust their positions closer to real time as demand and VRE generation forecasts are updated.

The majority of electricity traded in Europe is exchanged through day-ahead markets (Lago et al., 2021), making day-ahead prices a fundamental reference for both spot and forward market activity. In most European markets, day-ahead prices are determined through a uniform-price (UP) auction mechanism, where the intersection of aggregated supply and demand merit-order curves defines the market-clearing price. Because of this single-price property, prior research has largely focused on the clearing price forecasting problem while less attention was given to the broader merit order curves forecasting problem, though of high interest for market players (Pelagatti, 2013). In particular, the benefits of merit-order curve modeling for short-term price forecasting have been discussed, but not yet rigorously proven.

In contrast, technical markets govern transactions for the physical management of the power system between the system operator (SO) and dispatchable units, including both GUs and (more recently) consumption units (CUs). Their primary objective is system security rather than economic optimality, and their

³See Chapter 1’s introduction.

design is therefore strongly influenced by network constraints and operational considerations. Technical markets typically include *ancillary services* (AS) markets, which aim to resolve network congestions and procure reserve capacity, and *balancing* markets, which operate in real time to correct residual imbalances between generation and consumption. The AS market often plays a central preliminary role by re-dispatching GUs after commercial markets are cleared, in order to maximize available flexibility to be used in the balancing market to locally adjust power injection and withdrawal levels.

Compared to commercial markets, technical markets are often more country-specific and consequently more challenging to analyze within a unified framework. With the integration of non-dispatchable generation from VRE, the power grid faces increasingly frequent congestions and supply-side imbalances (Wang et al., 2014). As a result, a growing share of power volumes are moving from commercial to technical markets, which are therefore receiving increasing attention. Their outcomes, however, remain difficult to predict for participants due to an evident technical prioritization by SOs.

Outline of the thesis

The thesis contributions are structured around three pillars of market decision-making processes: *monitoring*, *forecasting* and *strategic bidding*.

In **Chapter 1** (*monitoring*), we address the sectoral load monitoring problem by proposing a blind source separation (BSS) framework to disaggregate high-voltage grid load measurements in sectors' contributions. The BSS problem is formalized as a constrained variant of non-negative matrix factorization (NMF) (Lee and Seung, 1999), which is termed linearly-constrained non-negative matrix factorization (LCNMF). We show how LCNMF allows to combine low-frequency disaggregated consumption information with high-frequency aggregated measurements to estimate the separate contributions of the different sectors to the total system load. Applying this method to Italian national load data from 2021 to 2023, we identify characteristic residential, commercial and industrial sectors' load profiles and obtain an accurate sector-wise disaggregation of the national power demand.

In **Chapter 2** (*forecasting*), we address the commercial markets forecasting problem by introducing a general framework for the day-ahead prediction of the supply and demand merit-order curves, from which both point and probabilistic clearing price forecasts can be derived. Specifically, we leverage functional data analysis (FDA) to efficiently represent a pair of supply and demand curves in a low-dimensional vector space and employ regularized vector autoregressive models for their prediction. We conduct a rigorous empirical comparison of price forecasting performance between the proposed *curve-based* model, i.e., derived from predicted merit-order curves, and state-of-the-art *price-based* models that

directly forecast the clearing price, using data from the Italian day-ahead market over the 2023-2024 period.

In **Chapter 3** (*strategic bidding*), we propose a probabilistic classification framework to model AS markets operating under a pay-as-bid (PaB) mechanism and derive an optimal bidding strategy for market participants. Specifically, we use a calibrated random forest classifier to estimate the acceptance probability of individual bids based on price and other explanatory variables including bid characteristics and indicators of grid conditions. Additionally, we employ XAI techniques to interpret the underlying machine learning system and provide insights into actual market clearing mechanisms. We then show how this bid acceptance probability estimator can be used by market participants to optimally price their bids and maximize expected marginal profit.

In the **Conclusion**, finally, we summarize the main contributions, discuss the limitations of the work, and outline directions for future research.

CHAPTER 1

Blind source separation for sectoral demand monitoring

In this chapter, we propose a blind source separation framework to disaggregate open-access high-voltage grid load measurements into sectoral contributions. The approach relies on a constrained variant of non-negative matrix factorization, termed linearly-constrained non-negative matrix factorization (LCNMF), which allows prior information to be incorporated as linear constraints on the factor matrices, thereby providing weak supervision of the separation process. The framework is evaluated using Italian national load data from 2021 to 2023.

1.1 Introduction

1.1.1 Motivations

Monitoring and understanding electricity consumption dynamics is essential for power system operation. At the national level, electricity consumption (and generation) is typically monitored by the TSO both at the level of geographical zones, such as states, regions or power zones, and across different sectors often grouped in three or four categories: industry, commercial & public services (or simply *commercial* or *services*), residential (or *households*) and sometimes agriculture (IEA, 2021; Mor et al., 2021). While zone-level consumption is di-

rectly accessible to the TSO at high temporal resolution via measurements on the high-voltage network, sector-level consumption is considerably more difficult to monitor. It must be retrieved from the medium- and low-voltage networks,¹ which fall outside the TSO’s scope and are managed by numerous distribution system operators (DSOs). This process thus involves aggregating numerous building-level load measurements, which are not centrally collected. As a result, sector-specific consumption data are often available only at very poor temporal resolution, mostly on an annual basis, and with significant reporting delays (Voulis et al., 2018).

However, estimating sector-specific consumption at finer temporal resolutions, such as monthly or daily, or even directly monitoring the load at the sector level, is of particular interest to TSOs and other power system actors for two reasons:

First, it obviously provides precious knowledge on sectors themselves. For instance, global sectors’ typical load profiles and granular consumption can be used for the calibration and validation of building stock energy models, especially in top-down approaches (Langevin et al., 2020). It is also valuable for demand-side management as it allows to understand load composition during peak hours – which may vary depending on the type of day or the season – and therefore help designing efficient sector-specific demand response programs (Wang et al., 2015). Real-time consumption monitoring can also help rapidly detect and quantify how the different sectors respond to economic shocks (e.g., energy price spikes) (Møller Andersen et al., 2024; Thoma, 2004), policy interventions (e.g., lockdowns) (Wang et al., 2021; Krarti and Aldubyan, 2021; van Zoest et al., 2023) or extreme weather events (e.g., heat waves) (Zhang et al., 2022), which is increasingly important in a context of rising demand-side and generation flexibility needs. Past and current efforts on sectors’ power consumption monitoring rely on small consumer samples, not always representative nor accessible (Voulis et al., 2018; Gerossier et al., 2017), and only available a posteriori. Finally, sector-level demand at higher frequency is obviously not only useful to the power sector but also to economic institutions and policymakers to monitor the national economic activity (Kraft and Kraft, 1978) with more precision.

Second, it can enhance aggregate load modeling and forecasting (Andersen et al., 2013a,b; Diaz-Iglesias et al., 2025). A significant portion of the variability in the national load curve arises from the dynamic contributions of different sectors (Andersen et al., 2013a). For example, weekday load profiles are typically flatter than those on weekends due to a stronger industry contribution, whose consumption pattern is rather stable over the day. It may also improve the modeling of weather or calendar effects on the aggregate load curve, as different load components (i.e. sectors) may respond differently to specific conditions.

¹With the exception of a few industrial consumers directly connected to the high-voltage grid.

For instance, the households and services sectors may show higher temperature sensitivity in summer due to widespread air conditioning usage, compared to the industrial sector (Pagliarini et al., 2019). Additionally, the long-term evolution of sector-specific load profiles may be easier to model and project in scenarios – especially in the context of electrification – than the aggregate national load (Andersen et al., 2013b, 2014; Møller Andersen et al., 2024).

1.1.2 Objectives of the Work

In this work, we demonstrate that, despite the absence of actual measurements, it is possible to monitor the hourly consumption of the residential, commercial and industrial sectors in real-time from the aggregate open-access TSO load measurements, by applying signal separation techniques. We formulate this unsupervised load disaggregation task as a blind source separation (BSS) problem, showing that the incorporation of sector-level information usually available – such as annual demand and monthly consumption proxies – as boundary conditions allows to identify characteristic sectors’ load profiles and obtain an accurate sector-wise disaggregation. The BSS problem is formalized as a constrained variant of non-negative matrix factorization (NMF) (Lee and Seung, 1999), which is termed linearly-constrained non-negative matrix factorization (LCNMF), and solved using a modified version of the multiplicative updates (MU) algorithm of Lee and Seung (2000). We apply this method to Italian national load data from 2021-2022 and show that the identified sector profiles for this period allow to accurately disaggregate in real-time new load curves for which no sector-level information is available, using 2023 data for validation.

The remainder of this chapter is structured as follows. Section 1.2 provides a detailed literature review and explains our contributions. Section 1.3 details the methodology, beginning with a description of the data sources (1.3.1) and a qualitative introduction to the problem (1.3.2) before presenting the mathematical and computational details (1.3.3 and 1.3.4). Section 1.4 presents the results, and Section 1.5 concludes with a discussion and final remarks.

1.2 Literature review

Statistical modeling and analysis of load curves is a major topic in power systems. A literature overview suggests that most works belong to one of the three following categories:

1. **Load forecasting:** The problem is to predict future load at various network scales from past measurements and influencing factors such as weather and calendar effects. Though the results of the present work may be relevant for such purposes, the problem we solve is not a forecasting one and the corresponding literature is therefore not further presented here.

2. **Load profiling/classification:** This category regards the analysis of typical patterns and similarities between load curves observed in different locations, for different day types, seasons or categories of consumers. The comparison with such works is relevant since our method involves identifying sector-specific load profiles.
3. **Load disaggregation:** We refer to load disaggregation as the task of separating unobserved devices, end-users or consumer groups contributions to observed aggregated load measurements. This is precisely the problem of interest. Since nearly all load disaggregation works are limited to device-level decomposition of buildings consumption, we selected those whose methodological framework is close to ours and this part of the review is essentially on the technical side.

In addition, we demonstrate that this work brings a genuine methodological novelty by extending the discussion to all contributions that have proposed constrained non-negative matrix factorization methods, regardless of the application.

1.2.1 Load profiling and classification

In load profiling and classification, a distinction can be made according to whether load measurements or models corresponds to end-users or an aggregated level (such as district or regional/national level).

At the end-user level, typical residential, commercial and industrial load profiles can be inferred by averaging individual consumers' load curves within each sector. This is typically done with a manual weekday/weekend, season, region and household characteristics stratification or in a more data-driven fashion with clustering techniques. As highlighted by Voulis et al. (2018), thanks to the large diffusion of smart meters, the residential sector has received most attention in the literature (Paatero and Lund, 2006; Räsänen et al., 2010; Teeraratkul et al., 2018; Andersen et al., 2021; Michalakopoulos et al., 2024; Jin and Sanders, 2024), while fewer works studied commercial and industrial consumers (Jardini et al., 2000; Bellinguer et al., 2023). Our study, instead, considers aggregate country-level profiles and their temporal variations rather than within-sector heterogeneity.

Regarding the aggregate level, most works use a so-called bottom-up approach (Langevin et al., 2020): The load profile of the aggregation level is built by aggregating load profiles of different consumer categories with weights reflecting the local consumer mix. For instance, Voulis et al. (2018) builds neighborhood, district and municipality-level load profiles, which are shown to be well-described by a three-level "residential", "business", and "mixed" classification. Their results show that at aggregate levels, only global sector categorization really makes the difference, rather than finer ones. Their study however lacks the inclusion of the industrial component, which is instead present in our work.

Andersen et al. (2013a) aggregates load curves of 4500 end-users in Denmark to derive national profiles for residential, commercial, public services, industry and agriculture sectors, finding strong weekly and seasonal variation except for industry. Later work by Møller Andersen et al. (2024) shows that sector profiles remain relatively stable over the 2019-2022 period. Finally, Andersen et al. (2013b, 2014) adopt more of a top-down approach as they estimate the consumer mix in local areas in Denmark by combining an end-users load profiles dataset and transformer station aggregate measurements. Note that, similarly to Voulis et al. (2018), they rely on the assumption that load profiles of specific consumer categories are constant across locations and, differently from us, do not properly model sectors' load profiles variations over time. Moving to the highest aggregation level, Behm et al. (2020) proposed a methodology to build weather-dependent load profiles for European countries. They however do not address the weather-independent temporal variations of load profiles and sector-specific consumption patterns.

1.2.2 Load disaggregation

Prior research on load profiling and classification has consistently shown that aggregate curves from the substation to the national level contain identifiable sector components. However, while building-scale load disaggregation is a flourishing field of study, which we will discuss first, disaggregation at higher grid scales remains largely unexplored. We will review the few existing works on this topic in a subsequent section.

Non-intrusive load monitoring

First introduced in Hart (1992), building-scale load disaggregation is referred to as non-intrusive load monitoring (NILM) or sometimes energy disaggregation. The purpose of NILM is to monitor a building's load at the appliance level without the necessity of installing dedicated sensors on every single device – which in this case is referred to as intrusive load monitoring (ILM) – but with a unique building-level energy meter, reducing therefore costs and addressing privacy issues. While this work is not about NILM, it's actually solving a similar load disaggregation problem at a higher level of the network, with the difference that it separates consumer categories rather than individual appliances. Because of this conceptual similarity, we believe it's relevant to compare our methodological approach to established NILM techniques.

According to a recent review by Schirmer and Mporas (2023), NILM methods fall into three primary families: machine learning, pattern matching, and source separation. It's important to note that most of these approaches, especially those in the first two families, are supervised, meaning they require access to ground truth, appliance-level measurements to which a learning procedure is applied. In contrast, our method operates in an unsupervised fashion, directly estimating

sector load profiles from a composite signal without needing any ground truth measurements.

In the source separation methods, three different categories are distinguished: independent component analysis (ICA) (Zhu and Lu, 2014; Semwal et al., 2013; Herath et al., 2023), non-negative matrix or tensor factorization (NMF/NTF) (Figueiredo et al., 2014; Matsumoto et al., 2016; Rahimpour et al., 2017; Miyasawa et al., 2019; Zarabie and Das, 2019) and sparse component analysis (SCA) (Singh and Majumdar, 2019; Piga et al., 2016; Kolter et al., 2010). All three methods have in common that they solve a low-rank matrix (or tensor) approximation problem, differing however in the assumptions they rely upon and the factor matrices properties they look for. A key point is that they are in practice unsuccessful when they are used in a fully blind, i.e. unsupervised, setting, as in this case there is often non-uniqueness of the solution and high dependence on the initialization strategy. In fact, Schirmer and Mporas (2023) states that source separation methods developed for NILM, despite being originally thought as unsupervised, are actually "semi-unsupervised" learning procedures, as they all require some degree of prior knowledge which can range from ground truth source signal data to little contextual or structural information on the sources or the mixture.

Notably, both Kolter et al. (2010) (SCA-based) and Rahimpour et al. (2017) (NMF-based) rely on training data consisting of individual appliance-level measurements. Their methods involve modeling each source load (appliance) using basis functions – predefined in Rahimpour et al. (2017) and learned in Kolter et al. (2010) – which are then used to disaggregate the aggregate load. Interestingly, Rahimpour et al. (2017) formulates an NMF problem with a linear constraint on the weights matrix, similar in form to ours. However, they treat this as a non-negative least squares (NNLS) problem, updating only the weights while keeping the basis fixed. Finally, Wytock and Kolter (2014) propose a contextually-supervised source separation framework. They incorporate prior information about appliances load via penalty terms, allowing for weak supervision in the absence of full training data. However, their paper does not provide details on how the resulting penalized matrix factorization problem is solved algorithmically.

Beyond the building level

To the best of our knowledge, only two works tackled a load disaggregation problem beyond the building scale, among which only one concerning a disaggregation in consumer categories:

- A recent study² by Dampeyrou et al. (2024) proposing an unsupervised method for separating the weather- and calendar-dependent components

²In which the authors confirm the literature gap we identify.

of the national load curve. Similarly to us, they don't rely on the observation of disaggregated measurements for model training. Nevertheless, their methodological framework does not resemble ours as they estimate the separate components as non-linear functions of weather and calendar variables with a deep-learning method. The authors themselves note a key limitation: the resulting decomposition lacks meaningfulness and practical utility. In contrast, our approach's decomposition is both identifiable and meaningful, as prior research on load profiling showed.

- Another, more relevant work is Gerossier et al. (2017), proposing a source separation framework to estimate consumer category profiles from transformer substation load data. Their method uses a customer database containing annual consumption for each consumer category, which is available for every area covered by the substations. The main difference on the application side regards the fact that they work at the low-medium voltage vs high-voltage network scale in our case. On the methodological side, since they already know the consumer categories weights in the aggregate curves, they solve a non-blind source separation problem – which boils down to NNLS – for which a global minimum exists. Our source separation problem instead is truly blind, as consumer categories weights are unknown and must be estimated simultaneously with their load profiles. This optimization problem is more complex as it is non-convex and therefore no global minimum can be found (Gillis, 2020). Though valuable for the specific problem tackled by the authors, their study has several limitations affecting results validity and its generalizability to similar yet different load disaggregation problems like ours:

1. It relies on a non-typical customer database – belonging to the French DSO – providing very local consumer information which may not be available for other applications. Our method, conversely, uses widely available country-level annual and monthly consumption data.
2. As noted by the authors in the conclusion, they do not have an automatic way to determine the number of source profiles that can be extracted from the aggregate curves. In our case, though we look for a three-level decomposition, we model the aggregate load with a higher number of source profiles, directly inferred from the data.
3. By assigning a single and fixed profile to each consumer category, they assume that average consumption behaviors at this geographical scale are temporally and spatially invariant. However, Andersen et al. (2014) highlights that, while this is a reasonable assumption for households, the industry profile is very likely to differ between areas depending on the local industrial mix. Moreover, the results of Andersen et al. (2013a) clearly show that households and services

profiles change over the year. The framework of the present work, instead, enables flexible profiles that can account for intra-week and seasonal variations.

4. They only test the ability of their model to accurately represent a new transformer aggregate load curve but they don't validate the accuracy of their decomposition. This is obviously due to the lack of ground truth measurements, like us, but in our case we validate the disaggregated estimates at the month level against sector-level consumption indicators and compare the load profiles with those reported in the literature.

1.2.3 Constrained NMF

Beyond contributions specifically addressing load disaggregation problems, several *constrained* NMF (CNMF) methods have been proposed in the literature, most of which extend the original multiplicative updates (MU) algorithm introduced in Lee and Seung (2000). A recent review can be found in Guo et al. (2024). However, most papers tackle highly specific cases which are not generalizable to our problem. For instance, Liu et al. (2006) and Liu et al. (2012) introduce a so-called CNMF method which is in reality more akin to nonnegative matrix *tri*-factorization (NMTF).

Several contributions – mainly in the hyperspectral unmixing literature – focus on *sparse* or *smooth* NMF (which can be both considered as belonging to CNMF, c.f. Guo et al. (2024)), where a L^p -norm or a more elaborate regularization term is incorporated into the objective function to promote sparsity or smoothness in the factor matrices (Pauca et al., 2006; Jia and Qian, 2009; Liu et al., 2011; Chen and Cichocki, 2004). In a different direction, Hosseinzadeh Aghdam (2022) proposes a specific CNMF method – classified as NMF *in manifolds* in Guo et al. (2024) – aiming at preserving pairwise similarities between observations in the reduced space. Finally, Peng et al. (2012) provide theoretical grounding for our LCNMF method by showing that the MU algorithm can be trivially tweaked to solve NMF with additional linear constraints on the factor matrices, although the modified MU are not explicitly detailed in their paper.

Consequently, this work introduces a novel algorithmic framework for solving a general NMF problem with linear constraints on the factor matrices.

1.2.4 Research gap and main contributions

All aforementioned studies on aggregate load modeling and profiling rely on the availability of large, costly, end-users load samples or atypical consumer information datasets, making them poorly generalizable. Besides, most of them put efforts in the definition of granular load-based consumer classifications while failing to account for the temporal variations of daily profiles. Moreover, none of

them address the problem of monitoring the total consumption of an entire consumer group, such as households, services or industry at the national level and at high frequency. Despite studies showing that highly aggregate load curves variations reflect a dynamic composition of distinctive sectors consumption profiles, no work explored the potential of load disaggregation methods for large-scale sectors consumption monitoring. Finally, no existing methodological framework makes it possible to solve a general NMF problem with linear constraints, such as the one considered in this paper. To address these limitations, this work makes the following contributions:

- (i) We are the first to propose a transmission-level load disaggregation method for monitoring sectoral power demand from national high-voltage grid load measurements.
- (ii) We introduce a novel top-down approach for identifying country-level sectors daily load profiles and their temporal variations which does not require the availability of consumer-level samples and relies on open-access TSO data only
- (iii) We provide an algorithmic framework to solve a general NMF problem with linear constraints on the factor matrices, which can be applied to unsupervised load disaggregation problems at any scale of the power grid.

1.3 Methodology

Fig. 1.1 provides an input-output diagram of the load decomposition process.

1.3.1 Data

The consumption data we aim to decompose come from the Italian TSO's (Terna) hourly measurements of the national load in Italy between 2021 and 2023 (Terna, 2024e). Additionally, available information on sectors' electricity consumption is used. Specifically, we use two data sources:

- (i) The yearly consumption per sector from Terna's annual consumption report (Terna, 2024a), which we refer to as the *annual sector demand* (ASD).
- (ii) Three unitless indicators (Terna, 2024b,c,d) describing the monthly evolution of the household, industry³ and services sectors consumption, designated as *monthly sectoral indicators* (MSI).

Table 1.1 shows a tabular view of the three different data sources and pre-processing steps applied.

³Which also includes the agricultural sector. Given that the agricultural sector's weight in total consumption is negligible compared to the three other sectors, and since we have no suitable indicator to act as a proxy for its monthly consumption, we have chosen to merge this sector with the industrial one (as we believe their consumption patterns are most similar).

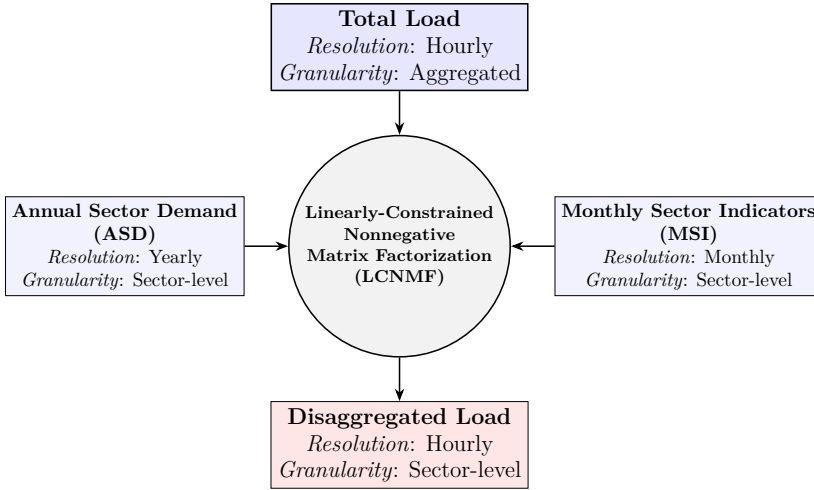


Figure 1.1: Diagram showing the load decomposition process. The input data are represented in blue rectangular boxes (lighter blue for secondary inputs) while the output of the LCNMF disaggregation procedure is represented as a red rectangular box.

Data	Resolution	Unit	Source	Preprocessing
Load	Hourly	MW	Terna (2024e)	Daylight saving time (DST) adjustment ⁴
ASD	Yearly	MWh	Terna (2024a)	Merge of Agriculture into Industry and rescaling to match load data ⁵
MSI	Monthly	None	Terna (2024b,c,d)	None

Table 1.1: Data sources and preprocessing steps

1.3.2 Qualitative description

Blind source separation (BSS) is a technique used to separate a set of mixed signals into their original, independent source components, without prior knowledge of the mixing process or the source signals. BSS can be best understood in the context of audio signal decomposition, in particular with the cocktail party problem (CPP), pictured in Fig. 1.2. Imagine a cocktail party where several conversations or devices generating some audio signal are happening si-

⁴The second occurrence of the duplicate timestamp (02:00) on hour gain days is deleted and the missing timestamp (02:00) on hour loss days is added with the load value observed at 01:00.

⁵Note that the decomposition, originating from load data, will inherently overestimate the sectors' consumption as it disaggregates a quantity that includes network losses, being computed from the generation side of the grid. To address this issue, one can either subtract the losses to the load data or add the losses proportionally to the ASD, before running the separation process.

multaneously. With multiple microphones capturing the overlapping voices, the challenge is to isolate the speech of just one person. BSS techniques analyze the combined audio signals to extract the different voices of individual speakers. In practice, they rely on the fact that (i) the microphones are positioned at different locations in the room such that each microphone receives a signal with some voices being perceived louder and some other quieter (ii) the individuals have distinct voices, which exhibit some regularity.

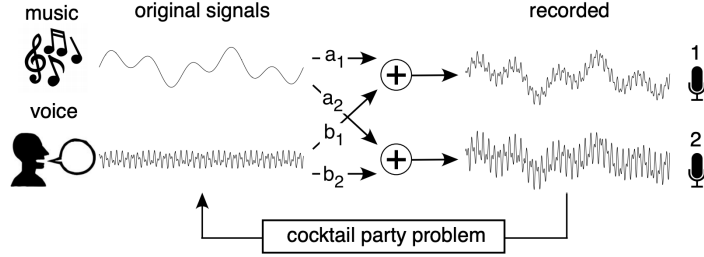


Figure 1.2: *Illustration of the cocktail party problem. Source: Shlens (2014)*

In our load disaggregation problem, we observe each day a load curve having a specific shape resulting from the mixture of the sectors' load profiles, analogous to the "voices" in the CPP (Fig. 1.3). We aim to leverage the fact that some sectors "speak louder" on certain days to identify their load profile, estimate their contribution to the overall mixture, and reconstruct their load signal. However, the assumption in point (ii) above – where speakers' voices are different and regular – poses a limitation in our context, as it assumes sectors have distinct and time-invariant load profiles. While it is reasonable to expect that the typical load curve of one sector differs from another, it is clear that these profiles can vary based on factors such as the day of the week (weekday vs. weekend) or the season (winter vs. summer) as shown by Andersen et al. (2013a). Therefore we cannot consider the sectors' load profiles as fixed and constant over time. To address this limitation, we assume that each sector's load profile is itself a mixture of multiple underlying sources, which may or may not be directly interpretable, and that these sources meet the conditions for separation. Note that the number of sources present in the aggregated signal can be inferred from the data with dimensionality reduction techniques such as principal component analysis (PCA). By identifying the contribution of these sources to the overall load signal, we can estimate the contribution of each sector by aggregating the corresponding sources.

However, without incorporating prior information on the sectors profiles or weights, the likelihood of finding the desired decomposition – where the sources are precisely the "voices" of our sectors – is quite low. Fortunately, we do have valuable data that can guide the separation process: the ASD and MSI. The



Figure 1.3: *Illustration of the load disaggregation problem. In this example, each sector is associated to one source only.*

annual sector demand (ASD) provides information on sectors’ total annual consumption, while the monthly sectors’ indicators (MSI) indicates how this consumption is distributed over the 12 months. Hence an estimate of each sector’s monthly consumption can be computed, and this can be compared to the results of the decomposition, aggregated at the month level. By constraining the decomposition to produce monthly consumption estimates that align with these known values, we can identify sources that are hierarchically linked to the sectors and hence reconstruct the sectors’ load.

Once we have identified the sources, any *new* set of daily load curves can be decomposed into contributions from these sources without requiring any further information about sectors’ consumption (which, by contrast, is necessary for the initial source identification). In other words, we can decompose new load data using only the load curves themselves (and the identified sources, obviously). The source separation process becomes “non-blind” because we only need to estimate the weights of the observed mixtures. To validate this nowcasting predictive approach, we decompose the load curves for the year 2023 using the sources estimated on 2021-2022, without using any further sector-level data. We then compute the predicted monthly consumption and compare it to the MSI for verification.

The next two subsections dive into the mathematical details of what was just described.

1.3.3 Model

Let $L_i : [0, T] \rightarrow \mathbb{R}^+$ represent the system's⁶ load as a function of time during a day $i \in \{1, \dots, n\}$. The total energy consumption e_i during day i is given by:

$$e_i = \int_0^T L_i(t) dt = \|L_i\|_1$$

We define $X_i : [0, T] \rightarrow \mathbb{R}^+$ as the load shape curve observed on day i :

$$L_i = e_i \cdot X_i$$

In other words, X_i is the normalised load curve of day i such that it corresponds to a total consumption of 1. For notation convenience, we may assimilate X_i to an element of $\mathbb{R}_+^p$, corresponding to evaluations of the underlying function at p regularly-spaced time instants spanning the interval $[0, T]$. For example, $p = 25$ corresponds to evaluations at 00:00, 01:00, $\dots$, 24:00.

Each curve X_i is assumed to be a mixture of K sources $S_1, \dots, S_K$ with $K \ll n$:

$$\begin{cases} X_1 &= c_{11}S_1 + c_{12}S_2 + \dots + c_{1K}S_K + \epsilon_1 \\ X_2 &= c_{21}S_1 + c_{22}S_2 + \dots + c_{2K}S_K + \epsilon_2 \\ &\vdots \\ X_n &= c_{n1}S_1 + c_{n2}S_2 + \dots + c_{nK}S_K + \epsilon_n \end{cases} \quad (1.1)$$

where for $i \in \{1, \dots, n\}$ and $k \in \{1, \dots, K\}$, the coefficient c_{ik} represents the *concentration* of source k in the load (shape) curve X_i and $\epsilon_1, \dots, \epsilon_n$ are additional noise terms. The profiles $S_1, \dots, S_K$ and the concentrations c_{ik} 's are constrained to be positive. Additionally, we impose that $S_1, \dots, S_K$ integrate to 1 (as the X_i 's) and thus $c_{i1} + \dots + c_{iK} = 1$, for all $i \in \{1, \dots, n\}$. The system of equations (1.1) can be re-written in a matrix form:

$$X \approx CS \quad (1.2)$$

where X is the $n \times p$ load curves matrix, C the $n \times K$ concentrations matrix and S the $K \times p$ sources matrix:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1K} \\ c_{21} & c_{22} & \dots & c_{2K} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nK} \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_K \end{bmatrix}.$$

The load can hence be decomposed by finding a pair of matrices $\hat{C}$ and $\hat{S}$ whose product $\hat{C}\hat{S}$ best approximates the data X . The estimated consumption $\hat{e}_{ik}$ of source k during day i is then deduced by multiplying the estimated

⁶The national power grid in our case

concentration $\hat{c}_{ik}$ with the total consumption e_i :

$$\hat{e}_{ik} = \hat{c}_{ik} e_i$$

K can be chosen by estimating the underlying dimension d of the space in which the load curves X are embedded, e.g. with functional principal component analysis (FPCA) (Ramsay and Silverman, 2005). Given that the fitted curves CS are convex linear combinations of the sources and live therefore in a $(K-1)$ -dimensional linear subspace of $\mathbb{R}_+^p$, we must choose $K = d + 1$.

The low-rank matrix approximation problem of equation (1.2), when both matrices are constrained to have all their coefficients positive, is known as *Non-negative Matrix Factorization* (NMF), with the difference that we have an additional constraint due to the fact that the profiles contained in S must integrate to 1. Moreover, as stated in the previous section, we would like to add a constraint on the solution such that it is consistent with the ASD-MSI sectors' monthly consumption estimates. To formalise this, let $E = [e_1 \ e_2 \ \dots \ e_n]'$ be the n -dimensional vector of daily consumption, for which we consider a block-form notation in which the vector elements are grouped by month:

$$E = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_m \end{bmatrix}$$

where m is the number of months in the period considered. We additionally define a $m \times n$ matrix B and re-write the $n \times K$ matrix C of concentrations in a similar block-form manner:

$$B = \begin{bmatrix} E'_1 & 0 & \dots & 0 \\ 0 & E'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & E'_m \end{bmatrix} \text{ and } C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_m \end{bmatrix}$$

Multiplying the two matrices gives us the $m \times K$ matrix of the monthly consumption of the K sources:

$$BC = \begin{bmatrix} E'_1 C_1 \\ E'_2 C_2 \\ \vdots \\ E'_m C_m \end{bmatrix}$$

The problem now concerns mapping the K sources to the g sectors. We may assign more sources to sectors where we expect the characteristic load profile to vary more significantly throughout the year. We formalise this mapping by

defining a surjection $\sigma : \{1, \dots, K\} \rightarrow \{1, \dots, g\}$ and the associated $K \times g$ matrix A such that, for source $i \in \{1, \dots, K\}$ and sector $j \in \{1, \dots, g\}$:

$$A_{i,j} = \begin{cases} 1 & \text{if } j = \sigma(i) \quad \text{i.e. if } i \text{ is a source of sector } j \\ 0 & \text{otherwise} \end{cases}$$

For instance, in the case $K = 5$ and $g = 3$, attributing source 1 to sector 1, sources 2 and 3 to sector 2, and sources 4 and 5 to sector 3, we have:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

We can finally obtain the $m \times g$ matrix Z of the estimated monthly consumption per sector:

$$Z = BCA$$

On the other side, let $Y = [Y_1 \dots Y_g]$ be the $m \times g$ matrix whose columns Y_j 's represent the MSI, which, as highlighted in subsection 1.3.2, only carry information on the *shape* that the columns of Z , seen as monthly time series, must assume.

Intuitively, we would like to drive the decomposition to yield a Z matrix with columns as *correlated* as possible with those of Y , regardless of the difference in scale. However, the introduction of a penalty term based on the Pearson correlation coefficient – or other relevant similarity measure quantifying statistical dependence – between Z_j and Y_j in the loss function may be computationally intractable. To address this issue, our approach is to use the ASD to *rescale* the matrix Y so that it can directly be compared to Z . More specifically, let w_j , for $j \in \{1, \dots, g\}$, be the total consumption given by the ASD of sector j over the period considered (which must necessarily be composed of entire years). Note that we should have $w_1 + \dots + w_g = \|E\|_1$. Performing the following scaling operation, for all $j \in \{1, \dots, g\}$:

$$Y_j \leftarrow \frac{Y_j}{\|Y_j\|_1} w_j$$

Y is now exactly an estimate of what Z should be. Note that the boundary condition $Y_1 + \dots + Y_g = [\|E_1\|_1, \|E_2\|_1, \dots, \|E_m\|_1]'$, i.e. summing the sectors' consumption should give the total consumption at the month level, may not be fully satisfied but we can similarly rescale Y row-wise to correct this effect.

We can write down the final optimisation problem:

$$\begin{aligned}
 (\hat{C}, \hat{S}) = & \arg \min_{C \in \mathbb{R}_+^{n \times K}, S \in \mathbb{R}_+^{K \times p}} \text{dist}(X, CS)^2 \\
 \text{s.t. } & BCA = Y \\
 & S \mathbf{1}_p = \mathbf{1}_K
 \end{aligned} \tag{1.3}$$

where $\text{dist}()$ is some distance measure on $\mathbb{R}_+^{n \times p}$ and $\mathbf{1}_r$ is the r -dimensional vector of ones.

1.3.4 Solution method

The multiplicative updates (MU) algorithm introduced by Lee and Seung (2000) allows to find a local minimum to the ordinary NMF problem, formulated with a distance measure being the metric induced by the Frobenius norm (or another distance measure similar to the Kullback-Leibler divergence). Essentially, from arbitrary initial values, the matrices C and S get alternatively updated by multiplying their current value by a factor that depends on the quality of the approximation, until convergence.

We propose a modified version of the MU algorithm to solve the *linearly constrained* non-negative matrix factorization (LCNMF) problem, i.e., NMF with additional equality constraints that take a linear form in the factor matrices, generalizing (1.3). Indeed, given any $B \in \mathbb{R}_+^{m \times n}$, $A \in \mathbb{R}_+^{K \times g}$, $Y \in \mathbb{R}_+^{m \times g}$, $F \in \mathbb{R}_+^{l \times K}$, $D \in \mathbb{R}_+^{p \times q}$, $Z \in \mathbb{R}_+^{l \times q}$, let

$$\begin{aligned}
 (\hat{C}, \hat{S}) = & \arg \min_{C \in \mathbb{R}_+^{n \times K}, S \in \mathbb{R}_+^{K \times p}} \text{dist}(X, CS)^2 \\
 \text{s.t. } & BCA = Y \\
 & FSD = Z.
 \end{aligned} \tag{1.4}$$

We reformulate (1.4) by choosing the metric induced by the Frobenius norm as the distance measure and by translating the constraints in terms of distance. Considering then its Lagrangian form, we get the following unconstrained loss function to be minimized, with constraints incorporated as penalization terms:

$$\mathcal{L}(C, S) = \|X - CS\|_F^2 + \alpha \|BCA - Y\|_F^2 + \beta \|FSD - Z\|_F^2 \tag{1.5}$$

for some $\alpha, \beta > 0$ penalizing violations of the constraints.

To minimise $\mathcal{L}$, we start from arbitrary initial values and update the factor matrices as follows, until convergence:

$$S \leftarrow S \cdot \frac{C'X + \beta F'ZD'}{C'CS + \beta F'FSDD'}, \quad C \leftarrow C \cdot \frac{XS' + \alpha B'YA'}{CSS' + \alpha B'BCAA'}$$

where $\cdot$ (multiplication) and $\frac{\cdot}{\cdot}$ (division) operations are coefficient-wise. These updates are derived following the same logic as in Lee and Seung (2000), further detailed in Appendix A.1.

A first consideration regards the determination of suitable α and β values as they define the weight that will be given to the terms penalizing the violation of the constraints. Given that the penalization loss terms may not be of the same magnitude as the factorization loss term, their primary purpose is a rescaling one. Analysing jointly the three loss terms at convergence obtained for different values of (α, β) , one can determine suitable values.

A second consideration concerns the initialization strategy, which is obviously a critical point. Our approach involves running the MU algorithm $N = 1000$ times, each time with a different initialization – keeping the same initial value $\left[\frac{1}{p} \quad \dots \quad \frac{1}{p}\right]'$ for the rows of S but obtaining each time a different initial value for C by uniformly sampling its rows on the simplex $\mathcal{S}(K) = \{x \in \mathbb{R}_+^K \mid x_1 + \dots + x_K = 1\}$ – generating a large set of solutions. This set of solutions is then clustered based on the loss function value at convergence and the cluster of solutions associated with minimum value is retained. This allows us to work with an ensemble of N^* solutions that can be combined to obtain more robust estimates and to quantify uncertainty. The ensemble-estimated consumption $\hat{e}_{ik}$ of source k during day i is hence deduced by averaging the individual estimates given by the N^* solutions:

$$\hat{e}_{ik} = \frac{1}{N^*} \sum_{l=1}^{N^*} \hat{e}_{ik}^{(l)} e_i$$

Additionally one can also retrieve source k 's estimated load during day i :

$$\hat{L}_{ik} = \frac{1}{N^*} \sum_{l=1}^{N^*} \hat{e}_{ik}^{(l)} e_i S_k^{(l)}$$

As mentioned in subsection 1.3.2, we can project a new load curve X_0 on the linear space spanned by the solution matrix $\hat{S}$ to estimate the sources concentrations hence sectors' contributions to X_0 . The source separation problem becomes "non-blind" as we only need to estimate the concentrations:

$$\hat{C}_0 = \arg \min_{C \in \mathbb{R}_+^{n \times K}} \|X_0 - C\hat{S}\|_F^2$$

This is a convex optimisation problem for which the global minimum can be found, as long as the sources are linearly independent (i.e. $\hat{S}$ is full rank). In practice, we use the standard MU of Lee and Seung (2000) to solve the problem, updating C only.

1.4 Results

1.4.1 LCNMF solutions analysis and interpretation

We applied the source separation algorithm to 2021 and 2022 data, acting as a training set, to identify the sources and consequently the sectors' profiles.

We hold 2023 data out as a test set on which we estimated the concentrations of the sources identified in the training set, without incorporating the ASD-MSI information. We then compared the sectors' monthly consumption of 2023 estimated from the decomposition with the ASD-MSI target.

Fig. 1.5 shows the scree plot of the FPCA performed on 2021-2022 load curves. We can see that 4 principal components explain $\approx 97\%$ of the variance and appear as a good choice for dimensionality reduction. We can therefore deduce that we should get a satisfying decomposition with $K = 5$ sources. Trials clearly indicate that the optimal sources/sectors mapping is obtained by assigning two sources to households, two sources to services and one source to industry. This finding implies that the households and services daily load profiles are variable while that of industry is fixed, which aligns with the results of Andersen et al. (2013a).

Concerning the penalization parameters, we chose $\alpha = 3 \cdot 10^{-10}$ and $\beta = 1$ as they were providing the best trade-off between matrix factorization quality and compliance with the constraints. Following the $N = 1000$ LCNMF runs, two compact and well-separated clusters of loss values at convergence were observed. We then simply used a threshold of $\mathcal{L} = 0.01$ to retrieve the cluster of $N^* = 271$ optimal solutions (Fig. 1.4). As underlined in subsection 1.3.4, we do not choose a specific solution but we analyse and use the entire set of N^* optimal solutions to get a robust ensemble estimation and quantify uncertainty.

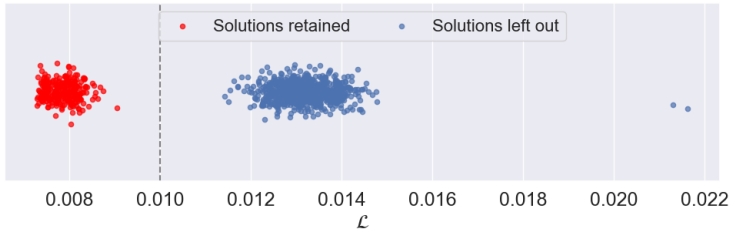


Figure 1.4: Value of the loss $\mathcal{L}$ at convergence for the $N = 1000$ LCNMF runs. The values are vertically jittered for legibility. The grey dashed line indicates the threshold used to isolate the red cluster.

A first consideration regards the quality of the approximation of the load curves that we get with the LCNMF decomposition. Table 1.2 compares the average low-rank matrix approximation quality of LCNMF with the approximation given by the projection of the curves on the space spanned by the first 4 functional principal components (FPCs). The comparison is relevant because both approximation methods rely on a 4-dimensional representation. We observe that LCNMF provides a better fit than FPCA, showing that the approximation quality is more than satisfactory.

$Loss$	$\ A - B\ _1$	$\ A - B\ _F$	$\ A - B\ _\infty$
FPCA	9.2	0.0096	0.0053
LCNMF	8.9	0.0077	0.0051

Table 1.2: *Comparison of the low-rank matrix approximation quality between FPCA and LCNMF*

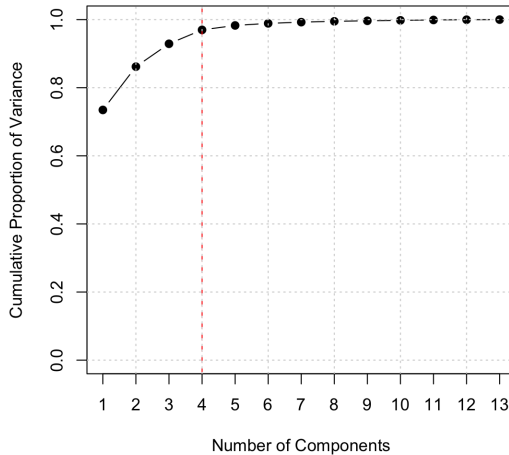


Figure 1.5: *Scree plot of the FPCA run on 2021-2022 load curves*

The extracted sources are shown in Fig. 1.6. We notice considerable variability across the different solutions, except for the source corresponding to the industrial sector. In this case, all solutions converge on the same load profile, marked by a sharp morning ramp beginning at 05:00 and three distinct peaks at 09:00, 16:00, and 20:00. This trimodal pattern is very similar to the profile that Andersen et al. (2013a) and Gerossier et al. (2017) identify for the industry, therefore validating this result. This consistency confirms the idea that a single source is enough to model the industrial sector’s load profile – which is in fact expected to be stable according to the results in Andersen et al. (2013a). Sources mapped to the households sector show a morning ramp that either starts steeply at 05:00 or more gradually at 08:00. The first one displays an evening peak around 18:00-19:00, while the second one shows it around 21:00-22:00. Beyond these few observations, the results are hard to interpret. The sources linked to the services sector are generally flatter than those of the household sector. They typically exhibit a morning ramp either at 05:00 or 07:00, depending on the solution. In particular, for most solutions, S_4 (Services) shows an afternoon peak around 17:00–18:00. S_5 (Services) seems to be characterized, in most cases,

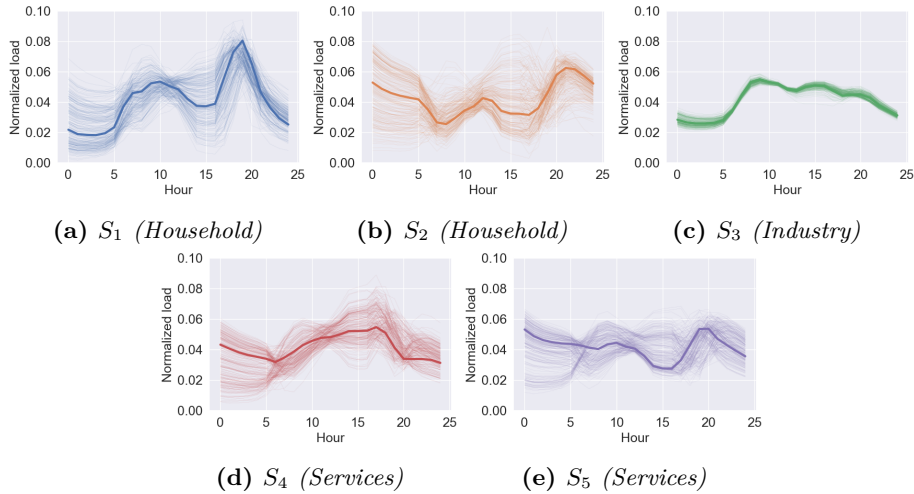


Figure 1.6: Sources corresponding to the $N^* = 271$ retained solutions. Each thin line corresponds to one solution while the thick line, added for visualization purposes, is the geometric medoid of the thin lines, which is the closest observation to the geometric median (in terms of Euclidean distance). Since inside a sector the source number is arbitrary, the sources of household and services were aligned on the order given by the first solution.

by a rather constant regime with an afternoon trough centered on 15:00 and an evening peak located around 20:00.

Overall, except for the industrial sector whose load profile is directly described by S_3 (Industry), the interpretability of the resulting sources for the two other sectors is limited as their load profile is given by a positive linear combination of their sources. For these two sectors, there is in fact a profile specific to each day as the weights of the linear combination are free to vary, causing therefore a modification of the profile whenever one weight increases in proportion to the other. Fig. 1.7 shows the average profile of households and services that we get depending on the type of day – Monday, Working day, Saturday and Holiday⁷ – and the season⁸. It is quite clear that enabling flexible profiles allows us to model weekly and yearly seasonal variations in these two sectors’ daily consumption patterns. Specifically, we can see that the households profile is represented by a double-hump curve with a first small hump at 10:00-11:00 and a second higher around 19:00-20:00. The curve is flatter and has a shallow and late morning ramp during the weekends compared to the weekdays. We can also note that the evening peak is delayed of one hour, occurring more around

⁷Sundays and Bank holidays

⁸The mapping is made at the month level so winter is composed of all days from January to March, spring from April to June, summer from July to September and fall from October to December.

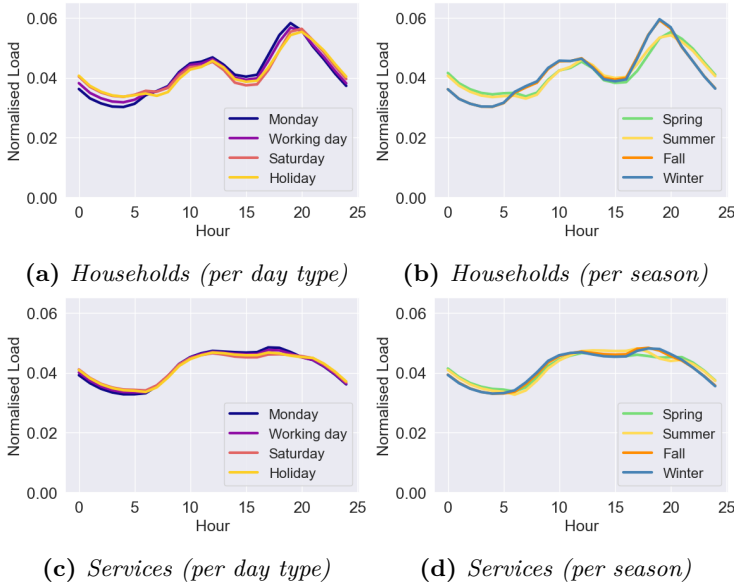


Figure 1.7: Average load profiles found for the households and services sectors depending on the day type and the season. The load profiles were first averaged at the day type and season level for each solution separately after which the average across solutions was retrieved. For legibility, we do not show the result for all individual solutions.

20:00 than 19:00. Concerning the season effect, we observe a fall-winter vs a spring-summer regime. Fall-winter profiles show higher amplitude, earlier and steeper morning ramp, and earlier evening peak around 19:00. On average, the services profiles show a nearly constant activity between 10:00 and 20:00 with no clear morning or evening peak, which is consistent with businesses and offices hours in Italy. While the day type effect is not so clear, we observe a season effect that is similar to that for households, with the difference that summer and spring show slightly diverging behaviors in the afternoon.

These households and services daily consumption patterns globally align with those observed in Andersen et al. (2013a) for Denmark, with some nuances likely due to climate and lifestyle differences.

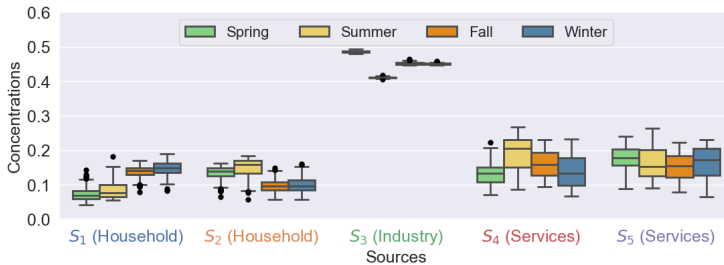
The interpretation of the identified sources is made easier by looking at their concentrations measured every day. Specifically, we can try to understand if a source is more or less active depending on the day type and the period of the year. Fig. 1.8 shows the distribution of the average concentration of each source per type of day and season. We can therefore appraise both the effect of day type/season, and the variability, or disagreement, among the solutions. Overall, the day type seems to have a clear effect on the activity of any source (fig.

1.8a): on average, all five sources contribute equally to holidays while Mondays and working days are clearly dominated by S_3 (Industry). Saturdays lie in-between. Besides, we can note the strong agreement between solutions – which can be called “models” at this point – on the average concentration estimated during working days compared to the other day types. On the contrary we observe considerable disagreement in the disaggregation of Mondays’ load curves, especially for what regards S_3 ’s contribution.

For what regards the season effect, fig. 1.8b highlights a clear opposition spring-summer vs fall-winter conditions for sources associated to households, S_1 and S_2 , as was observed previously in the profiles. This observation supports the existence of two load regimes characterizing household consumption, one under cold temperatures, driven by S_1 , and one under warm temperatures, driven by S_2 . A possible explanation is the quadratic effect of temperature on power consumption in Italy (Pagliarini et al., 2019) with electric heating in winter and air conditioning devices in summer, causing two distinct seasonal shifts in the daily consumption profile. Note this dual effect may not apply to countries where there are just either heating or cooling needs, e.g., Denmark. The season effect is also pronounced on S_3 (Industry), with a lower activity in summer (due to factory closures in August) and a highest activity in spring, indicating therefore that this effect is more a calendar than a climatic one. A summer effect opposed to the spring one is also perceived for S_4 (Services).



(a) Per day type



(b) Per season

Figure 1.8: Distribution of the average concentrations per day type and season across solutions.

1.4.2 Disaggregated consumption

The core output of the process, as indicated in Fig. 1.1, is the disaggregated load signal, with the three separate contributions of the sectors. We first compute the daily load curve of each source, obtained from the shape curve rescaled with the fraction of the daily consumption corresponding to that source, itself obtained by multiplying the concentration with the daily consumption. We then add up the sources curves to get the sectors curves, that we finally concatenate along days to get a continuous signal. Note that the signal presents an inherent discontinuity every day at midnight since the concentration varies non-smoothly from day to day. This discontinuity is obviously a mathematical artifact, which however does not affect much the aggregated daily quantities. Nonetheless, for applications where this discontinuity is deemed to be problematic, it can be overcome by smoothing the concentrations regarded as hourly data over time, keeping daily average concentrations equal to the original ones. In this work, we don't pursue the matter further.

The disaggregated load signal for the first week of February 2022 is shown on Fig. 1.9. Beyond the uncertainty that is quite high for the household and services sectors compared to the industry one, we can observe that the uncertainty is higher during the weekend and even higher on Monday. This confirms the earlier observed fact that Mondays have a different status from the other weekdays.

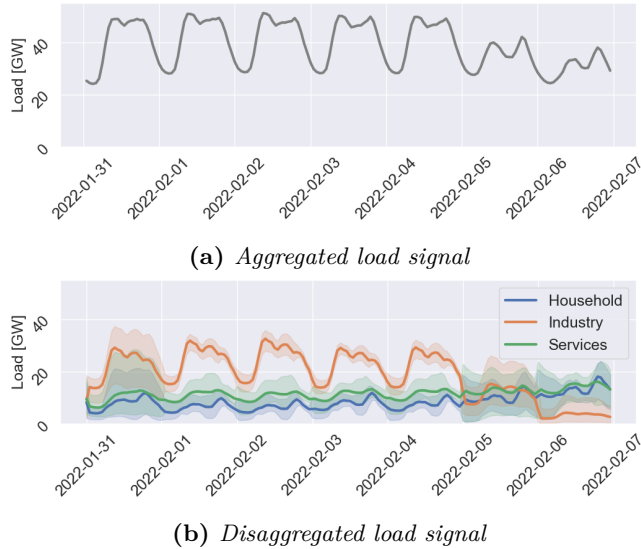


Figure 1.9: Result of the load disaggregation for the week starting on Monday January 31st, 2022. Shaded areas on the bottom panel represent the one-at-a-time 95% prediction intervals given by the 0.025 and 0.975 quantiles of the ensemble estimation.

Unfortunately, we don't have sector-level consumption data to validate the results of the disaggregation at the hourly, daily or weekly level. But moving to the monthly level, we have the MSI to which our estimates can be compared, which are the data, together with the ASD, that were used to guide the separation process. As 2023 was left out in the blind separation process – meaning that the sector-level information corresponding to that year is completely unknown to the model – we can reliably validate our disaggregation procedure by verifying whether monthly sectors' consumption estimates align with the MSI for that year. As a benchmark, we include a naive forecast corresponding to one-year-lagged values and a yearly-seasonal exponential smoothing model: Holt-Winter's (HW) additive method with no trend⁹ (Hyndman and Athanasopoulos, 2021). For the HW model, we forecast the 12 months of 2023 using 2021-2022 data only – similarly to the BSS approach.

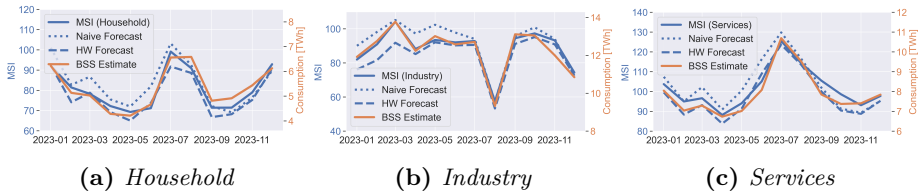


Figure 1.10: Comparison of the BSS monthly sector consumption estimates with the TSO's MSI and its naive and HW forecasts for the test year 2023. Note that the data are not on the same scale and are therefore represented with two different ordinate axis (left/blue for MSI and right/orange for BSS).

	Naive forecast	HW forecast	BSS
Household	0.914	0.959	0.951
Industry	0.962	0.961	0.974
Services	0.912	0.971	0.970

Table 1.3: Pearson's correlation coefficient with MSI for the test year 2023.

The comparison of these out-of-sample predictions with the MSI for 2023 is reported in Fig. 1.10 and Table 1.3. All three estimation methods align well with the observed values. While BSS appears to reproduce the dynamics of the observed series slightly better than the naive forecast, its performance remains comparable to that of HW. These results suggest that the BSS approach produces accurate monthly consumption estimates and is competitive for nowcasting sectors monthly consumption.

⁹Given the series are stationary.

1.5 Discussion and conclusion

As explained in the introductory section, monitoring electricity consumption at the level of residential, commercial and industrial sectors is of considerable interest to the TSO and other power system stakeholders. To date, no solution exists to perform it live, at country scale, high frequency and utilizing open-access national grid load measurements only. In this context, we introduced a novel blind source separation framework to monitor sectors consumption from open-access aggregate grid-scale load measurements collected by the TSO. The proposed LCNMF method is highly generalizable as it allows to solve a load disaggregation problem at any scale of the power network, for which prior information – such as low-frequency disaggregated consumption values or direct information on one or more load components – can be expressed as linear constraints on the factor matrices. While the current formulation addresses linear constraints only, the framework can be easily adapted to support more complex non-linear forms, provided they involve functions differentiable with respect to the factor matrices (see Appendix A.1).

Applied to the Italian grid data between 2021 and 2023, our method produced results revealing established features of household, industry, and services sectors' electricity consumption profiles. There is in fact a strong consistency between the profiles we estimate and those identified in previous studies (Andersen et al., 2013a; Gerossier et al., 2017; Voulis et al., 2018) relying on ground truth data, thereby supporting the validity of our results and the effectiveness of the proposed method. In particular, households consumption shows a daily profile that is dominated by an evening peak and is driven by two distinct regimes, in warmer vs colder seasons. Industrial consumption is particularly low during holiday periods and does not seem to be much affected by seasons. Finally the services sector is found to be especially active during the usual businesses and office hours and shows a singular summer daily load profile, which could be related to the air conditioning devices working at their maximum around 16:00, and only in summer. The results also showed that the industrial sector exhibits a particularly characteristic and consistent daily load profile, which allows for a clear distinction from the other two sectors. In contrast, households and services sectors profiles, whose variability was found to be well captured by linear combinations of two subprofiles each, appear to show a certain degree of similarity that makes them harder to dissociate. Besides, the monthly sectoral consumption estimates derived from disaggregated load were found to closely align with the TSO's official statistics – indices based on consumer samples – and be a competitive approach for nowcasting sectors' monthly consumption. These estimates are currently being assessed by the Italian TSO (Terna) as a potential faster and simpler way to monitor sectors' monthly consumption, complementary to the existing indices.

CHAPTER 2

Forecasting merit-order curves in day-ahead markets

In this chapter, we introduce a functional data analysis (FDA) framework for predicting day-ahead supply and demand merit-order curves, from which both point and probabilistic electricity price forecasts can be derived. We conduct a rigorous empirical comparison of price forecasting performance between the proposed "curve-based" model, i.e., derived from predicted merit-order curves, and state-of-the-art "price-based" models that directly forecast the clearing price, using data from the Italian day-ahead market over the 2023-2024 period.

2.1 Introduction

2.1.1 Context

Most wholesale electricity trading in Europe takes place on day-ahead markets, where producers and retailers exchange power one day before delivery. These markets generally operate under a double uniform-price (UP) auction mechanism, where the intersection of the supply and demand *merit-order curves* sets the market clearing price for all accepted offers and bids. These curves are formed from the aggregation of individual supply offers and demand bids following the merit order, i.e. increasing price for supply and decreasing price for

demand (Fig. 2.1). Offers (or bids) priced below (resp. above) the clearing price are accepted¹ while the remainder are rejected. For market participants, accurately estimating the clearing price is therefore sufficient to anticipate whether an offer/bid will be accepted and at which price.

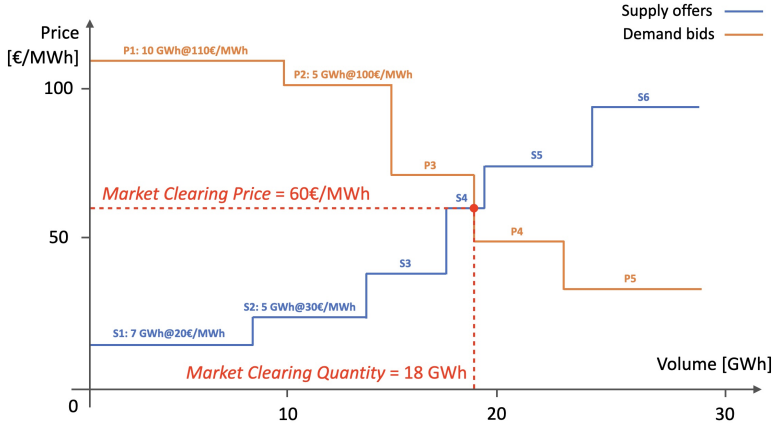


Figure 2.1: Example an electricity spot market clearing from the supply (in blue) and demand (in orange) merit-order curves.

Because of this single-price property, a large body of research has focused on forecasting the day-ahead clearing price itself, a discipline known as *electricity price forecasting* (EPF). The consensus is that data-driven approaches based on historical prices tend to outperform *fundamental* (or structural) ones which try to model exact market mechanisms, especially for short-term hourly price forecasts (Weron, 2014). In particular, a highly competitive day-ahead EPF benchmark based on regularized autoregressive models and feedforward neural networks was consolidated by Lago et al. (2021). This framework is still state-of-the-art and has been widely employed in recent studies (Olivares et al., 2023; Lipiecki et al., 2024; O'Connor et al., 2025; Cerasa and Zani, 2025).

2.1.2 Objectives of the work

Nevertheless, the problem of forecasting the entire merit-order curves underlying price formation is gaining momentum in the academic community (Petropoulos et al., 2022). The seminal work of Ziel and Steinert (2016) demonstrated the capacity of "data-driven fundamental" approaches based on historical auction data and their potential for generating competitive clearing price forecasts. Forecasting entire curves offers several advantages: For participants with market power, it enables precise quantification of how marginal changes in offers may influence

¹The intersection point may split a supply offer or demand bid which in that case would be partially accepted

the clearing price and allows realistic market simulations from which optimal bidding strategies can be derived (Pelagatti, 2013). For price-takers and other actors exposed to the day-ahead price, curve-based models offers a more comprehensive view of market results, enhances the interpretability of price predictions, and may improve the quantification of price predictions' uncertainty². Additionally, complex nonlinear dynamics observed at the price level may become more tractable when modeled at the curve level.

Despite a growing number of contributions, an essential question remains unanswered: Are merit-order curves forecasting models competitive with state-of-the-art price-centric EPF models in terms of clearing price prediction accuracy?

In this context, this work has two main objectives:

1. Introduce a general, simple, and computationally efficient framework for predicting day-ahead supply and demand merit-order curves, from which both point and probabilistic clearing price forecasts can be derived.
2. Rigorously compare the clearing price forecasting performance of the proposed *curve-based* models, i.e. derived from merit-order curves predictions, with that of state-of-the-art *price-based* models, which directly model the clearing price.

Our framework builds on Ziel and Steinert (2016) but eliminates the need for curves discretization by preserving their functional form. Specifically, we leverage functional principal component analysis (FPCA) to efficiently represent a pair of supply and demand curves in a vector space and employ regularized vector autoregressive models inspired by EPF literature for their prediction. We compare this functional approach with the original method of Ziel and Steinert and top-performing price-based point and probabilistic EPF models, on the Italian day-ahead market, during the 2023-2024 period.

The remainder of this paper is structured as follows: section 2.2 reviews previous contributions and presents the knowledge gap addressed by this work, section 2.3 details the forecasting and testing methodology, section 2.4 describes the dataset, section 2.5 presents and discusses the empirical results, and section 2.6 provides concluding remarks.

2.2 Literature review

2.2.1 A thematic review of prior work

Table 2.1 provides an overview of contributions on the topic. Rather than providing a linear and descriptive review, we organize it according to several key aspects that we believe are central to the merit-order curve forecasting problem.

²This latter aspect is investigated in this paper.

Handling the infinite dimensionality

A central aspect of the problem regards the fact that merit-order curves are essentially infinite-dimensional functional objects, in the sense that they represent quantities that vary over a continuum. It is therefore necessary to consider either an appropriate statistical framework, e.g. functional data analysis (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006) or a finite-dimension representation of the curves to which classic multivariate time series methods can be applied.

Within the functional approach, several generalizations of the autoregressive (AR) models (Box et al., 1994) to the functional setting have been used: Functional AR (FAR) models (Canale and Vantini, 2016; Solovieva and Vargiolu, 2021; Li et al., 2025), functional SARMA (Mestre et al., 2020) and nonparametric FAR (NPFAR) models (Shah and Lisi, 2020). Methodological reference on functional time series include Bosq (2000) and Hörmann and Kokoszka (2012). The advantage of functional approaches, beyond elegantly tackling the problem from a mathematical standpoint, is that they avoid the information loss inherent in finite dimension representations. Their main drawback, however, lies in the limited theoretical development and the lack of efficient, easy-to-use, open-access implementations of the models proposed in the literature (Aue et al., 2015).

Aside of the functional framework, there are essentially three strategies to address the dimensionality problem: discretization, parametrization and dimensionality reduction. For all three, the overall workflow is similar: (i) Transform the curves into a finite-dimension vector form, (ii) Forecast this vector using multivariate methods, and (iii) Invert the transformation to recover the predicted curves.

Although any software ultimately relies on discrete representations, we refer to the *discretization* approach when the vector representation consists of curve evaluations over a grid of reasonable size (no more than a few dozen points), enabling multivariate modeling. The functional form is then recovered using a chosen interpolation method. This strategy was adopted in the seminal work of Ziel and Steinert (2016) and its extensions in Kulakov (2020) and Ghelasi and Ziel (2024). *Parametrization* refers to methods that assume a parametric form (e.g. linear, logarithmic or polynomial) for the curves, with the finite-dimensional representation given by the fitted parameters; see for instance Yıldırım et al. (2023) and Ciarreta et al. (2023). Finally, *dimensionality reduction* refers to approaches where the finite-dimension representation is obtained by means of a low-rank approximation of the data through standard matrix factorization techniques (Guo et al., 2021; Tang et al., 2024), neural networks (Sinha and Lucheroni, 2025) or FPCA (Pelagatti, 2013; Li et al., 2025).

FPCA is a well-known, simple, and effective method for forecasting functional time series (Hyndman and Shahid Ullah, 2007; Hyndman and Shang, 2009; Shang, 2013; Aue et al., 2015). It is therefore a natural choice for a benchmark framework. Further arguments supporting this choice are provided in section 2.3.2.

Work	Model typology	Dim. reduction	Curve form	Price forecasting
Pelagatti (2013)	VAR	FPCA	Inverse	
Ziel and Steinert (2016)	VAR	Discretization	Inverse	✓
Canale and Vantini (2016)	Functional AR	✗	Original	✓
Mestre et al. (2020)	Functional AR	✗	Inverse	
Kulakov (2020)	VAR	Discretization	Inverse	✓
Shah and Lisi (2020)	Functional AR	✗	Original	✓
Guo et al. (2021)	Neural Network	PCA	Original	
Soloviova and Vargiolu (2021)	Functional AR	✗	Original	✓
Yildirim et al. (2023)	State-space	Parametrization	Inverse	
Ciarreta et al. (2023)	VAR	Parametrization	Original	✓
Tang et al. (2024)	Neural Network	PCA, NMF, SDL	Original	
Vivó and Alonso (2024)	Neural Network	✗	Inverse	
Ghelas and Ziel (2024)	VAR	Discretization	Inverse	
Li et al. (2024)	k-NN	✗	Inverse	
Ghelas and Ziel (2025)	Fundamental	✗	Original	✓
Sinha and Lucheroni (2025)	Neural Network	Autoencoder	Original	
Li et al. (2025)	ARIMA, Functional AR	FPCA	Inverse	✓

Table 2.1: Tabular view of (main) previous works addressing supply & demand curves forecasting, sorted by publication year.

Modeling the original vs inverse curve

Being a monotone step-function, a merit-order curve $P(q)$ expressing price as a function of quantity admits an inverse $Q(p)$ expressing quantity as a function of price which is also a monotone step function. These are just two different functional representations of the same data object, and one can easily switch from one representation to an other. Interestingly, a large number of contributions work with the inverse form Q – offering little discussion on the reasons behind this choice – though there is no clear consensus on what is the best representation strategy between the two forms. However, as we will see in section 2.3.1, the inverse form has a more natural mathematical definition and offers a convenient perspective on the main source of curves variability.

Addressing the monotonicity constraint

Merit-order curves are by definition monotone, increasing for supply and decreasing for demand and this must be ideally respected for predictions. To enforce this constraint, two strategies appear in the literature:

The first is to transform the curves to work in a space of unconstrained functions, usually achieved applying a composition of a logarithmic and differential operator on the smoothed curves (Pelagatti, 2013; Canale and Vantini, 2016), work with first-order differences over a discretization grid (Ziel and Steinert, 2016; Kulakov, 2020; Ghelas and Ziel, 2024) or assuming a monotone parametric form (Soloviova and Vargiolu, 2021; Ciarreta et al., 2023; Ghelas and Ziel, 2025). Note that works considering the two latter approaches do not exactly

enforce the monotonicity constraint as the vector representation – either first-order differences or parametric coefficients – components sign is not constrained in the prediction process. It doesn't seem however to be an issue in practice.

The second is to post-process the predictions, for instance by taking the predicted curve's nearest neighbor in the training set (Sinha and Lucheroni, 2025; Li et al., 2025) or applying a correction re-establishing monotonicity in the predicted curve (Mestre et al., 2020; Vivó and Alonso, 2024).

In our case, rather than working in the constrained functional framework, although appealing from a mathematical point of view, we adopt the post-processing approach as the log-differential transformation was found to introduce excessive noise in the FPCA during preliminary experiments.

Dealing with the step-like characteristic

Merit-order curves are step functions whose step heights and widths vary across both the domain and time. Most studies, either explicitly or implicitly, work with continuous or smooth approximations of these curves – a practice we consider appropriate. Nevertheless, this approach has been criticized by several authors (Vivó and Alonso, 2024; Li et al., 2024, 2025; Sinha and Lucheroni, 2025), who therefore develop techniques that preserve the step-like form, particularly for point forecasts. However, this can lead to inconsistent or misguided modeling choices: For instance, Li et al. (2024) use a k -NN algorithm to predict curves with $k = 1$ to avoid the smoothing induced when averaging curves for $k > 1$, hence departing from the usual bias–variance trade-off considerations usually driving the choice of k . Similarly, Li et al. (2025) use a functional basis of constant functions defined on equal-width, disjoint intervals, thereby imposing fixed step widths, which the data do not exhibit.

Beyond offering a more tractable form, working with smooth approximations is theoretically grounded. Indeed, as pointed out by Pelagatti (2013) and Diquigiovanni et al. (2024), step locations (the bid prices) are random; being real-valued, they can be modeled as realizations of a continuous random variable. Consequently the expected value of the curves process is a smooth function³ and smooth point forecasts – being conditional mean estimators – are therefore appropriate. This hypothesis is easily verified empirically by inspecting the smoothness of pointwise average curves in real merit-order curves datasets.

Joint prediction of supply and demand

Day-ahead power demand has historically shown little elasticity making the demand merit-order curve look more like a vertical line close located at the load forecast than decreasing steps. This point has undoubtedly motivated academic researchers to focus primarily on the supply side (Pelagatti, 2013; Mestre et al., 2020; Guo et al., 2021; Yıldırım et al., 2023; Tang et al., 2024; Li et al., 2025;

³Further arguments are provided in section 2.3.2.

Ghelasi and Ziel, 2025). However, first, this assumption does not hold for all markets (e.g. EPEX-DE, the German day-ahead market) and second, the increasing penetration of large-scale storage systems together with the growth of demand-side response is likely to increase day-ahead demand elasticity in the next years.

Although any methodology developed for supply curves is trivially applicable to demand curves by considering their opposite (which is a supply-like increasing function) the challenge is to account for the statistical dependence that may exist between the two objects. This point is crucial when the ultimate goal is price forecasting. Among the works that consider both supply and demand curves, only a few propose a method capable of making supply curves depend on past demand curves and vice-versa. A simple yet effective approach (which we adopt in this work) within dimensionality-reduction frameworks is to stack the supply and demand vector representations into a single vector representing the pair, to which multivariate forecasting models can then be applied (Ziel and Steinert, 2016; Ciarreta et al., 2023). This dependence structure is also incorporated by Canale and Vantini (2016) via a bivariate FAR model (a functional extension of a two-dimensional vector AR model) applied directly to the supply–demand function pair. A notable conclusion from their study is that modeling this dependence is indeed essential for achieving the best predictive performance.

Curves forecast evaluation and clearing price forecasting

A final crucial point regards the way merit-order curves forecasts are evaluated, since not all curves portions carry the same importance. Indeed, in a pay-as-clear auction, the bid price doesn’t matter so much as long as it is estimated to fall either below, above or around the market clearing price. As a result, as we move further from the clearing quantity, bid prices may be subject to an arbitrary variability which is irrelevant from an economic point of view. For instance, market participants with variable renewable energy assets may place bids either at $-500\text{€}/\text{MWh}$ ⁴ or $0\text{€}/\text{MWh}$, while still being certain to be cleared if the market clearing price never falls below $10\text{€}/\text{MWh}$. This variability is likely to be hard to capture and does not reflect variability in market outcomes.

A simple way to address this issue, adopted in several works, is to restrict the supply-demand curves pair to the narrowest domain (possibly with some margin) containing all historic intersections, i.e. clearing price-quantity points (Mestre et al., 2020; Shah and Lisi, 2020; Soloviova and Vargiolu, 2021). This is the approach we follow in this work. Since the conditional distribution of the clearing price exhibits smaller temporal variations than that of the clearing quantity – which has lower uncertainty and closely follows daily demand patterns, this approach is likely to be more effective at discarding low-significance bids when curves are represented in their inverse $Q(p)$ form. Another possible approach is

⁴The minimum bidding price in European markets.

to "clip" bid prices within a conservative range of plausible clearing prices, as done by Guo et al. (2021) and Tang et al. (2024). Alternative weighting-based evaluation methods have also been proposed, based on the distribution of bid prices (Vivó and Alonso, 2024; Li et al., 2024) or clearing prices (Tang et al., 2024). Their main drawback is that results tend to be highly sensitive to the chosen weighting scheme.

Finally, a desirable property of supply and demand curves forecasts is their ability to yield accurate clearing price predictions, and they must be evaluated accordingly. As shown in Table 2.1, most works addressing both supply and demand curves forecasting tested their ability to produce accurate price forecasts, but rarely following the EPF best practices established in Lago et al. (2021).

2.2.2 Knowledge gap and contributions

Merit-order curves forecasting benchmark

Given the absence of a unified evaluation framework – as in EPF – and the scarcity of direct and fair comparisons between proposed approaches, identifying top-performing merit-order curves forecasting techniques remains challenging. Moreover, only a small number of papers provide indicative algorithms running times, preventing a clear assessment of their performance/efficiency trade-off.

Ziel and Steinert (2016) were the first to propose a method addressing all aspects of the problem. Their framework is simple, general enough to be applied across different markets, computationally efficient, and relatively easy-to-implement. Yet, aside from a few direct extensions (Kulakov, 2020; Ghelasi and Ziel, 2024) and despite a large number of subsequent contributions, no alternative method has been conclusively shown to outperform the original approach of Ziel and Steinert. Their framework remains the de facto state-of-the-art, and whether it can be consistently improved remains an open question. We identified only one study, Yıldırım et al. (2023), that provides a direct comparison of forecast accuracy for supply curves. Their results suggest that their method and that of Ziel and Steinert perform similarly. A comparison is also reported in Sinha and Lucheroni (2025), but the authors state that they employed a modified version of Ziel and Steinert’s model without providing further details. Additionally, we are somewhat skeptical of their results as all considered benchmarks – including Ziel and Steinert’s model – are found to perform far worse than a naive method.

This work aims to provide a general, simple, and computationally efficient benchmark framework for day-ahead supply and demand merit-order curves forecasting, from which both point and probabilistic clearing price predictions can be derived. We conduct a rigorous comparison with Ziel and Steinert’s framework and provide an open-access Python implementation⁵ (github.com/gkchl

⁵We also implement the model of Ziel and Steinert (2016), considering the multivariate

`n/moc-forecast`) to facilitate replication and application to other merit-order curves datasets.

Comparison with state-of-the-art EPF models for price prediction

As mentioned earlier in the introduction, it remains unclear whether merit-order curves models are capable of yielding state-of-the-art point and probabilistic clearing price forecasting performance. While Ziel and Steinert demonstrated ten years ago that their approach outperformed existing price-based models for point forecasting, this result may no longer hold in light of recent advances in EPF.

Several studies have assessed the effectiveness of curve-based approaches for clearing price prediction by comparing them with price-based methods; however, the latter are often overly simplistic and not representative of top-performing EPF models. In addition, the rigorous forecast evaluation practices recommended by Lago et al. (2021), such as daily recalibration, are seldom followed. Furthermore, none of the post-2021 contributions compare curve-based price forecasts against the benchmark models of Lago et al., despite their availability through an open-access software implementation (the `epftoolbox` Python package). Shah and Lisi (2020) compare their curve-based price forecasts only against basic autoregressive models without exogenous covariates, even though variables such as load and renewable generation forecasts are known to be essential for achieving top performance (Weron, 2014). Similar limitations apply to Solovieva and Vargiolu (2021), Ciarreta et al. (2023) and Li et al. (2025). Ghelasi and Ziel (2025) consider an "expert" autoregressive model with exogenous covariates, but do not include regularized, parameter-rich autoregressive models with cross-hour dependence, despite evidence that such models consistently outperform expert ARX specifications (Ziel and Weron, 2018). For probabilistic forecasting, bootstrap-based Monte Carlo approaches similar to ours were proposed by Ziel and Steinert (2016) and Shah and Lisi (2020), but they were not compared with state-of-the-art probabilistic EPF methods.

A rigorous empirical comparison between our curve-based price forecasting approach and state-of-the-art price-based EPF models is conducted in this work. The comparison is restricted to regularized, parameter-rich autoregressive models (Ziel, 2016; Uniejewski et al., 2016; Ziel and Weron, 2018; Lago et al., 2021) as our curves forecasting procedure relies on a direct multivariate extension applied to the curves vector representation. In these experiments, all models share a common regularized *vector* autoregression implementation in Python. For probabilistic forecasting, we compare with quantile regression averaging (Nowotarski and Weron, 2018), conformal prediction (Kath and Ziel, 2021) and isotonic distributional regression (Henzi et al., 2021), using the open-access implementation

models described in this paper – very similar to Ziel and Steinert’s original ones – within a shared software architecture.

of Lipiecki et al. (2024).

2.3 Methodology

2.3.1 Curves definition

Formally, a merit-order curve represents price as a function of quantity $q \rightarrow P(q)$ and is defined from a set of bids characterized by a price and a quantity $(p_1, q_1), (p_2, q_2), \dots, (p_n, q_n)$ sorted following the merit-order, i.e., increasing order of price $p_1 \leq p_2 \leq \dots \leq p_n$ for the supply curve, decreasing order of price $p_1 \geq p_2 \geq \dots \geq p_n$ for the demand curve. Mathematically, a merit-order curve can be defined as follows:

$$P(q) = p_k, \quad \text{if } \sum_{i=0}^{k-1} q_i \leq q < \sum_{i=0}^k q_i, \quad k = 1, \dots, n$$

where we set $q_0 = 0$, for notational convenience. For supply (or demand), $P(q)$ is said to be the marginal price associated to a total offered (resp. demanded) quantity q .

From the definition we can notice that merit-order curves are right-continuous monotone step functions, non-decreasing for supply and non-increasing for demand. A direct consequence is that $q \rightarrow P(q)$ admits a *right-continuous generalized inverse* $p \rightarrow Q(p)$, which is itself a right-continuous monotone step function. It is interesting to note that this "pseudo-inverse" has an easier mathematical notation and interpretation. Indeed, its expression matches (for supply curves) the unnormalized⁶ cumulative distribution function (CDF) of a discrete distribution over the support $\{p_1, \dots, p_n\}$. For demand, the only difference is that it is the reverse CDF (or survival function). Using the superscript (s) for supply and (d) for demand, we have:

$$Q^{(s)}(p) = \sum_{i=1}^n q_i \mathbb{1}_{\{p_i \leq p\}}, \quad Q^{(d)}(p) = \sum_{i=1}^n q_i \mathbb{1}_{\{p_i \geq p\}}$$

That is, $Q(p)$ is the quantity that is offered (or demanded) at a price below (resp. above) p . Note that the definition above does not require the sequence of offers (or bids) to be price-ordered. It is more intuitive to work with the quantity function Q as it represents a more familiar mathematical object than the price function P . The quantity function simply tells us about the total offered quantity and how this offered quantity is distributed on the price domain. It also conveniently changes the perspective on the curves variability which is mainly attributed to their inelastic component: seasonal power demand variations for demand curves and renewable energy sources (RES) generation for supply curves.

From now on, unless otherwise specified, references to supply or demand curves will refer to the quantity function.

⁶Given the q_i 's don't sum to 1.

2.3.2 Curves representation

The goal is to model and forecast the time series of curves $Q_1^{(s)}, Q_2^{(s)}, \dots, Q_T^{(s)}$ and $Q_1^{(d)}, Q_2^{(d)}, \dots, Q_T^{(d)}$ over a time period composed of T hourly intervals.

FDA is a framework that extends traditional statistical methods to handle data that vary over a continuum, such as curves, surfaces, or other functional forms. Unlike classical approaches that analyze discrete points, FDA treats entire functions as the fundamental unit of analysis. It assumes however that only discrete and noisy samples of continuous functions are observed and therefore the first step of any FDA is to estimate the underlying functions from the data – a process known as *smoothing*.

In our case, things are a bit different: we observe the full functions and these functions are not continuous. Nevertheless, as we saw that the quantity function is expressed as the (unnormalized) CDF of some discrete distribution, with a non-fixed, random support, it may be interpreted as a realization of a Dirichlet process (Ferguson, 1973; Teh, 2011) with a continuous base distribution. The corresponding base distribution has a smooth CDF, which then becomes the natural object of analysis.

In order to forecast functional time series, we follow the guidance of Aue et al. (2015) that recommend the use of functional principal component analysis (FPCA) to derive an uncorrelated vector representation of the curves and apply standard multivariate time series forecasting methods to this representation, for finally obtaining a curves forecast using the Karhunen-Loève (KL) expansion (Horváth and Kokoszka, 2012). FPCA is a dimensionality reduction technique for functional data, extending classical PCA to infinite-dimensional spaces. It decomposes a set of observed functions into an orthonormal basis that captures the dominant modes of variations in the data. This decomposition yields the truncated KL expansion:

$$Q_t(p) \approx \bar{Q}(p) + \sum_{k=1}^K \beta_{tk} \xi_k(p)$$

where $\bar{Q}(p)$ is the mean function, $\xi_1(p), \xi_2(p), \dots, \xi_K(p)$ are the first K *functional principal components* (FPCs) and $\beta_{t1}, \beta_{t2}, \dots, \beta_{tK}$ are the *scores* of observation t on the first K FPCs. Intuitively, the FPCs can be seen as specific features common to all curves while the scores are measures of how pronounced are these features in a specific curve. The perfect equality holds when $K = \infty$ but in practice a satisfying approximation can be found with a finite K , typically chosen – as in traditional PCA – as the elbow point of the scree plot or such that the ratio of explained variance surpasses a certain threshold (Johnson and Wichern, 2007).

Since FDA assumes smooth functions, the estimated FPCs should also be smooth functions. This smoothness can be achieved either by incorporating

a roughness penalty directly into the FPCA estimation procedure or by pre-smoothing the functional data curves prior to applying FPCA. We adopt the latter approach for computational simplicity, utilizing kernel smoothing with the *Nadaraya-Watson* kernel estimator (Ferraty and Vieu, 2004). The bandwidth parameter of the kernel function, which controls the level of smoothness, is chosen using the generalized cross-validation criterion.

Aue et al. (2015) show that a vector autoregressive model of order p – $\text{VAR}(p)$ – applied to the K -dimensional vector representation given by the scores on the K first principal components is asymptotically equivalent to a functional autoregressive model (Bosq, 2000) of the same order – $\text{FAR}(p)$ – when $K \rightarrow \infty$. Therefore, they argue that forecasting the functional time series by forecasting the vector time series of the FPC scores (given that it provides a satisfying dimensionality reduction) with any multivariate forecasting method – being a VAR model or not – is theoretically justified and is easier than devising complex functional models. Note that the uncorrelation property of the FPC scores has motivated approaches where univariate models are separately fitted to each of the score series (Hyndman and Shahid Ullah, 2007; Hyndman and Shang, 2009). As highlighted by Aue et al. (2015), though it can turn out to be a simple and efficient way to forecast the vector of scores, it has the major drawback of overlooking the cross-correlation that can exist at lags greater than zero, since the FPCA procedure only guarantees that the vectors have no instantaneous (lag 0) correlation.

In our case, we would like to account for possible dependence between the supply and demand curves time series. To do so, we perform FPCA separately for the supply and demand curves and by concatenating the K_s scores for the supply curves with the K_d scores of the demand curve, we get a $K = K_s + K_d$ -dimensional vector representation $\mathbf{y}_1, \dots, \mathbf{y}_T$ of the paired functional series $(Q_t^{(s)}, Q_t^{(d)})_{t \in \{1, \dots, T\}}$, with

$$\mathbf{y}_t = \left[\beta_{t1}^{(s)}, \beta_{t2}^{(s)}, \dots, \beta_{tK_s}^{(s)}, \beta_{t1}^{(d)}, \beta_{t2}^{(d)}, \dots, \beta_{tK_d}^{(d)} \right]^\top$$

The advantage of working with $\mathbf{y}_t$ is triple: (i) we turn an infinite-dimension problem into an intrinsically finite-dimension one, (ii) we phase out the intrinsic dependence of functional data, (iii) we jointly model the supply and demand processes.

We forecast $\mathbf{y}_t$ with some multivariate forecasting model (detailed in the next section), with possibly a set of exogenous covariates $\mathbf{x}_t$. We then inverse-transform a h -step-ahead forecast $\hat{\mathbf{y}}_{t+h}$ using the truncated KL expansion to get a forecast for the curves pair $(\hat{Q}_{t+h}^{(s)}, \hat{Q}_{t+h}^{(d)})$.

It can happen that curves predictions are not perfectly monotonic. Though preliminary experiments suggested the occurrence of little deviations only and low impact on clearing price predictions, it is still desirable to respect this structural constraint. To do so, we post-process the curves predictions using isotonic

regression with the pool-adjacent violators algorithm (Leeuw et al., 2009). Being a highly efficient $O(n)$ algorithm, it has a negligible impact on computation times.

2.3.3 Forecasting the curves' vector representation

We consider here the day-ahead forecasting problem, regarding the vector hourly time series as 24 separate daily time series, one for each hour of the day. This is common practice in EPF as the 24 hours of the next day are simultaneously settled the day before. In addition, market dynamics vary a lot depending on the hour of the day, justifying to treat them separately. For a detailed motivation of this choice, the reader can refer to Ziel and Weron (2018). We hence change the time indexing of $\mathbf{y}_t$ to consider the value at day d and hour h , $\mathbf{y}_{d,h}$ and we solve 24 one-step-ahead forecasting problems.

We consider four variants of the popular parameter-rich ARX model estimated with LASSO widely used in the context of electricity price forecasting (Ziel, 2016; Uniejewski et al., 2016; Ziel and Weron, 2018; Lago et al., 2021; Uniejewski, 2024). First this framework was shown to be highly performing for price forecasting and second, it was used in Ziel and Steinert (2016).

Specifically, we consider two *univariate* models treating each component of $\mathbf{y}_{d,h}$ separately as suggested by Hyndman and Shang (2009) – the *concurrent* ARX and the *full* ARX – and two *multivariate* models jointly modeling the vector $\mathbf{y}_{d,h}$, the concurrent VARX and the full VARX. The term "concurrent" means the target variable at hour h is influenced only by past values from the same hour h . Conversely, "full" means the variable is also influenced by past values observed at different hours. The comparison between the univariate and the multivariate approaches allow to test the added value of modeling the cross-dependence between scores while the comparison between the concurrent and full approaches allow to test the added value of modeling the cross-dependence between hours.

The *AR* part refers to the fact that lagged values are used to predict future values. In our case, similarly to state-of-the-art EPF models (Lago et al., 2021) we include lags 1, 2, 3 and 7, considering therefore information up to 7 days before. Additionally, each model has the suffix *X* which means that it makes use of exogenous variables which refer to day d and available on $d - 1$ (i.e., they are known in anticipation or they are themselves one-day-ahead forecasts). This set of r exogenous variables is represented by the r -dimensional vector $\mathbf{x}_{d,h}$. Again, like Lago et al. (2021), we also include lags 1 and 7 of these exogenous variables, and a three-dimensional vector $\mathbf{z}_d$ of dummy variables flagging the day type: Mondays, Working days (from Tuesday to Friday), Saturdays and Holidays (Sundays and bank holidays)⁷.

⁷Note this is slightly different from Lago et al. (2021) who consider one dummy for each day of the week but similar to Mestre et al. (2020) and addresses the calendar effects issues

Finally, still following Lago et al.'s guidance, the L^1 -regularization parameter λ is selected as that minimizing the Akaike information criterion (AIC), exploiting the least angle regression (LARS) algorithm (Efron et al., 2004) for the parameter search. Once λ is selected, the model is estimated using the traditional coordinate descent algorithm (Tibshirani, 1996).

Concurrent ARX (**ARX**)

The concurrent ARX⁸ models each component $y_{d,h}$ of $\mathbf{y}_{d,h}$ as:

$$y_{d,h} = \phi_{1,h}y_{d-1,h} + \phi_{2,h}y_{d-2,h} + \phi_{3,h}y_{d-3,h} + \phi_{7,h}y_{d-7,h} + \beta_{0,h}^\top \mathbf{x}_{d,h} + \beta_{1,h}^\top \mathbf{x}_{d-1,h} + \beta_{7,h}^\top \mathbf{x}_{d-7,h} + \boldsymbol{\theta}_h^\top \mathbf{z}_d + \epsilon_{d,h} \quad (2.1)$$

where $\phi_{\cdot,h}$ are the autoregressive coefficients and $\beta_{\cdot,h}$ and $\boldsymbol{\theta}_h$ are the r -dimensional and 3-dimensional vectors of (lagged) exogenous and dummy variables coefficients, respectively.

Full ARX (**fARX**)

The full ARX⁹, very similar to its homonym in Uniejewski et al. (2016), models each component $y_{d,h}$ of $\mathbf{y}_{d,h}$ as:

$$y_{d,h} = \sum_{j=1}^{24} \phi_{1,h}^{(j)} y_{d-1,j} + \sum_{j=1}^{24} \phi_{2,h}^{(j)} y_{d-2,j} + \sum_{j=1}^{24} \phi_{3,h}^{(j)} y_{d-3,j} + \sum_{j=1}^{24} \phi_{7,h}^{(j)} y_{d-7,j} + \beta_{0,h}^\top \mathbf{x}_{d,h} + \beta_{1,h}^\top \mathbf{x}_{d-1,h} + \beta_{7,h}^\top \mathbf{x}_{d-7,h} + \boldsymbol{\theta}_h^\top \mathbf{z}_d + \epsilon_{d,h} \quad (2.2)$$

where $\phi_{\cdot,h}^{(j)}$ is the linear effect of the lagged hour j on the hour h .

Concurrent VARX (**VARX**)

The concurrent VARX models the full vector $\mathbf{y}_{d,h}$ as:

$$\mathbf{y}_{d,h} = \boldsymbol{\Phi}_{1,h} \mathbf{y}_{d-1,h} + \boldsymbol{\Phi}_{2,h} \mathbf{y}_{d-2,h} + \boldsymbol{\Phi}_{3,h} \mathbf{y}_{d-3,h} + \boldsymbol{\Phi}_{7,h} \mathbf{y}_{d-7,h} + \mathbf{B}_{0,h} \mathbf{x}_{d,h} + \mathbf{B}_{1,h} \mathbf{x}_{d-1,h} + \mathbf{B}_{7,h} \mathbf{x}_{d-7,h} + \boldsymbol{\Theta}_h \mathbf{z}_d + \boldsymbol{\epsilon}_{d,h} \quad (2.3)$$

Where $\boldsymbol{\Phi}_{\cdot,h}$ are $K \times K$ matrices of autoregressive coefficients and $\mathbf{B}_{\cdot,h}$ and $\boldsymbol{\Theta}_{\cdot,h}$ are the $K \times r$ and $K \times 3$ matrices of (lagged) exogenous and dummy variables coefficients, respectively. Contrarily to the two previous models, we allow for cross-dependence – described by the off-diagonal coefficients of $\boldsymbol{\Phi}_{\cdot,h}$ – between the components of $\mathbf{y}_{d,h}$.

raised by Ziel and Steinert (2016).

⁸Note that this model is equivalent to a SARIMAX(3,0,0) $\times$ (1,0,0)₇, with the difference that it is estimated with the LASSO and not with maximum likelihood.

⁹Similarly, this model is equivalent to a VARX(7), applied to the 24-dimensional vector $[y_{d,1}, \dots, y_{d,24}]^\top$ – with lags from 4, 5, 6 constrained to be null – with the difference that it is estimated with the LASSO and not with maximum likelihood.

Full VARX (**fVARX**)

The full VARX models the full vector $\mathbf{y}_{d,h}$ as:

$$\begin{aligned} \mathbf{y}_{d,h} = & \sum_{j=1}^{24} \Phi_{1,h}^{(j)} \mathbf{y}_{d-1,j} + \sum_{j=1}^{24} \Phi_{2,h}^{(j)} \mathbf{y}_{d-2,j} + \sum_{j=1}^{24} \Phi_{3,h}^{(j)} \mathbf{y}_{d-3,j} + \sum_{j=1}^{24} \Phi_{7,h}^{(j)} \mathbf{y}_{d-7,j} \\ & + \mathbf{B}_{0,h} \mathbf{x}_{d,h} + \mathbf{B}_{1,h} \mathbf{x}_{d-1,h} + \mathbf{B}_{7,h} \mathbf{x}_{d-7,h} + \Theta_h \mathbf{z}_d + \epsilon_{d,h} \end{aligned} \quad (2.4)$$

Where we impose the restriction that the off-diagonal terms of $\Phi_{\cdot,h}^{(j)}$ must be zero whenever $j \neq h$, that is, cross-dependence is allowed between hours, but only within the same component¹⁰. This restriction is imposed to have a manageable number of parameters. Indeed, without this condition we would have $4 \times [24 \times K] + 3r + 3 = 96K + 3r + 3$ parameters for each component of $\mathbf{y}_{d,h}$, while with this condition the number of parameters is only $4 \times [23 + K] + 3r + 3 = 4K + 3r + 95$. For instance, for $K = 10$ and $r = 4$, the condition implies 147 parameters instead of 975.

2.3.4 Clearing price point and probabilistic forecasting

The clearing price *point* prediction $\hat{p}_t$ is simply obtained by determining the zero¹¹ of the function $p \rightarrow \hat{Q}_t^{(d)}(p) - \hat{Q}_t^{(s)}(p)$.

Regarding the *probabilistic* prediction, the goal is to forecast the quantiles of the conditional distribution of the price at day d and hour h given all information available at $d - 1$. The strategy we adopt is rather straightforward and based on Monte Carlo simulations:

- (1) Estimate the predictive distribution of $\mathbf{y}_t$ nonparametrically using a residual-based bootstrap on a rolling calibration window
- (2) Simulate a vector $\mathbf{y}_t^{(\text{sim})}$ from that distribution
- (3) Inverse transform $\mathbf{y}_t^{(\text{sim})}$ to obtain $(Q_t^{(s,\text{sim})}, Q_t^{(d,\text{sim})})$
- (4) Find the clearing price $p_t^{(\text{sim})}$
- (5) Repeat (2), (3) and (4) N times
- (6) Take the quantiles of the empirical clearing price distribution $\hat{F}_t$ obtained.

The choice of the calibration window is a classic trade-off as short windows have the advantage of capturing potential non-stationarity and correcting eventual

¹⁰In that sense, "semi-full VARX" could be a more precise denomination but we keep "full VARX" for simplicity.

¹¹Computationally speaking, we take the linear interpolation of the two prices concerned by the sign change.

forecasts bias while long windows will allow for a more accurate estimation of the errors distribution. To get "the best of both worlds", we proceed as in Lipiecki et al. (2024): We consider four different window sizes: 28 days (4 weeks), 56 days (8 weeks), 91 days (13 weeks) and 182 days (26 weeks) yielding four CDF estimates $\hat{F}_t^{(28)}$, $\hat{F}_t^{(56)}$, $\hat{F}_t^{(91)}$ and $\hat{F}_t^{(182)}$. We then compute an ensemble prediction $\hat{F}_t^{(\text{ens})}$ considering *vertical* (probability) averaging (Lichtendahl et al., 2013):

$$\hat{F}_t^{(\text{ens})} = \frac{1}{4} \left(\hat{F}_t^{(28)} + \hat{F}_t^{(56)} + \hat{F}_t^{(91)} + \hat{F}_t^{(182)} \right)$$

Regarding step (1), we estimate the multivariate distribution of the K -dimensional forecast error vector $\epsilon_{d,h} = \mathbf{y}_{d,h} - \hat{\mathbf{y}}_{d,h}$. To robustly estimate this distribution which we assume stationary within the calibration window, we consider hour-dependent variances for the marginals but a shared correlation structure: We assume the errors $\epsilon_{d,h}$ for a given hour h are independent and identically distributed (i.i.d.) according to some K -dimensional distribution $\mathcal{D}(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h)$ with mean $\boldsymbol{\mu}_h$ and a covariance matrix $\boldsymbol{\Sigma}_h$ which can be decomposed as follows:

$$\boldsymbol{\Sigma}_h = \mathbf{V}_h^{1/2} \mathbf{R} \mathbf{V}_h^{1/2}$$

where $\mathbf{V}_h$ is the $K \times K$ diagonal matrix containing the hour-specific marginals variances and $\mathbf{R}$ the correlation matrix, assumed identical for all hours. Hence, by bootstrapping the *marginal*-standardized residuals

$$\boldsymbol{\eta}_{d,h} = \hat{\mathbf{V}}_h^{-1/2} (\epsilon_{d,h} - \hat{\boldsymbol{\mu}}_{d,h})$$

across *all hours*, we obtain simulations $\boldsymbol{\eta}^* \sim \hat{\mathcal{D}}(\mathbf{0}, \mathbf{R})$ which can be transformed in simulations of $\epsilon_{d,h}$:

$$\epsilon_{d,h}^{(\text{sim})} = \hat{\boldsymbol{\mu}}_h + \hat{\mathbf{V}}_h^{1/2} \boldsymbol{\eta}^*$$

This modeling choice allows us to account for possible hour-dependent uncertainty without having to estimate 24 distinct distributions.

2.3.5 Benchmark curves and price forecasting models

Naive model

The naive model (**Naive**) works as follows:

$$\hat{Q}_{d,h}^{\text{naive}} = \begin{cases} Q_{d-7,h} & \text{if } d \text{ corresponds to a Monday, Saturday or Sunday} \\ Q_{d-1,h} & \text{otherwise} \end{cases} \quad (2.5)$$

Ziel and Steinert (2016)

While our approach differs from that of Ziel and Steinert (2016) regarding the curves vector representation, the multivariate forecasting techniques applied to that representation are essentially the same. For that reason, to guarantee a fair comparison between the two approaches, we apply the models detailed in the previous section to (i) our vector of supply and demand FPCA scores (ii) the vector of supply and demand price class volumes obtained with the method described in Ziel and Steinert (2016). We invite the reader to refer to the authors' paper for a rigorous and detailed description of this procedure (as well as for curves reconstruction), which we refer to as *Ziel-Steinert Transformation (ZST)*, as opposed to our *functional principal component analysis (FPCA)*.

Briefly, **ZST** consists of taking the first-order differences of discrete evaluations of the quantity curve on a specific price grid. This price grid is constructed by defining an equispaced *quantity* (or volume) grid and transforming it into a non-uniform *price* grid via the mean price curve $P(q)$. For inverse transformation, the differences are cumulatively summed to recover the discrete values and the continuous curves are reconstructed using interpolation based on the mean quantity curve. Fig. 2.2 shows an example of a transformed and reconstructed supply-demand curves pair.

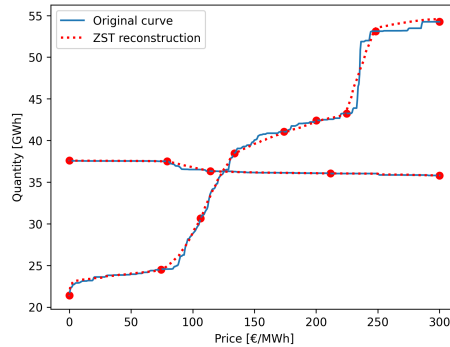


Figure 2.2: Example of **ZST** reconstruction for a supply-demand curves pair with $K_s = 9$ dimensions for supply and $K_d = 5$ for demand. The vector representation is given by the first-order differences of the y-coordinates of the red dots. The fixed price grid corresponds to their x-coordinates.

Price-based point forecasting models

We consider three benchmark *price-based* clearing price forecasting models: **ARX**, **fARX** and the Lasso-Estimated AutoRegressive model (**LEAR**) of Lago et al. (2021). The first two are already described in the previous section (equations

2.1 and 2.2). When both curve-based and price-based models are involved, to distinguish the clearing price forecasting model deriving from the curves model (*curve-based*) from that directly forecasting the price (*price-based*), we prefix the curve-based models with the curves representation used, e.g. **FPCA-ARX** or **ZST-ARX**.

Finally the **LEAR** model is very similar to **fARX** with the difference that it adds cross-hour dependence in the exogenous variables effects and accepts only two of them $\mathbf{x}_{d,h} = [x_{d,h}^1, x_{d,h}^2]^\top$ (usually the load and combined wind-solar generation day-ahead forecasts):

$$\begin{aligned}
 y_{d,h} = & \sum_{j=1}^{24} \phi_{1,h}^{(j)} y_{d-1,j} + \sum_{j=1}^{24} \phi_{2,h}^{(j)} y_{d-2,j} + \sum_{j=1}^{24} \phi_{3,h}^{(j)} y_{d-3,j} + \sum_{j=1}^{24} \phi_{7,h}^{(j)} y_{d-7,j} \\
 & + \sum_{j=1}^{24} \beta_{0,h}^{(j)\top} \mathbf{x}_{d,j} + \sum_{j=1}^{24} \beta_{1,h}^{(j)\top} \mathbf{x}_{d-1,j} + \sum_{j=1}^{24} \beta_{7,h}^{(j)\top} \mathbf{x}_{d-7,j} + \boldsymbol{\theta}_h^\top \mathbf{z}_d + \epsilon_{d,h}
 \end{aligned} \tag{2.6}$$

As in Lago et al. (2021), for all three models, prices are transformed with median-based scaling and the *area hyperbolic sine* variance stabilizing transformation.

Price-based probabilistic forecasting models

We test the three (price-based) clearing price probabilistic forecasting methods of Lipiecki et al. (2024), using the Julia package `PostForecasts.jl` (Lipiecki and Weron, 2025) developed by the authors: *Quantile Regression Machine* (**QRM**), *Conformal Prediction* (**CP**), *Isotonic Distributional Regression* (**IDR**). We also include the naive *Normal* (**N**) benchmark¹². As these methods are based on post-processing of point predictions, we use the best performing price-based model as a point predictor. For more information, we invite the reader to directly refer to Lipiecki et al. (2024).

2.3.6 Forecasts evaluation

Curves forecasts

As suggested in Ramsay and Silverman (2005), we measure curves forecasting accuracy with the *squared correlation function*:

$$R^2(p) = 1 - \frac{\sum_{t=1}^T [\hat{Q}_t(p) - Q_t(p)]^2}{\sum_{t=1}^T [Q_t(p) - \bar{Q}(p)]^2}$$

which represents the proportion of variance explained by the model across the price domain, and the *average squared correlation*:

$$R^2 = \frac{1}{p_{\max} - p_{\min}} \int_{p_{\min}}^{p_{\max}} R^2(p) dp$$

¹²Note that differently from Lipiecki et al. (2024) paper, we consider the version which does not suppose zero-mean errors and therefore accounts for potential forecasts bias.

Additionally, to test whether a model significantly outperforms another, we run the *Diebold-Mariano* (DM) test (Diebold and Mariano, 1995) on the L_2 -norm of daily functional forecast errors considering (i) the average L^2 error across the 24 hours (ii) separate tests run on each of the 24 hours¹³. Two one-sided tests are run for each pair of models.

Clearing price forecasts

We evaluate point predictions with the *mean absolute error* (MAE), the *root mean squared error* (RMSE) and the *relative MAE* (rMAE) as defined in Lago et al. (2021).

Price probabilistic predictions are evaluated using the *continuous ranked probability score* (CRPS):

$$\text{CRPS}(\hat{F}_{P_t}, p_t) = \int_{-\infty}^{\infty} (\hat{F}_{P_t}(x) - \mathbb{1}_{\{p_t \leq x\}})^2 dx$$

where p_t is the realized price at time t and $\hat{F}_{P_t}$ is the estimated conditional CDF we want to evaluate.

Probabilistic forecasting performance is composed of a *reliability* (or calibration) and *sharpness* (or resolution¹⁴) component (Nowotarski and Weron, 2018). As any proper scoring rule, the CRPS presents the advantage of simultaneously evaluating both but it can be interesting to understand how models perform on each of these two components. We assess therefore reliability with the *probability integral transform* (Gneiting and Katzfuss, 2014)

$$\text{PIT}_t = \hat{F}_{P_t}(p_t)$$

If p_t is actually drawn from $\hat{F}_{P_t}$, then PIT_t should be uniformly distributed, which can be verified visually with an histogram (or quantitatively with standard goodness-of-fit statistical tests). Sharpness, when forecasts are reliable, intuitively relates to the capacity of modeling the variability of the response's uncertainty. A natural way to assess this property is to look at the distribution of a dispersion parameter of the predictive distributions, for instance prediction interval (PI) width or standard deviation. In our experiments, we analyze the joint distribution of the PI width and the magnitude of point forecasts errors to understand if the predicted uncertainty variations are accurate (i.e. larger predicted uncertainty correlates with larger point forecasts errors).

Similarly to the curves forecasts, we perform DM tests for both point and probabilistic forecasts considering average and per-hour daily errors. For point forecasting we consider the absolute (ℓ_1) errors as in Lago et al. (2021) while we

¹³Considering daily forecast errors avoids the problem of the strong intraday error autocorrelation, typical in day-ahead price forecasting, which would violate Diebold-Mariano (DM) tests assumptions (Weron, 2014).

¹⁴The two concepts are equivalent when probabilistic forecasts have perfect reliability.

use the CRPS, as was done in Nowotarski and Weron (2018), for probabilistic forecasting.

2.3.7 Daily recalibration

As advised by Lago et al. (2021) and in order to simulate a live forecasting setting, the entire modeling pipeline is recalibrated on a daily basis using a rolling window of 364 days (52 weeks), since ≈ 1 year appears to be the best single-window choice for EPF according to Marcjasz et al. (2018) and Hubicka et al. (2019). This means that for every day d of the testing period:

- (1) We extract the vector representation of the $364 \times 24 = 8744$ curves pairs observed on $d - 1, d - 2, \dots, d - 364$
- (2) We fit the multivariate forecasting model on these vector observations
- (3) We predict the 24 curves vector representations for day d
- (4) We inverse transform the predicted vectors to obtain the 24 curves pairs forecasts

Regarding step (1) this implies that the **FPCA** or **ZST** transformation is re-run from scratch every day. This dynamic "representation learning" procedure is clearly preferable to a time-invariant representation learned on a fixed initial time window but adds a layer of complexity: We must choose the supply and demand dimensions (K_s, K_d) in an automatic manner.

For **FPCA**, we combine the elbow and threshold methods (see section 2.3.2) by taking the maximum of the two suggested number of FPCs to retain, considering a ratio of 99% for the threshold method. While automating the latter is straightforward, the elbow method relies on a qualitative visual assessment of the scree plot and requires therefore a more sophisticated approach. To automatically find this elbow point in the recalibration procedure, we use the popular knee-point detection method of Satopaa et al. (2011).

For **ZST**, Ziel and Steinert (2016) do not provide discussion on their choice for the number of price classes to consider. Since our goal is to compare the efficiency of the two curves representation strategies, we should ideally match it on a daily basis with the dimension picked by **FPCA**. For simplicity though, we take fixed K_s and K_d equal to their maximum value observed across the **FPCA** recalibration procedure, leaving therefore a slight advantage to **ZST**.

Note that this same daily recalibration procedure applies to the price-based models.

2.4 Data

2.4.1 Supply and demand curves

We analyze the supply and demand merit-order curves of the Italian day-ahead electricity market – *Mercato del Giorno Prima* (MGP) – over the 2023-2024 period. The testing period covers the whole year 2024 (366 days). Note that probabilistic price predictions require at least 182 days of out-of-sample point predictions so the testing period for probabilistic forecasting starts on July 1st, 2024, covering a period of 6 months. The quantity curves are built according to the procedure described in section 2.3.1, using the *DomandaOfferta* dataset retrieved from the FTP server of the Italian NEMO – Gestore dei Mercati Energetici (GME) – which contains all supply offers and demand bids submitted on the MGP. Three points must be highlighted:

Copper plate model

Around 35% of the time, the country-level economic optimum is not feasible due to transmission capacity constraints among the seven bidding zones composing Italy, which means that the market cannot be cleared from the country-level supply and demand curves. In this case, GME uses a *zonal pricing* mechanism, which splits the entire country-level market pool in a minimum number of "sub-pools" of market zones such that, in each subpool, the economic optimum is feasible. As a result, there can be several pairs of supply and demand curves – hence several market clearing prices – for a single hour, one for each subpool. To keep the analysis simple, we did as if the country-level economic optimum was always feasible and built the country-level supply and demand curves for any hour, ignoring the problem of transmission congestion, i.e., considering a *copper plate* model. As a result, the constructed curves reflect the true market clearing 65% of the time, the remaining 35% corresponding to a "virtual" market clearing only. Note the same curves construction procedure was considered in Mestre et al. (2020). The associated market clearing price is what GME calls the "National" or "Italy (unconstrained)" price and reports together with the zonal prices (see <https://www.mercatoelettrico.org/en-us/Home/Results/Electricity/MGP/Results/ZonalPrices>).

Market coupling

Italy is integrated into the European electricity market through market coupling with several neighboring countries which are France, Austria, Slovenia and Greece. Market coupling allocates cross-border trading capacity in order to harmonize European prices. It therefore imposes the cross-border exchanges between coupled countries. As a result, the purchased quantity in MGP does not necessarily equal the sold quantity: if Italy imports more than it exports, the (domestic) sold quantity will be lower than the purchased quantity because

a part of the demand is already satisfied by the imports. Conversely, if Italy exports more than it imports, the sold quantity will be higher than the purchased quantity because part of the supply must satisfy the exports. As a consequence, the market clearing price is not rigorously at the intersection of the supply and demand curves. However, everything happens as if imports were entering in the supply offers at the minimum price – such that the underlying offer is necessarily accepted – causing a right shift of the price *supply* curve, while foreign exports enter in the demand bids at the maximum price – such that the underlying bid is necessarily accepted – causing a right shift of the price *demand* curve. Therefore by incorporating the imports and exports in the supply and demand curves, the market clearing is found exactly at the intersection. In our case, the import and export quantities within the scope of market coupling are retrieved from the `MarketCoupling` GME dataset and the difference between imports and exports is added to the quantity supply curve, which is equivalent to adding them separately to both curves (from the point of view of market clearing).

Restriction of the curves domain

A last observation regards the domain on which the quantity curves are analyzed. The entire price domain ranges from -500€ to 3000€. Obviously, as stressed in the introduction, the part of that domain where the intersection point has negligible chances to occur is of lower interest, may be subject to arbitrary variability that has no impact of the market outcomes and could just add undesired noise in the model. Therefore, we decided to restrict the quantity curves to a price domain that contains at least all prices in the 2023-2024 period with a generous margin: indeed, we restrict the curves to the domain comprised between 0€/MWh and 300€/MWh.

Three example pairs of supply and demand curves of the dataset are showed on Fig. 2.3

2.4.2 Exogenous predictors

We collect a total of 4 exogenous variables for both curves and price forecasts which are similar to those of Mestre et al. (2020) and commonly chosen in EPF:

- *Day-ahead load forecast* (1 variable): Country-level day-ahead load forecast obtained from ENTSO-E (2025).
- *Day-ahead net transfer capacities* (2 variables): We consider the import net transfer capacity (NTC) with France and Switzerland obtained from ENTSO-E (2025). These two variables are relevant as Italy (nearly) systematically imports cheaper electricity from France and Switzerland, in the limit of the NTC. Therefore, variations of the latter directly affect market results – specifically the coupling imports from France and the regular imports (subject to the merit-order) from Switzerland. Note that the other

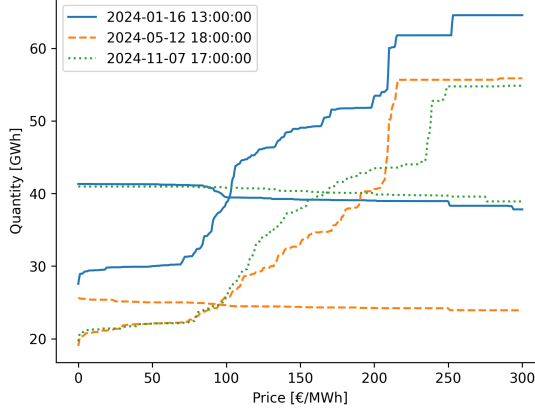


Figure 2.3: *Three example supply-demand curves pairs during the testing period.*

neighboring countries were not considered because of smaller NTC and the absence of a dominant flow direction.

- *Day-ahead forecast of wind and solar generation* (1 variable): Combined country-level day-ahead forecasts of wind and solar generation obtained from a private data provider (LSEG Data & Analytics, 2025), derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) and Global Forecast System (GFS) weather models.

Note that all these predictors are available at least one hour before the day-ahead market gate closure at 12:00 and can therefore realistically produce market results forecasts informing trading strategies "in time".

2.5 Results

2.5.1 Functional principal component analysis

The dynamic number of supply and demand principal components K_s, K_d selected across the test period by our combined elbow-threshold method is pictured on Fig. 2.4. It turns out this number is rather stable, for both sides with $K_s = 8$ or 9 and $K_d = 4$ or 5 depending on the period.

Fig. 2.5 shows the dynamic functional principal components across the test period. The stability suggested by the nearly constant number of selected FPCs is confirmed as we can see that each single FPC roughly captures a fixed mode of variation along time. For supply, FPCs subsequent to the fifth start to be noisy with less consistency. Nevertheless, there is a clear shift in most FPCs, justifying the daily recalibration of FPCA (also motivated by the probable mean shift that is not represented here).

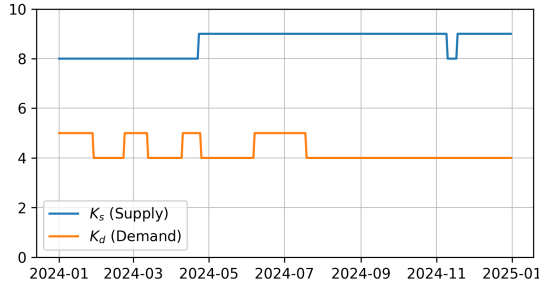


Figure 2.4: *Evolution of the number of functional principal components used for prediction across the testing period*

The first five and four supply and demand FPCs can be better interpreted on Fig. 2.6, which displays the effect a perturbation caused by FPCs on the mean curve for a snapshot of the dynamic FPCA at the middle of the test period. In particular for demand, FPC1 explains the quasi-totality of the variability and concerns the magnitude of the total demanded quantity. The contribution to the total variance beyond FPC1 is negligible. Subsequent FPCs were however selected because the elbow method was in this case less conservative than the threshold method. The additional consideration of their temporal consistency highlighted in 2.5 suggests their significance and goes against the hypothesis that these FPCs just capture noise. It is likely instead that they could capture the small scale variability in a domain of the curve where the intersection with the supply curve is likely to occur, and that is therefore of interest. For supply, FPC1 measures the overall supply level at any price, while FPC2 contrasts situations where there is higher supply at low price (below 100€/MWh) compared to high price (above 200€/MWh) and vice versa. FPC3 and FPC4 regard the distribution of the prices above (resp. below) 180€/MWh, and similarly contrasts situations when these prices are higher or lower. FPC5 is more difficult to interpret and regards more complex features of the price/volume distribution.

2.5.2 Curves forecasting

An example of curves (and clearing price) forecast for each of the nine curve-based models can be found in Appendix B.1. The average squared correlation R^2 and the squared correlation function $R^2(p)$ of the nine curves forecasting models are presented in Table 2.2 and Fig. 2.7, respectively. The naive model is clearly outperformed by all the other models for both sides. For demand, all models (except naive) perform almost equally: This is very likely due to the low demand elasticity on the Italian day-ahead market meaning that the quantity demand curve is more or less constant equal to the load forecast – predictor to which all models have access to. Concerning supply, the FPCA representation (**FPCA**) clearly outperforms that of Ziel and Steinert (2016)

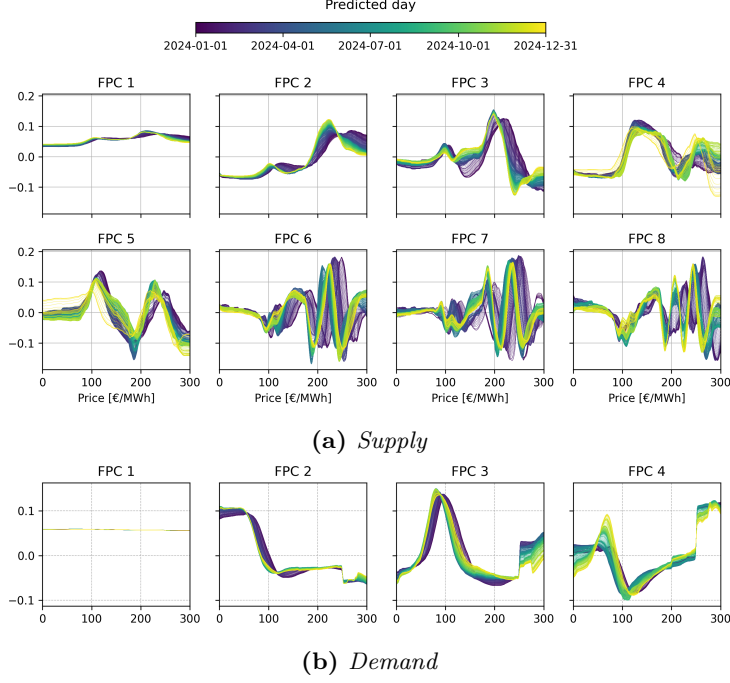


Figure 2.5: *Functional principal components across the test period*

(**ZST**), while the selected variant of the multivariate forecasting model does not show much impact, especially for **FPCA**. Looking at the entire squared correlation curve, we notice the performance drop of **ZST** models in the high prices range, which partly explains their lower average R^2 .

2.5.3 Clearing price forecast

The global clearing price forecasting performance is reported on Table 2.3 with the associated DM tests in Fig. 2.8a, for the four *price-based* and eight *curve-based* models. All models perform well compared to the naive approach. The DM tests reveal that all **ZST** models are significantly outperformed by all **FPCA**, regardless of the multivariate model variant considered. Besides, none of **ZST** models significantly outperform any of the price-based models, while all of them but one are significantly outperformed by at least one of the price-based models. The **FPCA** curve-based models instead are the best performers as all of them significantly beat all **ZST** and price-based approaches¹⁵. A key result is that the consideration of the cross-scores dependence (**VARX** models) appear to have a significant (positive) impact while the cross-hour dependence (full models) does not show any improvement, whether it be for **ZST** or **FPCA**. Instead,

¹⁵At the 5% level except for the univariate FPCA models vs **fARX** for which it is 10%.

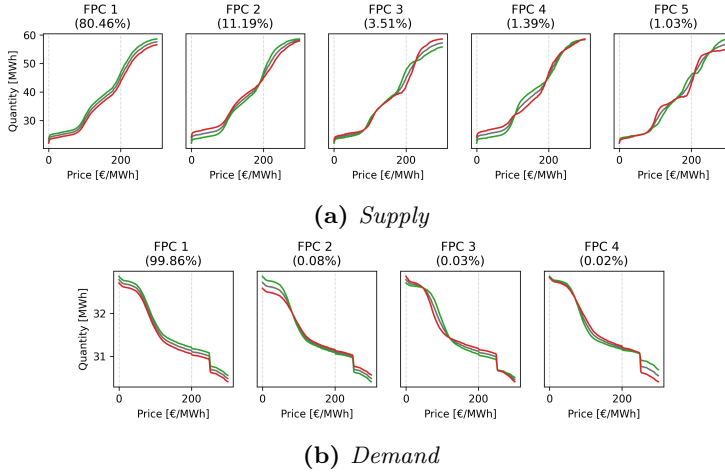


Figure 2.6: Visualization of the effects of the first 5 (for supply) and 4 (for demand) FPCs on the mean curve (in grey) estimated for prediction of 2024-07-01. The green curve corresponds to mean curve to which a multiple (20 for supply and 1 for demand) of the FPC is added, while the red curve corresponds to the mean curve to which a multiple of the FPC is subtracted. The number in parenthesis above each plot is the proportion of variance explained by the FPC.

for the price-based models, the cross-hour dependence is necessary to reach top performance.

Note the large performance gap between **ZST-ARX** and **ZST-VARX** (almost 1 €/MWh) compared to that observed between **FPCA-ARX** and **FPCA-VARX** (around 0.20€/MWh). This observation confirms a priori beliefs on the efficiency of the functional vs discretization representation method for capturing the statistical dependence that may exist between different portions of curves domain.

Fig. 2.9 shows the hour-level performance of each category's best model: **fARX** for price-based, **ZST-VARX** for **ZST** curve-based and **FPCA-VARX** for **FPCA** curve-based, while DM tests performed at the hour-level can be found in Fig. 2.8b. We understand that the **FPCA** model performs particularly well around mid-day, between 12 and 16, matching the period of simultaneous low-demand and high RES generation occasioning price drops, while the price-based model shows a slight advantage – confirmed by DM tests – during the night.

To understand more precisely the difference between the price-based and curve-based forecasts and visualise performance across the price range, we can refer to Fig. 2.10 where predicted prices are plotted against true prices for each model. Without much surprise, largest errors occur at the tails of the

	R^2_{supply}	R^2_{demand}
Naive	0.755	0.908
ZST-ARX	0.853	0.991
ZST-fARX	0.847	0.992
ZST-VARX	0.869	0.990
ZST-fVARX	0.860	0.991
FPCA-ARX	0.912	0.991
FPCA-fARX	0.916	0.992
FPCA-VARX	0.913	0.991
FPCA-fVARX	0.915	0.992

Table 2.2: *Curves forecasting performance measured with the average squared correlation.*

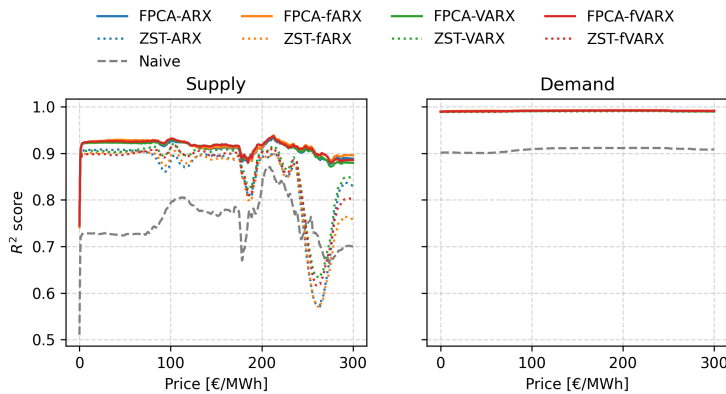


Figure 2.7: *Performance of curves prediction measured with the squared correlation function $R^2(p)$*

distribution, particularly on the left-side i.e. for prices below 100€/MWh. Note the two opposed biases between price-based and **FPCA** curve-based models as the first systematically overestimate prices below ≈ 50 €/MWh while the second tend to underestimate them. This observation suggests that considering an ensemble of price-based and curve-based models is likely to provide significant improvements.

2.5.4 Price probabilistic forecasting

The average CRPS of the clearing price probabilistic forecasting models are reported in Table 2.4 with the associated DM tests in Fig. 2.12. Again, all models outperform the naive strategy but the curve-based models based on **ZST** representation do not perform well compared to that based on **FPCA** and price-based models. While **FPCA** curve-based models perform globally better than price-based approaches, **fARX-QRM** strongly benefits from the ensemble averaging effect and only two of the curve-based models significantly outperform it at the

		MAE	rMAE	RMSE
<i>price-based</i>	Naive	11.34	1.000	17.24
	ARX	8.35	0.736	11.89
	fARX	7.97	0.703	11.37
	LEAR	8.24	0.727	11.65
<i>curve-based</i>	ZST-ARX	9.12	0.804	13.10
	ZST-fARX	8.91	0.786	12.92
	ZST-VARX	8.22	0.725	11.82
	ZST-fVARX	8.52	0.751	12.31
	FPCA-ARX	7.79	0.687	11.52
	FPCA-fARX	7.80	0.688	11.61
	FPCA-VARX	7.61	0.671	11.19
	FPCA-fVARX	7.66	0.676	11.51

Table 2.3: Clearing price prediction performance

		28D	56D	91D	182D	Avg.
<i>Price-based</i>	Naive-N	9.294	9.359	9.377	9.284	9.239
	fARX-N	6.214	6.189	6.186	6.207	6.147
	fARX-QRM	6.201	6.111	6.122	6.154	5.968
	fARX-CP	6.151	6.142	6.167	6.183	6.138
	fARX-IDR	6.807	6.663	6.529	6.518	6.347
<i>curve-based</i>	ZST-ARX	7.567	7.413	7.283	7.248	7.182
	ZST-fARX	7.351	7.234	7.121	6.930	7.040
	ZST-VARX	7.188	7.097	7.009	6.711	6.870
	ZST-fVARX	7.278	7.128	7.058	6.858	6.979
	FPCA-ARX	5.776	5.744	5.764	5.778	5.735
	FPCA-fARX	5.761	5.747	5.762	5.796	5.733
	FPCA-VARX	5.847	5.838	5.884	5.841	5.815
	FPCA-fVARX	5.928	5.954	5.975	5.944	5.923

 Table 2.4: Average CRPS for probabilistic clearing price forecasting across different calibration windows (in *Days*) and their ensemble obtained through vertical (probability) averaging

10% level regarding daily average errors. The top performers for point forecasting were the multivariate **FPCA** models, but for probabilistic forecasting, results suggest a slight advantage for univariate versions.

Fig. 2.11 shows the hour-level average CRPS of the best models of each category. As for point forecasting, the best curve-based model performs better in the early afternoon between 12 and 16 than the price-based approach. The performances are nearly equal for the rest of the day.

Looking at the calibration and resolution components (Fig. B.2 and B.3 left in Appendix B.2), we notice the poor reliability of the curve-based **ZST** and price-based **fARX-IDR** models. The first is strongly over-dispersed while the tails of the second are too light. **fARX-N** forecasts are reliable but show a poor resolution. The other models show comparable calibration – with a slight under-dispersion and clear skew towards lower prices – but heterogeneous

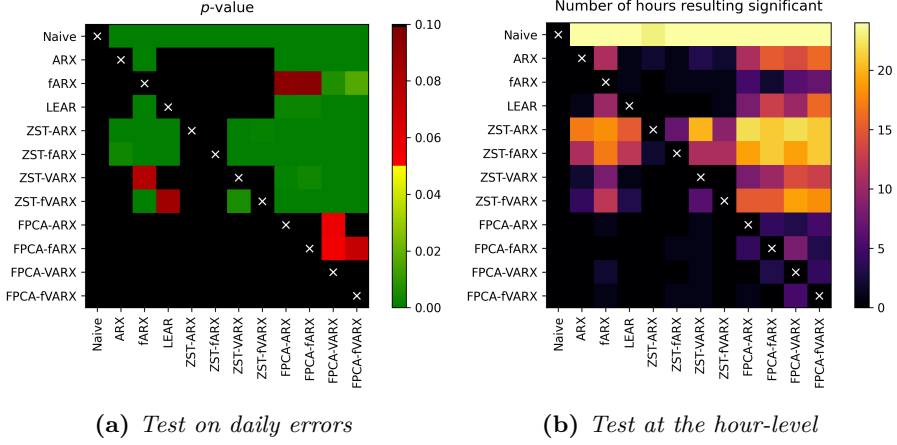


Figure 2.8: Results of the Diebold-Mariano test for the difference in price point forecasting performance performed (a) on average daily errors (b) at the hour level. The alternative hypothesis is that models on the x-axis outperform those on the y-axis (one-sided test).

resolutions: low for **fARX-CP**, intermediate for **fARX-QRM** and higher for **FPCA** models.

2.5.5 Computational efficiency

The computational time requirements for point and probabilistic forecasting models (Table 2.5 and 2.6) were evaluated on an Apple M2 Pro chip with a 12-core CPU and 16GB of RAM. Note that the **ARX** model could be made more efficient by parallelizing model fits across hours or vector components. The remaining models benefit from NumPy’s built-in multithreading for large matrix operations and therefore execute in a multicore fashion.

Price-based point forecasting models are definitely faster as they require fitting only 24 individual regression models, compared with the $24 \times K$ fits needed for curve-based models. Nevertheless, the overall computational cost remains low across all approaches. For probabilistic forecasting, the price-based normal (**N**), conformal prediction (**CP**), and isotonic distributional regression (**IDR**) models are highly efficient, whereas the quantile regression machine (**QRM**) is substantially more expensive. However, **QRM** is the only price-based method that was found competitive against curve-based models, which exhibit comparable computational efficiency.

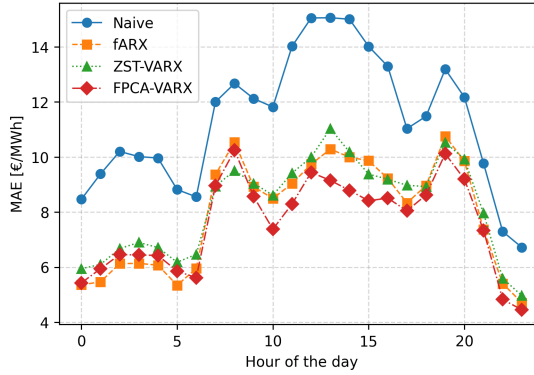


Figure 2.9: Mean absolute error of the three best performing models within each category (price-based and curve-based with either Ziel and Steinert (2016) or FPCA representation) for each hour of the day

	Wall time
ARX	0.2s
fARX	0.6s
LEAR	3s
ZST/FPCA-ARX	2s
ZST/FPCA-fARX	9s
ZST/FPCA-VARX	6s
ZST/FPCA-fVARX	16s

Table 2.5: Average computational time required for one daily iteration of price point forecasting. A daily iteration consists of (i) scaling/transforming the data & fitting the model to the last 364 days (ii) predicting the 24 prices of the next day.

2.6 Discussion and Conclusion

This work introduced a general framework for forecasting day-ahead merit-order curves, leveraging functional principal component analysis to efficiently represent a pair of supply and demand curves in a vector space and employing regularized vector autoregressive models for their prediction. The application to the Italian day-ahead market during the 2023-2024 period not only demonstrated the method’s effectiveness in forecasting merit-order curves but also in producing highly accurate clearing price point and probabilistic forecasts.

We tested four variations of our model, each treating the hourly time series as 24 independent daily time series – one for each hour – and differing in whether they accounted for (i) cross-dependence between hours and (ii) cross-dependence between components of the vector representation. Our findings indicate that accounting for cross-component dependence slightly improves clearing

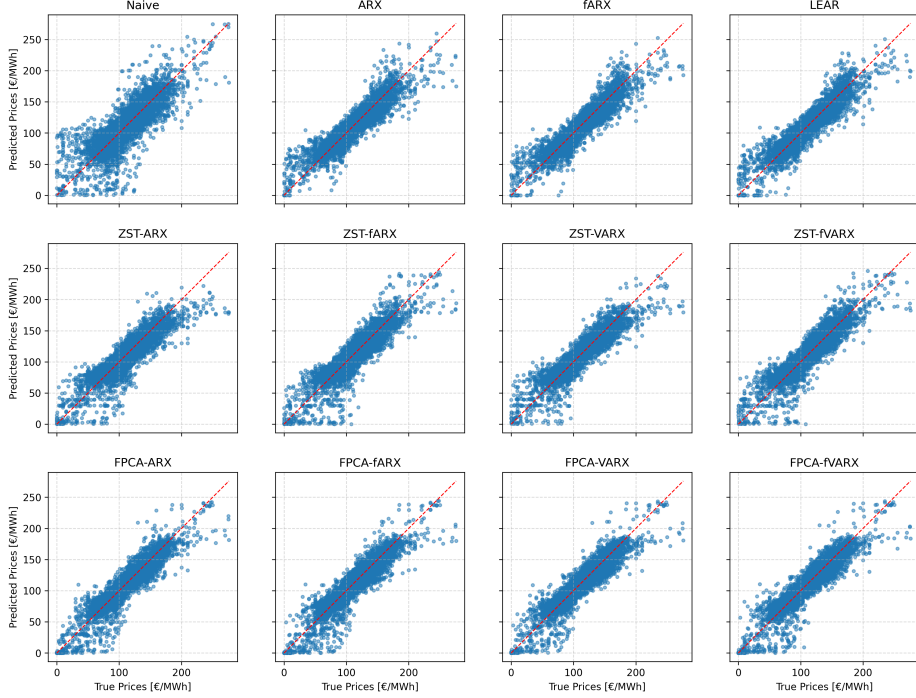


Figure 2.10: *Predicted vs actual clearing prices for all price-based and curve-based models. Each point represents one of the 8744 hourly observations in the test period. The dashed red line is the $y = x$ line.*

price forecasting performance, while cross-hour dependence does not yield significant gains.

Our curve-based model was rigorously compared with that of Ziel and Steinert (2016) – which share the same multivariate forecasting framework and differs only in the curves representation strategy – and state-of-the-art price-based models for clearing price forecasting. The results show that considering a functional principal component representation brings consequent improvements over a discrete representation and that our curve-based model – even in its simplest form – significantly enhances point and probabilistic clearing price forecasting performance, especially around mid-day when price occasionally drops due to high renewable generation and low demand. Specifically, our curve-based model reduces the mean absolute error by 5% and the continuous ranked probability score by 4%. If we restrict the comparison to the 10:00-17:00 hour range, the reduction is of 10% and 7%, respectively.

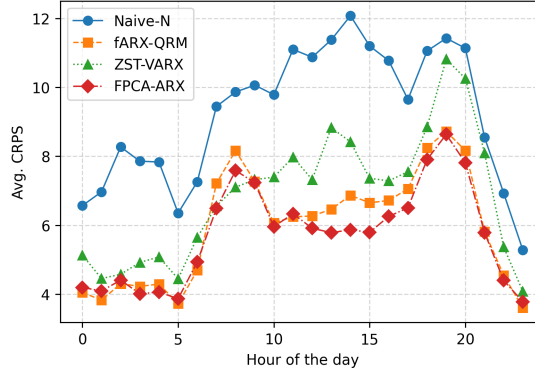


Figure 2.11: Average CRPS error of the two best performing models within each category, ensembled across the four time windows (price-based and curve-based with either Ziel and Steinert (2016) or FPCA representation) for each hour of the day.

	Wall time
N	<0.1s
QRM	5s
CP	<0.1s
IDR	<0.1s
ZST/FPCA	3s

Table 2.6: Average computational time required for one daily iteration of price probabilistic forecasting using a 182-day calibration window. A daily iteration consists of fitting the probabilistic model to the point forecasts and observed prices on the last 182 days and predicting the 24×99 quantiles for the next day.

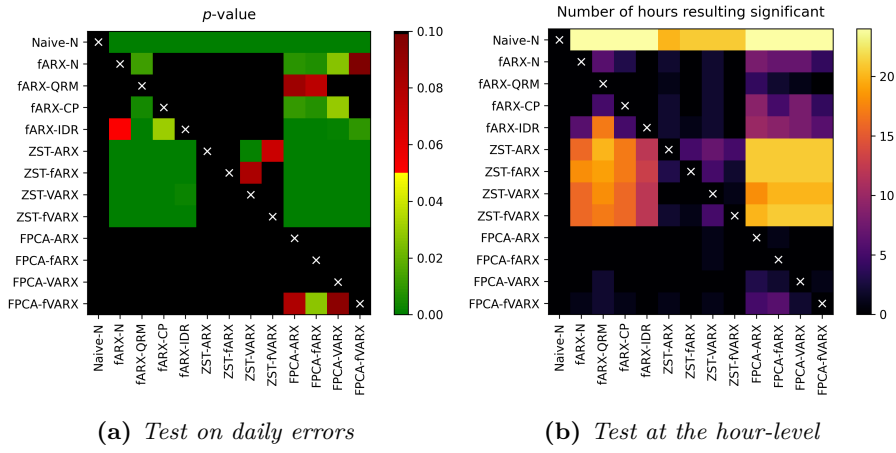


Figure 2.12: Results of the Diebold-Mariano test for the difference in price probabilistic forecasting performance as measured by the CRPS. The ensemble model obtained by averaging the four 28, 56, 91 and 182-days calibration windows is considered. The alternative hypothesis is that models on the x -axis outperform those on the y -axis (one-sided test).

CHAPTER 3

Strategic bidding in pay-as-bid reserve markets

In this chapter, we propose a probabilistic classification approach to model pay-as-bid ancillary services market (ASM) and derive an optimal bidding strategy for market participants. The proposed methodology is tested on the Italian ASM over the 2018-2022 period.

3.1 Introduction

3.1.1 Context

Power system management, often referred to as power system dispatching or only dispatching, is the process that SOs exploit to continuously ensure secure power system operations. It relies on provision of the so-called *ancillary services* which are a set of technical operations undertaken by generation units (and more recently consumption units) aiming at maintaining a constant balance between power generation and consumption at every node of the grid and resolving congestion issues. Without entering into technical details, most ancillary services simply correspond to an increase or decrease of power injection or withdrawal and are traded as energy supply or purchase offers characterized by a quantity in MWh and a price in €/MWh. In a traditional setting with

fully programmable generation units, dispatching is a low-cost activity, mainly dealing with network's faults and plants' failures, with limited opportunities for an actually competitive market for ancillary services. However, the diffusion of VRE – wind and solar generation essentially – is pushing SO to implement smarter strategies to decrease dispatching costs, as VRE's inherent variability is responsible of continuous supply-side system imbalances increasing the need for balancing services. Indeed, as a result of the ongoing energy transition, large volumes of power trading liquidity are slowly moving from energy-related markets to services-related ones, reflecting the growing relevance of when and where electricity is produced, rather than how efficiently (Li and Ho, 2022; Viola et al., 2024). According to Transparency Market Research (2024), in 2022 the ASM had a global value of \$8b in 2022, which is expected to grow in the next decade with an 8% compound annual growth rate, reaching \$17b in 2031. Europe owns the largest market share (40%), mainly because of its highly interconnected power network, which currently spans from Ukraine to Portugal as a single synchronized area. The expansion of ASM, together with the emergence of distributed energy resources (DERs) capable of providing ancillary services, represents a big opportunity for market operators. Both academics and practitioners agree on the possibility for diverse technologies to carve-out a relevant revenues stream from ASM dynamics (Merten et al., 2020; Baltputnis et al., 2024; S&P Global Ratings, 2021; McKinsey & Company, 2022), which however differ significantly from those of commercial (day-ahead and intraday) markets. Indeed, while commercial markets are explainable with economic fundamentals, ASM outcomes are obviously influenced by network technicalities, often at the expense of economic optimization (Lin and Magnago, 2017). SOs, acting as single buyers, may prioritize technical feasibility over economic efficiency to prevent blackouts, making ASM results difficult to predict. This creates challenges for market participants, especially newcomers with limited trading experience and low market power.

3.1.2 The ancillary services market design

ASM dynamics are strongly dependent on its market design, which refers to the combination of variables governing market's operations, typically determined through policy, regulatory and normative acts. In particular, these include: the type of services traded in the ASM and their technical characteristics, the remuneration mechanism and the trading timeline (Lin and Magnago, 2017). While a detailed discussion of ASM design is beyond the scope of this work, two main dispatching approaches can be identified: self-dispatching and central dispatching. According to Article 2 (17-18) of the Electricity Balancing Guideline (EBGL) (European Union, 2017), in a self-dispatching model (SDM), market operators determine the schedule of dispatchable GUs and CUs through a sequence of market-based processes. In contrast, in a central dispatching model (CDM), generation and consumption units submit cost and availability data to the SO,

who then optimize their schedules to minimize the total cost of securely meeting the demand at every node in the network. Therefore, in a CDM, power units can be considered as network assets controlled by the SO, while in a SDM they correspond to market participants freely determining their schedules according to their own objective function. Although the EBGL suggests the implementation of a self-dispatching model in Europe, peripheral regions, such as Italy, Ireland and Poland often adopt a central dispatching approach due to their low interconnection capacity with foreign regions, which leads to more frequent congestions and greater security needs. Currently under a SDM, the UK is even considering a return to a CDM (Department for Energy Security and Net Zero, 2024).

In CDMs – the model assumed in this work – GUs are required to offer their entire available capacity from the prospective load level (resulting from day-ahead and intraday market sessions) (i) to their nominal (maximum) power for upward regulation (UR) (ii) to their minimum technical power limit for downward regulation (DR). Each bid is defined by its price (expressed in €/MWh) and its quantity (MWh). Bids are either accepted or rejected by the SO based on bids’ competitiveness as well as prevailing network conditions. If accepted, only a portion of the offered quantity may be awarded, depending on specific network requirements at that time. An accepted bid may be either compensated at the bid price, under a PaB mechanism, or at the maximum price among all accepted bids, under a UP mechanism. Specifically, for downward bids, the GU compensates the SO for the purchased volume, whereas for upward bids, the GU receives a payment from the SO for the sold volume. Though much attention has been given in the literature to the UP setting, several ASM (such as Germany and Italy) are settled under a PaB scheme (Mazzi et al., 2018). It is important to note that market operators have a very limited inference capability over ASM’s results, because SOs’ decisions often depart from the economic optimum. As a result, its exact clearing mechanisms are not easily understood by market participants, since they follow a grid-specific optimization process that is largely unknown to them. Consequently, participants are not strongly incentivized to incorporate the ASM into their energy-trading strategies. Nevertheless, some GUs may choose to base their entire spot trading strategy on the ASM (Koechlin et al., 2024), capitalizing on its attractive PaB remuneration system and the typically higher selling prices it offers. This leads to high market concentration and significant barriers to entry, resulting in lower competition and reduced overall welfare.

While this issue is becoming more and more relevant, hindering relevant economic opportunities for newcomers such as DERs, the literature on this topic is still scarce. This work intends to fill this gap by offering a two-fold contribution. First, we propose a machine learning-based approach to model and predict individual bid acceptance in an ASM operating under a PaB remuneration mechanism. Second, we specify an ASM trading strategy applicable to any

market participant, based on the bid acceptance prediction model developed. Specifically, we address the complex problem of bid pricing, which – given that the offered quantity is automatically determined from the GUs’ binding program after day-ahead and intraday trading – remains the only strategic lever available to participants once day-ahead and intraday positions are settled. In fact, a trade-off must be found in bid competitiveness: a more competitive bid will increase the probability of acceptance but shrink potential profits, and vice versa. The proposed strategy is easy to implement and can be adopted independently of day-ahead/intraday trading schemes, making it particularly appealing to participants who wish to enhance their ASM performance without overhauling their global spot trading strategy.

3.1.3 Related works and research novelty

Predicting spot market results is closely related to electricity price forecasting (EPF). Indeed, most spot markets work under a single or double UP auction mechanism in which the results are entirely defined by the clearing price. As a consequence, most of the literature concerned with market results prediction has been focusing on EPF so far. Yet, a considerable number of ASMs, such as the Italian or the German one, work under a PaB auction design in which there is no such notion of price-based merit-order or clearing price. Indeed, market results in that case are characterized by individual bid acceptance status, following a more or less transparent clearing mechanism, as explained previously.

To the best of our knowledge, this study is the first to forecast ASM results at the individual bid level. In addition to filling this gap in the literature, our predictive ASM modeling serves two key objectives: (i) to provide market participants with a quantitative tool to support and improve their operational decision-making, and (ii) to shed some light on the opaque clearing mechanisms that characterize the ASM – particularly for CDMs, such as the one implemented in Italy. We achieve this by developing a dynamic random forest-based probabilistic classifier, enhanced with probability calibration techniques, that estimates the acceptance probability of any bid submitted to the market based on network conditions, participant and bid attributes, while employing XAI techniques to ensure the model’s interpretability.

A few works already tackled ASMs results forecast in a PaB design, but they actually concern an "aggregate" market outcome only. Moscetti et al. (2014) and Kraft et al. (2020) model an average of accepted bids’ price while Bâra and Oprea (2024) predict global power system imbalance which is the total quantity traded on the balancing phase of the ASM. Our approach therefore differs in the sense that we forecast *granular* market results, at the individual bid level. The Italian ASM was studied by Lisi et al. (2023) who introduced the concept of "cost-at-risk", defined as the α -quantile of the total costs incurred by the TSO over a future period of time, but also concern an aggregation of market results. Falvo et al. (2021) review the state-of-the-art resolution methods on

unit commitment and economic dispatch problems in the Italian ASM setting, but do not address the forecasting problem.

On the rich literature of EPF, Weron (2014) provided a complete review of the state-of-the-art, with a focus on spot markets. Already in 2014, Weron stated that modeling and forecasting of ASM prices is rather rare in the literature; this consideration is still valid today because no relevant literature has been developed around ASM price forecasting, despite a broad consensus on the impact of VREs and smart grids diffusion on the growth of services markets. According to Weron, this is linked to the characteristics of ASM prices that own particular features, such as high variability and frequent spikes, that makes their modeling with classic statistical approaches more challenging. We believe that the prevalence of PaB designs and CDM among ASMs and the more-technical nature of ASM trading is also a reason. To address price volatility and spiking behaviors, Wang et al. (2014, 2018) propose using mean-reverting jump diffusion models and Markov regime-switching models. These incorporate spike-related regimes into deterministic and stochastic price components and are shown to capture both structural shifts and sudden changes in market dynamics, caused by varying grid conditions. Computational intelligence approaches also appear efficient in modeling volatile and spiky regimes: support vector machines, random forest and neural networks were explored in Cardo-Miota et al. (2023); Giovanelli et al. (2018); Kraft et al. (2020), compared to traditional statistical models. In particular, Giovanelli et al. (2018) and Kraft et al. (2020) both observe that neural networks outperform the other approaches, as they are able to capture the fast-changing nature of ASM. It is worth mentioning Merten et al. (2020), that derive ASM bids acceptance probability in a UP design from a probabilistic forecast of the marginal price, obtained by bootstrapping various statistical and machine learning point prediction models. They conclude that computational intelligence approaches do not always provide better performance, and come with the cost of poorer interpretability.

Regarding the design of bidding strategies in ASM, prior research has largely focused on either theoretical market modeling using agent-based models or stochastic optimization for specific categories of market participants.

For instance, Siddiqui et al. (2001) analytically solve the optimal clearing quantity and price in a theoretical ASM setting under a UP scheme, albeit with simplifying assumptions that limit real-world applicability. Federico and Rahman (2003), study how bidding behavior shifts when moving from a UP to a PaB scheme under extreme cases (perfect competition and perfect collusion), providing policy-relevant insights but not addressing bid-level optimization under realistic conditions. Swider and Weber (2007) do propose an optimal bid pricing method under a PaB mechanism that is structurally similar to our own. They model acceptance probability based on bid price, quantity, and an assumed distribution of competitor bids. However, their method is not data-driven and relies on strong parametric assumptions. Furthermore, they do not validate their

proposed bidding strategy using actual market simulations.

Several studies propose stochastic optimization approaches to ASM bidding using mixed-integer linear programming, including Wang et al. (2017); Mazzi et al. (2018); Maneesha and Swarup (2021); Bovera et al. (2023); Li et al. (2023); Lokhande et al. (2023); Chen et al. (2024). These works often aim to jointly optimize the bidding strategy across the different spot markets (day-ahead, intraday, and AS markets), and are typically tailored to specific market participants such as DERs aggregators, virtual power plants or battery energy storage systems. Our approach differs in that it focuses exclusively on the ASM bidding strategy, after the clearing of day-ahead and intraday markets. At this stage, the quantities offered for upward and downward reserve are already determined, and participants need only to optimize the corresponding bid prices. As such, our method is complementary to those of the previously mentioned works, as pre- and post-day-ahead/intraday-clearing strategies can co-exist. However, most of these prior works rely on overly simplified models of ASM behavior – an important limitation that our data-driven framework seeks to overcome.

All of the aforementioned studies share a common limitation: they do not implement data-driven models to predict ASM results at the bid level, nor do they learn optimal bidding strategies directly from market data. This work aims to bridge the gap between research focused on data-driven market forecasting and that centered on the design of optimal bidding strategies. Specifically, we demonstrate how an accurate and granular forecasting model can be effectively transformed into a tool for optimal bid pricing. Our contributions may be concisely summarized as follows:

1. We introduce the first bid acceptance prediction model for ASM results in a PaB auction format
2. We propose a novel data-driven optimal pricing strategy for PaB ASM participants

In addition, a tabular view of the comparison of our work to the existing literature is presented in Table 3.1.

The remainder of the paper is organised as follows: Section 3.2 describes the dataset used in the analysis while Section 3.3 details the methodology. In particular, subsection 3.3.1 concerns the bid acceptance prediction model and subsection 3.3.2 the optimal bidding strategy. Section 3.4 presents the results and finally, followed by Sections 3.5 and 3.6, which provide a discussion and concluding remarks, respectively.

Reference	Bid-level Modeling	PaB Design	Data- Driven	Bidding Strategy
Moscetti et al. (2014)		✓	✓	
Kraft et al. (2020)		✓	✓	
Bâra and Oprea (2024)		✓	✓	
Lisi et al. (2023)		✓	✓	
Wang et al. (2014, 2018)			✓	
Cardo-Miota et al. (2023)			✓	
Giovanelli et al. (2018)			✓	
Merten et al. (2020)	✓		✓	✓
Siddiqui et al. (2001)	✓			✓
Federico and Rahman (2003)	✓	✓		✓
Swider and Weber (2007)	✓	✓		✓
Wang et al. (2017); ...	✓	✓		✓
This work	✓	✓	✓	✓

Table 3.1: Tabular view of contributions on ASM modeling

3.2 Data

3.2.1 Target variable

The dataset under study includes all¹ 5,751,703 bids submitted by the 131 GUs located in the *North* bidding zone to the Italian ASM – *Mercato dei Servizi di Dispacciamento* (MSD) *ex-ante* – over the 2018-2022 period. The data were collected from GME’s FTP server.

The target variable of our analysis is bid *acceptance* or its probability. Naturally, the bid price and quantity are introduced as covariates. Beyond bid-related data, additional factors that may describe the grid context in which the GUs operate, or that may directly influence the acceptance of bids, have also been collected. They are divided in four categories: *structural*, *exogenous*, *calendar* and *market*² data. In our analysis, they will be considered as covariates.

3.2.2 Structural data

Structural data covers all boundary conditions that characterize the operations of the power system, defining a specific context in which system operators and market participants are called to operate. We divide them in three groups: Generation, load and reserve capacity (group A), net transfer capacity (group B), and scheduled commercial exchanges (group C). The data have been collected

¹It has been chosen to exclude from the analysis the units participating to the ASM within the scope of the so-called UVAM project, as they are subject to different rules. For a better understanding of the choice please refer to Gulotta et al. (2020).

²Which also includes the bid price and quantity.

from ENTSO-E (2025), except for reserve capacity which has been calculated following the methodology outlined in Bovera et al. (2021). Data are referred to the last available forecast update or to the most recent market clearing results. Related variables are listed in Table 3.2.

Group	Variable	Source	Resolution	Unit
A	Load	ENTSO-E (2025)	hourly	MW
	<i>Generation Per Technology</i>			
	Biomass	ENTSO-E (2025)	hourly	MW
	Fossil Coal	ENTSO-E (2025)	hourly	MW
	Fossil Gas	ENTSO-E (2025)	hourly	MW
	Fossil Oil	ENTSO-E (2025)	hourly	MW
	Hydro	ENTSO-E (2025)	hourly	MW
	Photovoltaic	ENTSO-E (2025)	hourly	MW
	Wind	ENTSO-E (2025)	hourly	MW
	Water Reservoir	ENTSO-E (2025)	weekly	MWh
	<i>Reserve Capacity</i>			
	Secondary Reserve (upward)	Bovera et al. (2021)	hourly	MW
	Secondary Reserve (downward)	Bovera et al. (2021)	hourly	MW
	Tertiary Reserve (not spinning, upward)	Bovera et al. (2021)	hourly	MW
	Tertiary Reserve (total, upward)	Bovera et al. (2021)	hourly	MW
	Tertiary Reserve (total, downward)	Bovera et al. (2021)	hourly	MW
B	<i>Net Transfer Capacity</i>			
	North ↔ Center North	ENTSO-E (2025)	hourly	MW
	North ↔ Slovenia	ENTSO-E (2025)	hourly	MW
	North ↔ Switzerland	ENTSO-E (2025)	hourly	MW
	North ↔ Austria	ENTSO-E (2025)	hourly	MW
C	<i>Commercial Exchanges</i>			
	North ↔ Center North	ENTSO-E (2025)	hourly	MW
	North ↔ Slovenia	ENTSO-E (2025)	hourly	MW
	North ↔ Switzerland	ENTSO-E (2025)	hourly	MW
	North ↔ Austria	ENTSO-E (2025)	hourly	MW
	North ↔ France	ENTSO-E (2025)	hourly	MW

Table 3.2: *Structural data*

Due to the large number of variables collected, we opted to reduce the dimensionality of the structured data and summarize their key information using PCA for each of the three groups, retaining two principal components per group (Table 3.3).

3.2.3 Exogenous data

Exogenous data represents external factors that, while originating from markets other than the ASM, have a direct or indirect influence on it. For the purposes of this study, three key exogenous variables pertaining to the day-ahead market have been considered:

- **Volume of energy purchased and sold** (PurchMGP and SellMGP) on day-ahead market; these volumes indicate the load level of the power sys-

Component	Group	Variance Explained	Interpretation
IN_PC1	A	34%	Captures the overall scale of generation, load, and reserve capacity.
IN_PC2	A	16%	Highlights the relative contribution of intermittent renewable sources (solar and wind) versus conventional generation (oil, coal, hydro).
CT_PC1	B	34%	Measures whether the aggregate NTC between the North and neighboring zones is globally high or low.
CT_PC2	B	22%	Distinguishes contexts where NTC has a North-South rather than a South-North orientation.
SC_PC1	C	29%	Reflects the pattern of commercial exchanges, where high values denote situations where North imports from the south and exports to foreign countries, and low values when North imports from foreign countries and exports to the south.
SC_PC2	C	20%	Similar to SC_PC1 but focused on exchanges with the South and Slovenia.

Table 3.3: *Group-specific principal components of structural data and their interpretations*

tem, which usually corresponds to a specific configuration of power flows along the network.

- **Marginal technology** (**MargTech**) fixing the day-ahead market price; this variable is a good indicator of power system conditions, not only in terms of generation-demand equilibrium, but also regarding the available production flexibility.

3.2.4 Calendar data

Calendar information was incorporated with the following variables:

- **Hour** (**Hour**) of the day, treated as a numerical variable.
- **Day type** (**Workday**), differentiating between standard working days and non-working days³.
- **Solar declination angle** (**DeclAngle**), which is the angle between the sun's rays and the Earth's equatorial plane, varying throughout the year and encoding therefore the yearly seasonality.
- **Solar hour angle** (**SolarAngle**), which expresses the local solar time, encoding the daily seasonality.

3.2.5 Market data

These data are the bid and GU attributes:

³Saturdays, Sundays and bank holidays.

- **Bid price** (`PriceDiff`): Which is introduced through its normalised difference with the day-ahead market price, considered as the reference price, to have a normalised measure of bid competitiveness and avoid non-linear effects due to price inflation. Note that this covariate plays a special role in our analysis as it is the only parameter the operator can directly control to influence bid acceptance.
- **Bid quantity** (`Qty`): The upward and downward offered quantity in MWh. The upward (UR) quantity is equal to the difference between the nominal and the prospective operating capacity while the downward (DR) quantity is equal to the difference between the prospective operating and the technical minimum capacity.
- **Binding program** (`BP`): This variable indicates the prospective operational status of a unit before entering the ASM, following the prior day-ahead and intraday market sessions. It is equal to the ratio of the prospective operating capacity over the nominal capacity. A value of 0 indicates the GU is switched off (and consequently has no downward reserve) while a value of 1 indicates it is operating at its nominal capacity (and therefore has no upward reserve). It somehow carries the information of the bid quantity but relative to the nominal capacity of the power unit.
- **Unit location** (`Prov`): The location of each unit, represented by the province (second administrative level in Italy).
- **Generation technology** (`Tech`): Generation technology of the unit with six different categories used (see Table 3.4).

Technology	Type	GUs count
Fossil gas	Fossil	45
Fossil hard coal	Fossil	7
Thermoelectric	Fossil	7
Hydro pumped storage	Hydro	24
Hydro run-of-river and poundage	Hydro	14
Hydro water reservoir	Hydro	37

Table 3.4: *Detail of the `Tech` variable containing the GUs technology.*

The list of covariates used in the model is summarized in Table 3.5.

Name	Type	Description
IN_PC1	Numerical	First principal component for generation, load and reserve capacity
IN_PC2	Numerical	Second principal component for generation, load and reserve capacity
CT_PC1	Numerical	First principal component for NTC
CT_PC2	Numerical	Second principal component for NTC
SC_PC1	Numerical	First principal component for scheduled commercial exchanges
SC_PC2	Numerical	Second principal component for scheduled commercial exchanges
PurchMGP	Numerical	Volume of energy purchased on day-ahead market
SellMGP	Numerical	Volume of energy sold on day-ahead market
MargTech	Categorical	Marginal technology fixing the day-ahead market price
Hour	Numerical	Hour of the day
Workday	Categorical	Day type, differentiating between working and non-working days
DeclAngle	Numerical	Solar declination angle
SolarAngle	Numerical	Solar hour angle
PriceDiff	Numerical	Normalised difference between the bid and the day-ahead market price
Qty	Numerical	Bid quantity
BP	Numerical	Binding program
Prov	Categorical	Province where the unit is located
Tech	Categorical	Unit generation technology

Table 3.5: *List of the covariates included in the models. Except for the unit characteristics **Prov**, **Tech** and the day type **Workday**, all variables have an hourly resolution.*

3.3 Methodology

A graphical summary of the methodology can be found in Figure 3.1.

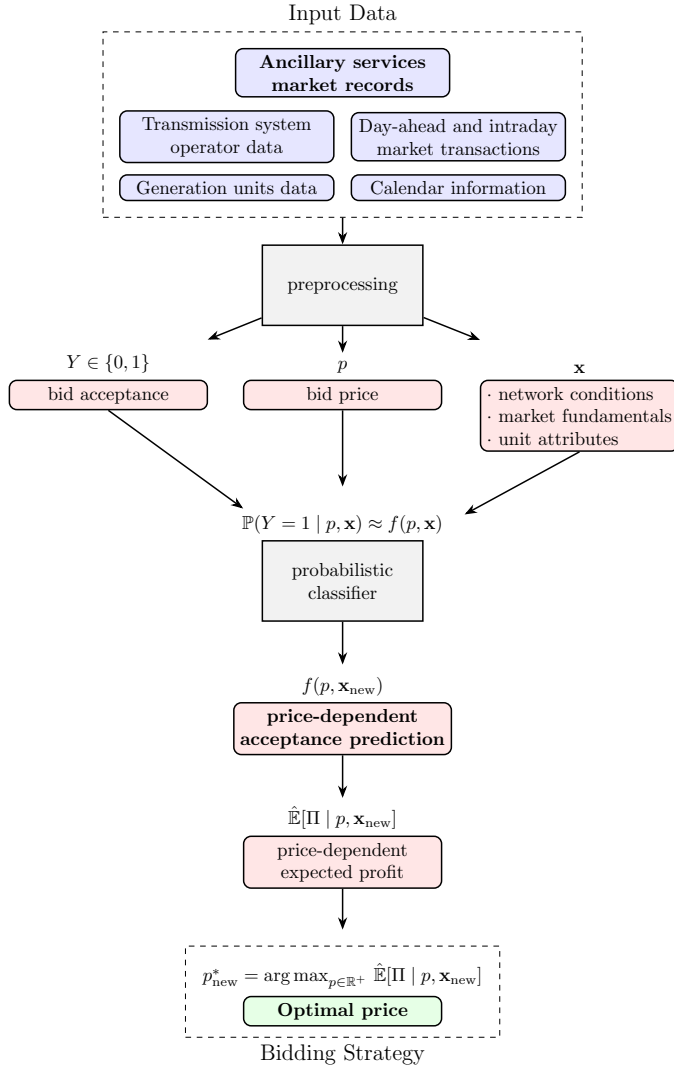


Figure 3.1: Flowchart diagram summarizing the proposed methodology

3.3.1 Bid acceptance prediction

Given the distinct characteristics and acceptance mechanisms of upward (UR) and downward (DR) bids, the dataset is divided into two separate subsets: one

for upward bids and another for downward bids. For each bid type, separate models are developed. As a preprocessing step, numerical variables are scaled and categorical variables are one-hot-encoded. The columns of the resulting design matrix are designated as *features*. The characteristics of the two preprocessed datasets are presented on Table 3.6.

	UR	DR
Number of observations	3,574,411	2,177,292
Number of features	67	67
Positive class frequency (acceptance rate)	3.6%	15.9%

Table 3.6: *Binary classification settings*

Classification model

The probabilistic classification pipeline we propose is based on *random forests* (**RF**) (Breiman, 2001), where the bid acceptance probability is estimated by averaging the binary outputs of the individual trees in the forest. This choice, among all possible classification methods usually considered for supervised learning problems on structured data, is motivated by the results of the study carried by Bovera et al. (2019), in which the authors perform a review of possible different statistical learning techniques to tackle this specific problem. The hyperparameters of the model (Table 3.7) were tuned on the out-of-bag sets of the model trained on 2019.

Hyperparameter	Value	Description
n_estimators	100	Number of trees in the forest
max_features	$\sqrt{r}$	Number of features to consider when looking for the best split
min_samples_split	2	Minimum number of samples required to split an internal node
min_samples_leaf	1	Minimum number of samples required to be at a leaf node

Table 3.7: *Hyperparameters of the RF model. r is the total number of encoded features ($r = 67$ in our case).*

Dynamic modeling approach

To simulate the behaviour of the classifier and to assess its performance in a live setting, the model is updated and tested on a monthly basis. We tested it on every month from 2019 to 2022, considering a sliding window of 12 months backward for the training set (see Fig. 3.2). This is a common practice in the EPF literature, identified as *recalibration*⁴ in Lago et al. (2021). It also allows to properly cross-validate any model selection procedure in a setting where

⁴We however here avoid this term not to make confusion with probability calibration.

observations are time-dependent. It is indeed very close to the *time-series cross validation* procedure as defined by Hyndman and Athanasopoulos (2021) with the difference that the training set is restrained to 1 year backward. Having a bandwidth of at least one year is necessary to take advantage of yearly seasonal effects while keeping it narrow enough allows to quickly adapt to changes in the acceptance mechanisms, which can be of utmost importance in market with fast-changing dynamics. Another motivation for choosing a narrow window is the computational cost. The resulting time-dependent model is called *dynamic model*. Note that the dynamic model's performance was compared to the *static models* obtained by fitting the random forest on specific years (identified with the suffix corresponding to the training year added to the model acronym: e.g. **RF2019**).

Probability calibration

In the realm of probabilistic classification/forecasting, a classifier is said to be *well-calibrated* if the predicted class probabilities are reliable estimates of the true class conditional probabilities (DeGroot and Fienberg, 1983). Some classifiers naturally have this property but some other, especially non-parametric and maximum margin methods, do not. The class probabilities estimated by random forests, though being overall good estimates of the true posterior probabilities compared to other methods, can be subject to slight distortions, having difficulty in predicting probabilities close 0 or 1, which is known as *under-confidence* (Niculescu-Mizil and Caruana, 2005). This distortion can however be efficiently corrected with *probability calibration* methods such as Platt scaling (Platt, 2000) or isotonic regression (Zadrozny and Elkan, 2002), essentially consisting in a univariate regression of the probability estimates on the observed class frequencies using binning. Attention should be given to the fact that the data subset used to train the model and the one used to fit the calibrator should be disjoint⁵, naturally justifying a cross-validation approach. In our case, we use isotonic regression as it is known to perform better when the sample size is large (Niculescu-Mizil and Caruana, 2005) and we use the probability predictions (and not the fitted values) of the dynamic model in a sliding window of 12 months backward to fit the calibrator. In this way, we manage to have a dynamic probability calibration that adapts rapidly to possible market regime changes but still makes use of enough data to estimate the correction to bring to the model outputs. Note that this design allows us to get calibrated probability predictions from 2020 only.

⁵To avoid calibrating the classifier on the fitted probabilities which can be distorted because of overfitting.

Comparison with benchmark methods

Two optimized benchmark classification methods have been considered for comparison with the proposed method.

The first benchmark is a purely price-driven model, inspired by the "marginal price" approach that was used by Rancilio et al. (2022) to model the balancing market results. This model, referred to as *historical prices* (**HP**) model, employs a non-parametric Bayes classifier with the bid price as a unique covariate. The prior probabilities are the relative frequencies observed in the data, while the price distributions conditional on the acceptance or refusal (i.e. the likelihood) are estimated non-parametrically using the empirical price distributions in each class from the training period. This approach serves as a baseline to assess the added value of incorporating additional covariates beyond just price.

The second benchmark is a *generalised linear model* (**GLM**) with a Bernoulli response, commonly known as *logistic regression* (Cox, 1958). This model includes all covariates without introducing interaction terms or non-linear transformations of the original variables. The purpose of such a benchmark is to assess the degree of non-linearity required to effectively model the acceptance.

Note that we consider the dynamic versions of the two benchmark models, as for **RF**.

Performance assessment and model explainability

The prediction performance is assessed in two ways: quantitatively with a scoring rule and qualitatively with reliability diagrams (Murphy and Winkler, 1977). From the perspective of binary classification, given the high unbalance of the two classes caused by a low proportion of accepted bids (especially for the upward bids), the emphasis is put on the precision-recall (PR) metrics (Saito and Rehmsmeier, 2015). In our context, precision is the probability that a bid predicted as accepted is actually accepted, recall is the probability that an accepted bid is predicted as accepted and F_1 -score is the harmonic mean of precision and recall. Since we want to evaluate the probability prediction, and not just the binary output obtained for a certain probability threshold, we use the version of the area under the receiver operating characteristic curve (AUC-ROC) suited to unbalanced classes, the area under the precision-recall curve (AUC-PR). We estimate the AUC-PR with the average precision method⁶ of Boyd et al. (2013). Though the AUC-PR is not a proper scoring rule⁷ and is therefore usually not recommended for evaluating probabilistic forecasts, it remains a popular and interpretable performance metric when the misclassification costs are undefined

⁶We are aware of the criticisms formulated by Flach and Kull (2015) but we believe they are mitigated in our case because of the very large sample size.

⁷One could argue that the Brier score (Brier, 1950), which is a proper scoring rule, instead should be used. But the drawback is double: The Brier score lacks interpretability and moreover gives the same importance to each class which is an undesired property in case of class imbalance.

or when the target is the decision function output, and when there is a strong class unbalance.

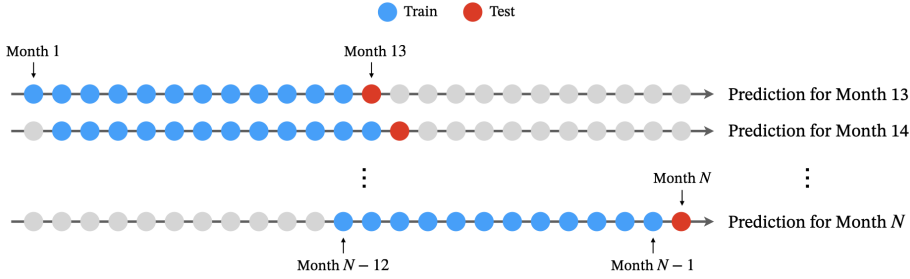


Figure 3.2: *Illustration of the dynamic modeling approach. The blue dots designate the months used for fitting the model (training set), the red dots the ones for which predictions are obtained (test set).*

Although random forests are often classified as black-box models – since their behavior cannot be easily interpreted by inspecting their fitted parameters⁸ – the application of XAI techniques can nonetheless provide some degree of interpretability. We use the well-established SHapley Additive exPlanations (SHAP) framework to interpret the model. SHAP (Lundberg and Lee, 2017) is a *local* interpretability technique, meaning it estimates the effect of one or more covariates on the model’s output for a single observation. This contrasts with *global* interpretability techniques, which aim to explain the overall influence of covariates on the model’s predictions (Molnar, 2020). However, local explanations can be extended to a global perspective by aggregating them across many observations and analyzing the distribution of the effects. We briefly introduce the intuition behind SHAP here, but refer the reader to the relevant chapter in Molnar (2020) for a more comprehensive explanation.

SHAP values are based on concepts from cooperative game theory, specifically Shapley values (Shapley, 1951), which provide a way to fairly allocate the total *payout* of a game among its *players*. In the context of XAI, the game is the prediction task, the payout is the difference between the model’s prediction and a baseline (e.g. the average model output) and the players are the input features (or covariates). Intuitively, SHAP computes the marginal contribution of a feature to the model’s prediction by looking at the model’s prediction shift when adding or removing that feature in the prediction process. While SHAP relies on strong theoretical results, it has the drawback of being computationally expensive. Various approaches to address this issue have been proposed, including TreeSHAP (Lundberg et al., 2020) that allows to compute SHAP val-

⁸While variable importance measures can highlight which covariates contribute most to the classification process, they don’t explain how these variables influence the model’s predictions.

ues in a polynomial instead of an exponential time for tree-based methods such as gradient boosted trees (Friedman, 2001) or random forests.

Given our large sample size, to keep computational times reasonable, we computed the SHAP values of all observations using TreeSHAP with the "tree-path dependent" estimation method, which is a sort of faster approximation of exact SHAP values. The quality of the approximation was tested on small subsets and was found to be satisfactory. We then derived (i) global feature importance by averaging the absolute SHAP values across all observations (ii) global feature effects by examining the relationship between feature values and their corresponding SHAP values.

3.3.2 Optimal bidding strategy

In this section, we explain how a market participant can make use of the model to price its bids on the ASM, in order to maximize its expected profit. We first explain how an optimal bid price can be derived from the bid acceptance probability estimate and the GU's marginal cost, after which we test the subsequent pricing strategy against the one that was actually adopted by two (arbitrarily chosen) gas-fired GUs (see Table 3.8), running a market simulation over the whole 3-years test period.

Market ID	Technology	Province	Capacity	Tested for
UP_LA_CASELL_2	Fossil-gas	Piacenza	370 MW	UR
UP_VADOTERM_5	Fossil-gas	Savona	780 MW	DR

Table 3.8: *Characteristics of the two GUs considered for the market simulation*

Optimal bid price

Let Y be the Bernoulli random variable denoting the acceptance of a randomly picked downward or upward bid on the market. Let p be the bid price and c the marginal cost that the submitting GU assumes at that moment. Following the pay-as-bid scheme, if an upward bid is accepted, the *marginal profit* associated to the transaction is simply obtained by subtracting the marginal cost c to the selling price p . In the case of downward bids, the profit is implicit: if accepted, the GU has to pay p to the TSO but saves the marginal cost c of the quantity that it will finally not have to produce, so the profit is obtained by subtracting the purchase price p to the marginal cost c . Formally, the marginal profit can be represented as a random variable:

$$\Pi = \begin{cases} \begin{cases} p - c & \text{if UR} \\ c - p & \text{if DR} \end{cases} & \text{if } Y = 1 \\ 0 & \text{if } Y = 0 \end{cases} \quad (3.1)$$

In our case, we developed a probabilistic classifier which models the conditional distribution of Y given the bid price p and the remaining $r - 1$ covariates designated by the vector $\mathbf{x}$. Consequently, we introduce the conditional expectation of the marginal profit:

$$\mathbb{E}[\Pi \mid p, \mathbf{x}] = \begin{cases} (p - c)\mathbb{P}(Y = 1 \mid p, \mathbf{x}) & \text{if UR} \\ (c - p)\mathbb{P}(Y = 1 \mid p, \mathbf{x}) & \text{if DR} \end{cases} \quad (3.2)$$

At the moment of bids submission, c along with $\mathbf{x}$ are known and fixed. Moreover a plug-in estimate $\hat{\mathbb{E}}[\Pi \mid p, \mathbf{x}]$ of the expected profit can be derived from the classifier probability output $f(p, \mathbf{x}) \approx \mathbb{P}(Y = 1 \mid p, \mathbf{x})$. Therefore a market player can estimate its optimal price p^* as maximizing the estimate of the conditional expectation of the marginal profit:

$$p^* \approx \arg \max_{p \in \mathbb{R}^+} \hat{\mathbb{E}}[\Pi \mid p, \mathbf{x}] \quad (3.3)$$

In practice, we apply a Savitzky-Golay filter (Savitzky and Golay, 1964) to the probability curve $p \rightarrow f(p, \mathbf{x})$ to de-noise the classifier output and obtain a smooth profit function to maximise. Additionally, the search range for p is restricted relative to the day-ahead market price p_{DA} by limiting `PriceDiff` to typical ranges observed in the data and therefore known to the model. For UR, we restrict `PriceDiff` $\in [0, \frac{1}{3}]$, implying $p \in [p_{\text{DA}}, 2p_{\text{DA}}]$, and for DR, `PriceDiff` $\in [-\frac{1}{2}, 0]$, giving $p \in [\frac{1}{3}p_{\text{DA}}, p_{\text{DA}}]$.

Market simulation

To validate the efficiency of a bidding strategy based on the optimal pricing process described above, we simulate the adoption of this strategy by the two GU and compute their resulting profits over the 3-years test period, using the elaborated model to simulate market results. We then compare these profits to those cleared under the GU's actual bidding strategy (i) using real market results (ii) using the model's simulated market results. While this comparison serves to validate the proposed pricing strategy, the comparison between (i) and (ii) also provides an "economic" validation of the model, complementing the statistical validation. Indeed, if expected market results, simulated through the model based on actually presented bids, are similar to real market results, then this proves that acceptance probabilities provided by the model are reliable.

Marginal costs are approximated as constant function of the GU's power output, estimated on an hourly basis through the GU's day-ahead market bid price,

under the assumption that fossil-fuel plants bid on average at their marginal cost. These estimates were found to be consistent with an other proxy computed from the gas price, carbon emissions price and average efficiency of gas-fueled generators. Outliers below 10€/MWh were removed and missing values were imputed using linear interpolation.

As noted in section 3.2.1, when a bid is accepted, only a fraction of the offered quantity may be awarded. Our motivation for disregarding this aspect of market results in the prediction task is that it was found to exhibit almost no dependence with any of the covariates, contrarily to the acceptance. In other words, the ratio of offered quantity that is actually awarded when a bid is accepted, appears to be completely random. We therefore deemed reasonable to simulate the awarded quantity of a bid simulated as accepted as the product between the offered quantity and an awarded quantity ratio sampled from its empirical marginal distribution⁹.

In this setting, we can simulate market results and therefore resulting profits resulting as follows:

For every submitted bid $p, \mathbf{x}$:

1. Sample acceptance y from a Bernoulli with parameter $f(p, \mathbf{x})$
2. Sample awarded quantity Q as indicated above
3. Compute the profit resulting from the bid as:

$$\text{profit} = \begin{cases} Q(p - c)y & \text{if UR} \\ Q(c - p)y & \text{if DR} \end{cases} \quad (3.4)$$

We perform $N = 1000$ Monte Carlo simulations, aggregating profits at the quarterly level to obtain both an expected value and a prediction interval for profits in each quarter.

3.4 Results

3.4.1 Comparison with benchmark methods

The comparison of the performance of the **RF** model vs benchmarks (**HP** and **GLM**) for each testing year of the dataset is pictured on Fig. 3.3. Whether it be for downward (3.3b) or upward (3.3a) bids, **RF** and **GLM** largely outperform **HP**, highlighting the relevance of the covariates considered beyond the price. Interestingly, while **RF** and **GLM** performances are comparable in the case of DR, we observe that **RF** considerably outperforms **GLM** in the case of UR. This result may underline a quasi-linear influence of the covariates on the acceptance probability in the case of DR, and a more complex non-linear effect for UR.

⁹Which was separately estimated for each year.

Moreover, we observe higher performance for DR (around 0.7) for which the prior acceptance probability is in the 15-20% range across years than for UR (around 0.45), for which it falls in 2-5%. An additional observation could regard the performance stability for UR, opposed to the performance decrease for DR from 2021, suggesting a variation in market dynamics affecting DR only.

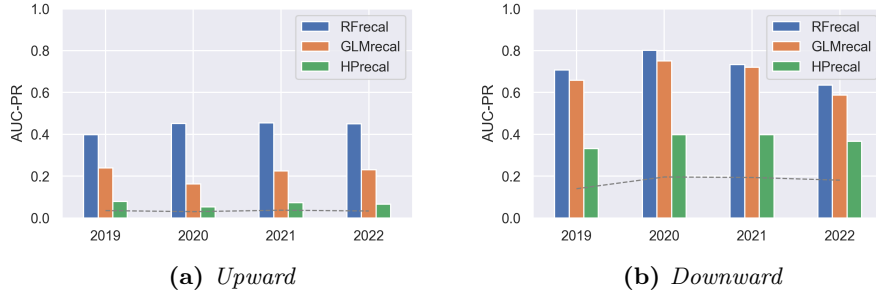


Figure 3.3: Comparison of the model performance vs benchmark for different years. The dashed line indicates the prior acceptance probability per year, considered as the baseline performance.

3.4.2 On the importance of frequent model updating

Fig. 3.4 shows the year-wise comparison between the AUC-PR of the dynamic model and the static models fitted on specific years.

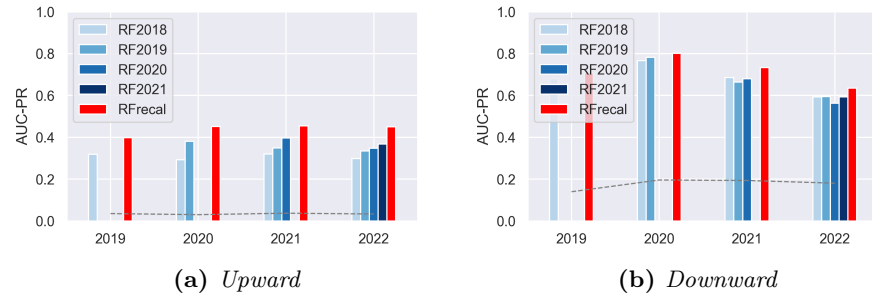


Figure 3.4: Year-wise comparison of the performance of the dynamic model with the static models. Static models trained on specific years are tested on subsequent years only (e.g. the model fitted on 2020 data is only tested on 2021 and 2022 data). The dashed line indicates the prior acceptance probability per year.

Beyond the difference in performance magnitude and trend between UR and DR, which was already commented previously, the necessity of frequent model updates turns out to be clearer for UR, despite the performance stability. Indeed,

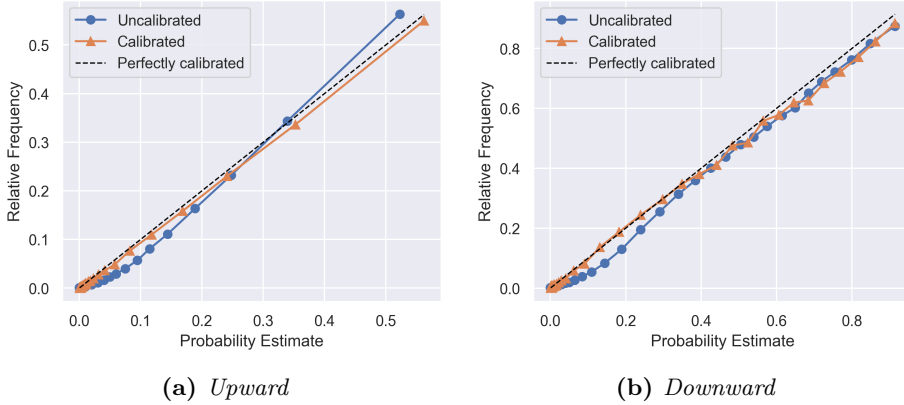


Figure 3.5: Reliability diagrams for the estimated probabilities. The black dashed line corresponds to an ideal perfectly calibrated classifier.

we observe that the static models keep up performances slightly lower but close to the dynamic model in the case of DR, but are clearly overtaken by the dynamic model right from the year following the fitted year, in the case of UR. For UR, furthermore, newer models are systematically outperforming older ones, while for DR the most recent static model does not necessarily perform better than an older one.

3.4.3 Generalisation error assessment

The reliability diagrams, pictured on Fig. 3.5, show that the model produces overall reliable probability estimates. It tends however to be slightly overconfident (predicted probability is higher than the observed relative frequency), except for the highest probabilities of UR where it shows an underconfidence. We can see that the probability calibration procedure is able to correct these distortions when predicting low acceptance probabilities (below 0.3) both for DR and for UR, but does not succeed in correcting the mild overconfidence observed in the case of DR for probabilities above 0.5. The knowledge of model's overconfidence at specific probability ranges can be valuable for market participants aiming to calibrate their strategies according to their risk tolerance. For instance, a risk-averse agent could account for this overconfidence and make its bid slightly more competitive – increasing bid price in the case of DR – whenever its target acceptance probability is in the overconfidence range.

Reliability diagrams provide a complete picture on the probabilistic prediction performance but do not allow to summarize this performance in a meaningful metric. Since our corrected probability predictions are well-calibrated, we know that the optimal binary prediction is obtained by applying a threshold

of 0.5 to the probability prediction. Therefore we can "pick" the corresponding point on the PR curve and consider the associated PR metric values for final model performance evaluation. Since we have about 50,000 samples per month in the test period, we can have a precise PR point estimate for each of the 36 test months on which the dynamic model is evaluated. Fig. 3.6 shows that the

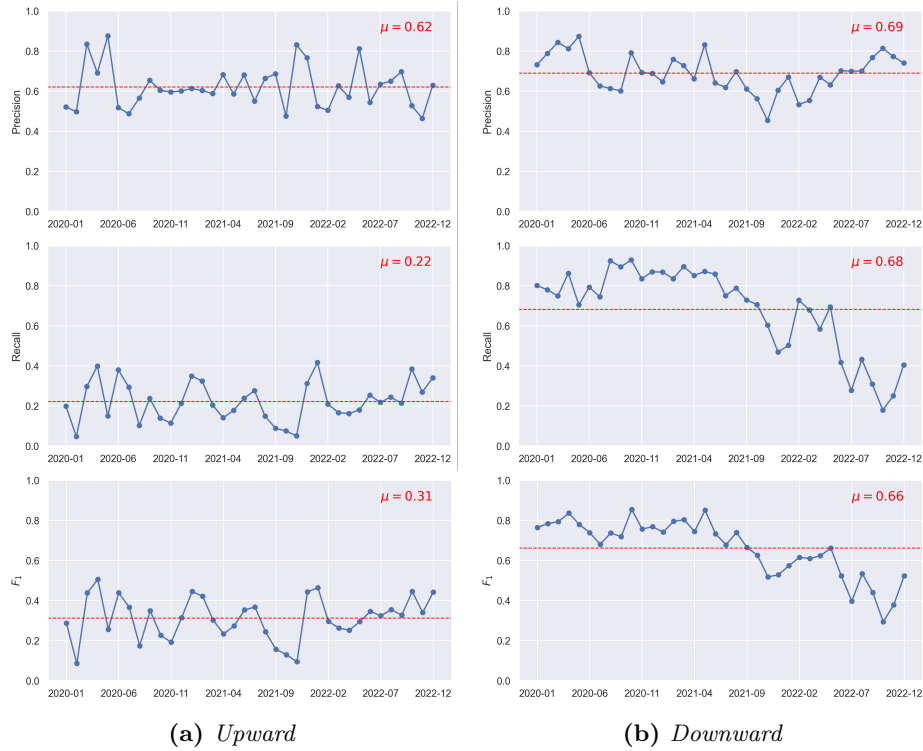


Figure 3.6: *PR metrics (precision, recall and F_1 -score) measured during each month of the test period. The red dashed line indicates the estimated mean in the hypothesis of stationary series, which doesn't seem to hold for downward bids.*

PR metrics vary consequently across months. These variations are however of different nature in the downward or upward case. An augmented Dickey-Fuller test (Dickey and Fuller, 1979) performed on each series of Fig. 3.6 systematically indicates a non-stationarity at the 5% significance level in the case of DR and a stationarity at any level of confidence for UR. This means that we get a good estimate of the short-term future performance of the model from the observed mean performance for UR (0.62 for the precision and 0.22 for the recall) but not for DR. Additionally, we find out that the AUC-PR performance drop observed earlier for DR is very likely to be due to the recall's behaviour that starts plummeting from the middle of 2021. This observation logically implies

that the model faced the emergence of accepted bids which acceptance cannot be explained with the covariates, denoting a clear market shift from that period.

3.4.4 Model explainability

Globally, both for DR and UR, the model essentially relies on the following variables (Fig. 3.7):

- **Tech** (unit technology)
- **Qty** (bid quantity)
- **BP** (measure of the binding program)
- **PriceDiff** (measure of the discrepancy between the bid price and the day-ahead market price)
- **Prov** (proxy for the location of the GU on the grid)

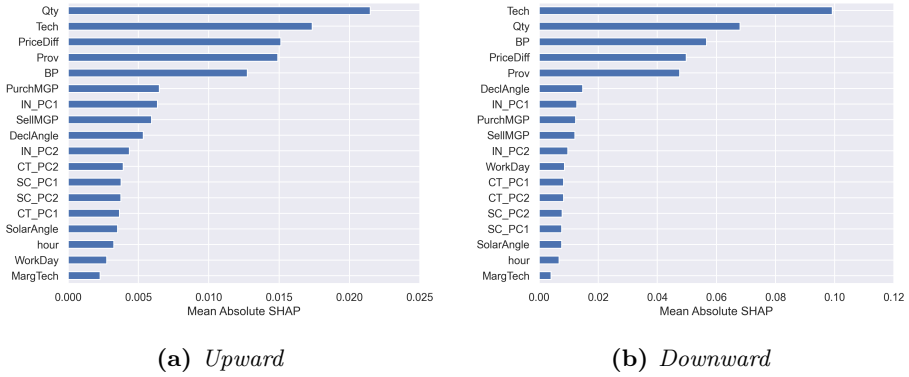


Figure 3.7: Feature importance over the entire test period

Given the model’s dynamic nature, we also examined how the importance of these four variables evolved over time (Fig.3.8). We can observe that their importances remain relatively stable over time, with however some notable variation concerning the unit technology (**Tech**) in the case of DR (Fig. 3.8b). Indeed **Tech**’s importance temporarily increased during the 2020-2021 period, potentially indicating the presence of technology-specific COVID-19 effects. Additionally, the rising importance of price (**PriceDiff**) over time may indicate increasing competition in the market. For UR, we can note instead a slight progressive increase of the unit technology importance (Fig. 3.8a).

The effect of the unit technology is shown on Fig. 3.9. Fossil gas units are (almost) systematically associated with higher acceptance probabilities, particularly in the case of DR. Contrarily, any other technology shows a negative

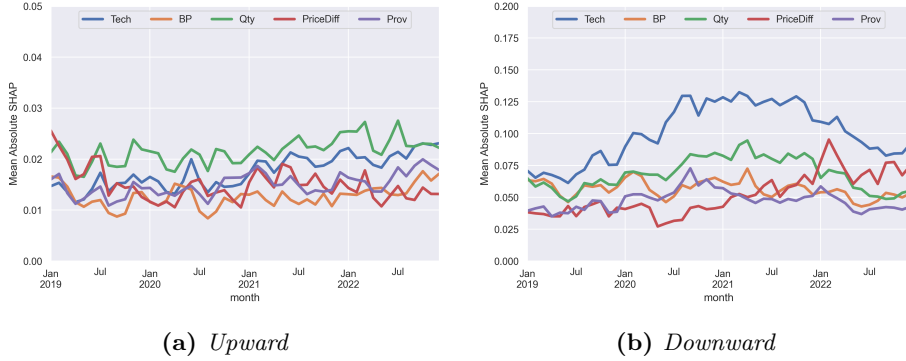


Figure 3.8: *Evolution of SHAP feature importance for the top four variables over the test period*

influence on bid acceptance, with a very distinct effect for hydropower technologies' downward bids. This effect is slightly more nuanced for UR since we observe numerous outlying SHAP values which are both positive and negative for all technologies. This aligns with earlier findings concerning the more complex nature of upward bids acceptance.

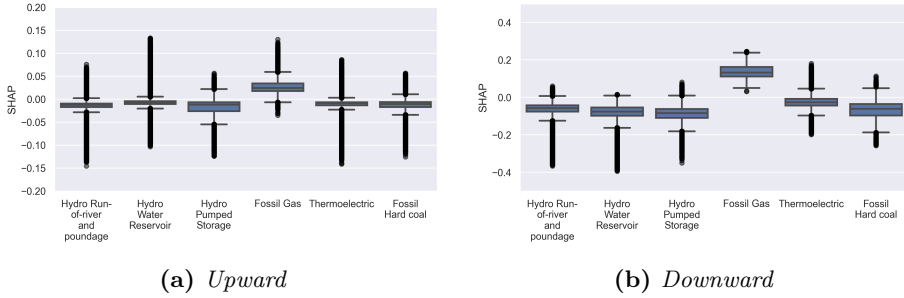


Figure 3.9: *Effect of the unit technology on the model's output*

While it is unlikely that the TSO applies explicit technology-based preferences, this result clearly shows that market dynamics are strongly influenced by the technical characteristics of different production technologies; more precisely, we can state that bids' acceptance is influenced by how GUs' characteristics are modelled by the TSO within the ASM. Although it is difficult to extract more information from the conducted analysis, a possible explanation of this phenomenon resides in some constraints affecting a lot of hydropower plants, which are often characterised by a limited energy capability linked to water availability (either stored or fluent), though this confounding effect should be controlled as the bid quantity is included in the covariates.

Turning to price, Fig. 3.10 reveals that its influence varies by technology.

For both UR and DR, the price effect is more pronounced for fossil gas units, especially in DR, where other technologies show a much weaker price sensitivity. As expected, a higher bid price reduces acceptance probability for UR and increases it for DR. The relationship is nearly linear for DR, again consistent with the good performance of the DR GLM model. In contrast, UR exhibits more that of a non-linear regime: beyond a threshold of $\text{PriceDiff} > \frac{1}{3}$ (i.e., a bid price more than twice the DA price), further increases do not significantly lower the acceptance probability. This pattern suggest that when bids are accepted while having a large price difference, thus being strongly inefficient for the TSO from the economic point of view, the reason behind their acceptance is mainly technical, overcoming possible financial optimization issues. This finding aligns with that of Spiller et al. (2022) which already highlighted the acceptance of a few bids at extreme prices.

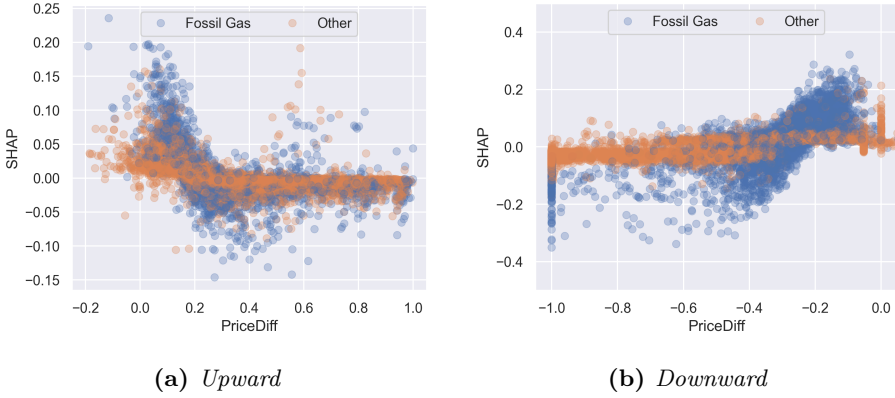


Figure 3.10: *Effect of the price difference (PriceDiff) on the model's output*

A similar pattern emerges with the binding program variable (BP), whose influence also depends on technology. For gas units, a BP in the 40-60% range maximizes the upward bid acceptance probability, while a BP between 70% and 100% range minimizes it. Interestingly, a null BP does not correlate with higher acceptance, possibly because switching such units on would also require compensating for startup costs, despite the additional downward flexibility they could offer. For DR, any bid placed by a GU which BP is lower than 80% is less likely to be accepted whatever the exact BP value, especially for the gas plants. For $\text{BP} \geq 80\%$, fossil gas units bids acceptance probability linearly increases with BP. This is coherent with the possible need, by the TSO, to create an upward reserve margin by decreasing the power setpoint of GUs that are producing at their maximum (nominal) capacity; indeed, a downward call during the ISP would reduce the GU scheduled power, creating therefore upward flexibility that the TSO can exploit for real-time balancing.

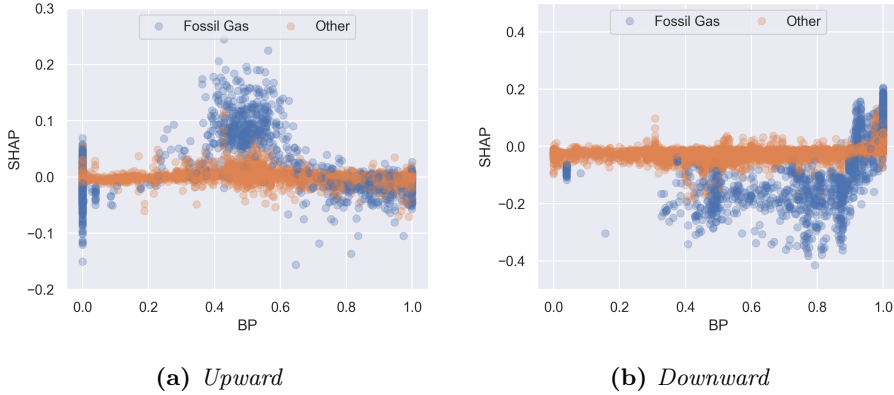


Figure 3.11: *Effect of the binding program (BP) on the model's output.*

Looking at the influence of the bid quantity (Fig. 3.12), we note that small offered volumes are less likely to be accepted than larger ones, with a rather clear threshold of 200 MWh. A possible explanation is that it is more convenient for the TSO to balance the network with fewer units and therefore larger per-unit volumes.

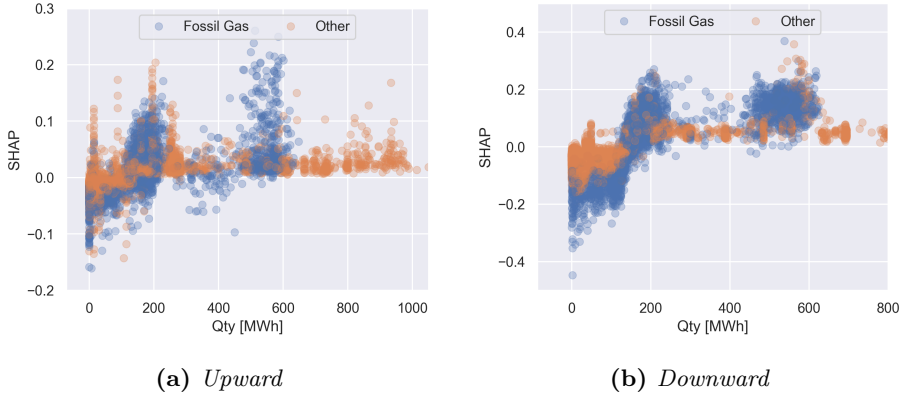


Figure 3.12: *Effect of the price quantity (Qty) on the model's output*

The influence of location is pictured on Fig. 3.13. For UR, we can see that a location in Pavia province (at the south of Milan) has a strong positive effect on bid acceptance. A similar, though weaker, positive influence is observed for eastern provinces near Bologna and Venice. In the case of DR, the provinces of Savona (west of Genoa) and Vercelli (northeast of Turin) show a clear positive effect, whereas the Ferrara province (northeast of Bologna) displays a negative

effect. These spatial effects are more likely to be related to the topology of the network rather than recurring generation patterns as the latter are already captured through variables such as the binding program (BP) and structural features (IN_PC1, IN_PC2, CT_PC1, CT_PC2, SC_PC1, SC_PC2). This spatial information could be particularly valuable for new market entrants, such as small-scale DERs or storage systems, when selecting optimal grid connection locations.

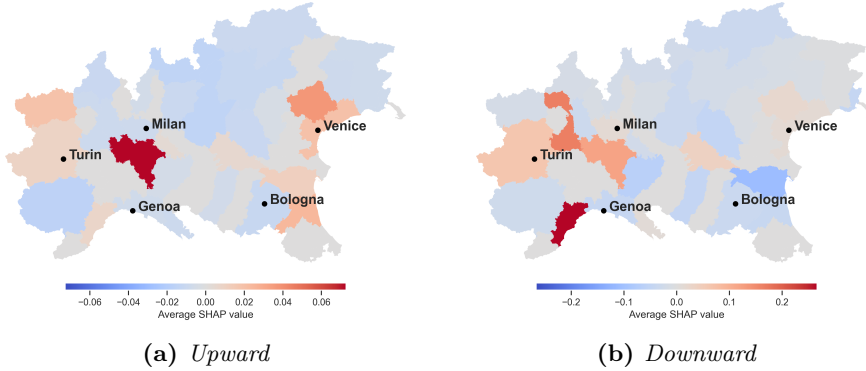


Figure 3.13: *Effect of the unit location (Prov) on the model's output. To enhance visualisation, provinces absent of the dataset but in the North region are represented with a null SHAP value*

Overall, the relevance of covariates other than the price confirms the initial assumption about the prevalence of technical constraints and grid-specific considerations in CDMs, at the detriment of economic optimality.

3.4.5 Optimal bidding strategy

Fig. 3.14 shows the estimated conditional acceptance probability and marginal profit as a function of the bid price p for an upward and a downward bid respectively submitted by the two selected GUs in January 2022. A marginal cost of 200€/MWh¹⁰ was assumed. For the downward bid, the GU bade at a price around 180€/MWh corresponding to an estimated expected profit of 15€/MWh while the estimated optimal bidding price is around 120€/MWh and leads to an estimated expected profit of 50€/MWh. It means that the GU has been "too competitive" and shrunk its expected profit with a bid price too close to its marginal cost, though guaranteeing a high acceptance probability. For the upward bid, the GU was not competitive enough: the actual bid price was around 285€/MWh, yielding an expected profit of 15€/MWh, whereas the optimal bid price – at around 260€/MWh – would have nearly doubled the expected profit.

¹⁰Which is roughly equal to the average day-ahead market price during that period.

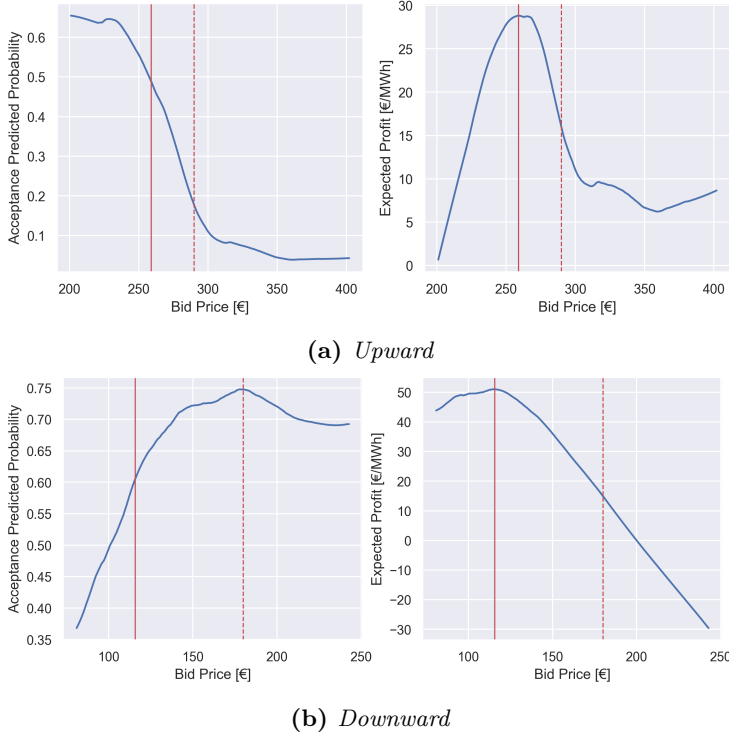


Figure 3.14: *Estimated acceptance probability and expected profit depending on the bid price for UR (top) and a DR (bottom). The red vertical solid line indicates the estimated optimal bidding price while the red vertical dashed line indicates the actual price that was submitted for that bid.*

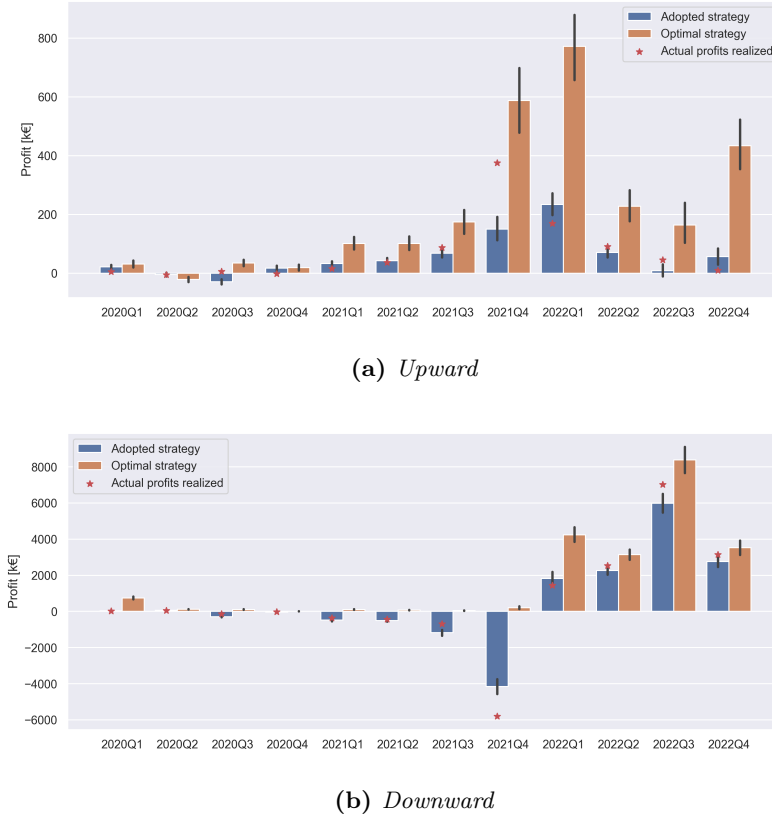


Figure 3.15: *Estimated quarterly profits cleared by UP_LA_CASELL_2 from UR (top) and UP_VADOTERM_5 from DR (bottom), resulting from the market simulations. Bars height represent the average value and vertical black segments indicate the 95% prediction intervals. The profits resulting from the true market results are indicated with a red star.*

The results of the market simulation are presented on Fig. 3.15. The first observation concerns the comparison between actual and simulated profits under the adopted strategy. While it is clear that the prediction intervals do not contain the true value at the nominal 95% coverage – which is expected given that the classifier is not perfect – we still observe that the model provides reasonably accurate estimates of quarterly profits. Furthermore, the underlying profit estimator appears to be unbiased, as it does not show a consistent tendency to overestimate or underestimate "true" profits. However, we note a large error in the fourth quarter of 2021 for both UR and DR, aligning with the previously identified market shift period and confirming a change in market dynamics

mainly induced by high spot prices beginning in the end of 2021 due to the energy crisis in Europe. A comment is needed regarding the negative profits for UR in 2021: though the GU appears to be "losing" in its ASM activity, it may still benefit from its overall spot trading activity, including day-ahead and intraday markets trading plus other monetary token received from the balancing activity. Specifically, volumes repurchased on the ASM may have been initially sold at a higher price on the day-ahead market, making this loss an opportunity cost. In this period, the optimal strategy only helps avoid losses, likely because the optimal bid price is very close to the marginal cost and/or any price below the marginal cost is not competitive enough. Finally, we see that the optimal strategy consistently results in significantly higher profits. The total profits cleared over the three-years period are reported in Table 3.9.

	Expected	Realized	PI 95%
Adopted strategy [M€]	0.67	0.83	[0.60, 0.74]
Optimal strategy [M€]	2.63	-	[2.42, 2.83]
% change	+292%	-	-

(a) Upward (*UP_LA_CASELL_2*)

	expected	realized	PI 95%
Adopted strategy [M€]	6.3	6.8	[5.5, 7.2]
Optimal strategy [M€]	20.7	-	[19.8, 21.7]
% change	+228%	-	-

(b) Downward (*UP_VADOTERM_5*)

Table 3.9: *Estimated profits cleared from ASM-ISP activity during the 2020-2022 period under the pricing strategy adopted by the unit and the optimal strategy proposed.*

3.5 Discussion

Along with the analysis of its performance over time, the comparison of the model to its benchmarks and static counterparts gave us valuable information on market dynamics. First, the consideration of unit, market and network information – beyond just bid price – is necessary to satisfyingly forecast market results, supporting evidence that they deviate from the economic optimum. Second the upward bids acceptance mechanisms are more complex than those of the downward bids, and are not well modeled through linear effects in the covariates, contrarily to DR. Moreover, the results suggest that the acceptance mechanisms caught by the model remain relatively stable over time for DR, whereas they

appear to undergo frequent changes for UR. Finally, we observed that DR was subject to a perturbation from the middle of 2021, as the model failed to identify a large part of the new accepted bids. Future work may investigate this aspect further.

The explanation techniques applied to the random forest model reveal clear effects. First, the unit's generation technology is a highly predictive factor of its bids acceptance but also conditions the effect of the other variables on the acceptance probability. Specifically, a fossil-gas technology appears to significantly raise the likelihood of a bid to be accepted, and increases the sensitivity of bid acceptance to the other relevant variables. Second, there appears to be a threshold in the offered volume – around 200 MWh – below which a bid is less likely to be accepted. Note that this insight is of interest for DERs aggregators who may see an indication of the critical size to reach in order to be competitive. We also find out that the binding program is a key determinant: the TSO logically prefers to call GUs near their minimum capacity for UR and near their maximum capacity for DR to maximize flexibility in both directions. However, in the case of UR, the estimated effect appears more nuanced. Rather than selecting plants operating very close to their minimum capacity – possibly to avoid the additional cost of the start-up service – the TSO tends to call plants operating around half of their maximum capacity. As expected, bid competitiveness is also crucial. There is a low competitiveness range where no significant effect is observed, followed by a range where the bid becomes competitive, exhibiting an almost linear effect. We also highlighted a clear spatial effect that may be related to the network topology and which could help new entrants in the choice of their geographic location.

Finally, while cross-validation effectively addresses the uncertainty in PR estimates, quantifying the uncertainty in individual acceptance probability estimates is more challenging. Although the reliability diagrams indicate low bias by showing that average probability estimates align with observed relative frequencies in the out-of-sample set, they do not provide guarantees regarding the estimator's variance at the individual observation level. Bootstrapping could be a way to estimate this variance, but it is definitely too costly in our case, given the sample size. Regarding the bidding strategy, quantifying this uncertainty could provide confidence intervals for either a bid's optimal price or its associated expected profit at a fixed price, both of which would be valuable for the operator. Furthermore, this would produce more reliable prediction intervals for the estimated profit in the market simulation as we would take into account the additional variability due to the estimation error of the Bernoulli parameter.

3.6 Conclusion

Power reserve markets are gaining increasing importance as power systems face growing balancing needs driven by the integration of variable renewable energy

sources and the rise of flexible loads. In this context, we proposed a machine learning approach to model pay-as-bid power reserve markets and derive an optimal bidding strategy for market participants. Specifically, we developed a probabilistic classification pipeline to estimate the acceptance probability of individual bids based on price and other explanatory variables including bid characteristics and indicators of network conditions. We then showed that this bid acceptance probability estimator can be used by a participant to optimally price its bids, maximizing the expected marginal profit.

The proposed methodology was validated on historical data from the Italian ASM, where it accurately estimates bid acceptance probabilities and allows market players to significantly increase their profits – by up to a factor of three – making the conclusions of this work highly relevant for them. Beyond statistical and economic performance, emphasis was placed on the interpretability of the underlying machine learning system, enabling some level of understanding of actual market clearing mechanisms. In particular we identified a systematic preference for bids from gas-fired power plants – controlling for other influencing factors – compared to those from other generation technologies such as hydropower. This asymmetry raises concerns about fairness and may hinder the future integration of alternative balancing service providers such as DERs. This finding is relevant for regulatory authorities, who may consider policy interventions to ensure more equitable treatment across technologies. Overall, we believe that the proposed decision-making tool and the findings of this study will foster competition in power reserve markets, reduce market concentration, and ultimately enhance overall welfare – benefiting end electricity consumers, who continue to face rising bills because of increasing balancing costs.

Conclusions

IN this thesis, we develop data-driven models for the analysis of electricity spot markets in decarbonizing power systems. We address three key steps of the data-driven decision-making pipeline of a market actor: *monitoring* market fundamentals, predicting participants' behaviors to *forecast* market results, and finally defining *bidding strategies* to optimize revenues. The proposed approaches rely on statistical models for binary, time series, and functional data, adopting a “statistician’s” perspective with a particular focus on uncertainty quantification and probabilistic prediction. While the primary contribution consists of proposing models intended to be deployable as operational tools, we also believe that the conclusions drawn from the proposed application – the Italian electricity market – provide valuable insights for power industry stakeholders, ranging from market participants to regulatory authorities.

Summary of contributions

In Chapter 1, we propose a blind source separation framework based on a constrained variant of non-negative matrix factorization to monitor the consumption of residential, services and industrial sectors at high frequency from aggregate high-voltage grid load measurements. Applying the method to Italy’s national load curve over the 2021-2023 period, we reconstruct accurate hourly consumption profiles for each sector. The results reveal that both household and services consumption exhibit two distinct daily regimes related to seasonality and day type, whereas industrial demand follows a single, stable daily profile. Furthermore, the monthly sectoral consumption estimates derived from the disaggregated load closely align with official sample-based statistics and prove therefore competitive with current approaches for real-time monitoring. A central takeaway of this work is the potential of statistical decomposition methods to address

electrical load profiling or disaggregation problems, even in the absence of high-frequency disaggregated consumption measurements.

Chapter 2 introduces a general framework for predicting day-ahead supply and demand merit-order curves in auction-based markets, from which both point and probabilistic electricity price forecasts can be derived. We leverage functional data analysis techniques to represent pairs of supply and demand curves in a low-dimensional vector space and employ regularized vector autoregressive models for their prediction. We conduct a rigorous empirical comparison of price forecasting performance between the proposed *curve-based* approach, derived from predicted merit-order curves, and state-of-the-art *price-based* models that directly forecast the clearing price, using data from the Italian day-ahead market over the 2023-2024 period. The results show that the curve-based approach significantly improves both point and probabilistic price forecasting accuracy relative to price-based methods, with average gains of approximately 5%, and improvements of up to 10% during mid-day hours, when prices occasionally drop due to high renewable generation and low demand.

Finally, in Chapter 3, we propose a machine learning approach to model pay-as-bid ancillary services markets (ASMs) and derive an optimal bidding strategy for market participants. Specifically, we develop a probabilistic classification pipeline to estimate the acceptance probability of individual bids based on price and additional explanatory variables including bid characteristics and indicators of network conditions. We further employ XAI techniques to interpret the underlying machine learning system and provide insights into actual market clearing mechanisms. We then show how this bid acceptance probability estimator can be used by market participants to optimally price their bids and maximize expected marginal profit. Applied to the Italian ASM over the 2018-2022 period, the results show that bid acceptance probabilities are accurately estimated and that the proposed bidding strategy enables market players to significantly increase their profits – by up to a factor of three. Furthermore, the analysis partially reveals the opaque clearing apparatus of the ASM, identifying a systematic bias in favor of gas-fired generation technologies. These findings are relevant not only for market participants but also for regulatory authorities.

Limitations and future research

We first discuss the limitations and directions for future research specific to each of the three contributions, before considering broader perspectives related to the PhD work as a whole.

A primary limitation of the contribution presented in Chapter 1 is the lack of ground truth data at fine temporal resolutions, such as hourly or daily, which would enable (i) improved supervision of the source separation process and (ii) a direct assessment of disaggregation accuracy, thereby complementing the

literature-based validation. Besides, the model assumes a hierarchical relationship between the identified sources and the sectors, an assumption that may not hold in practice. Indeed, a given source may be associated with multiple sectors – for example, air conditioning usage in both residential and commercial consumption. Future research could extend the empirical validation to other countries or smaller power zones where high-frequency ground truth consumption data are available, potentially considering a more granular consumer segmentation.

Regarding Chapter 2, further analyses could examine the effect of (i) alternative curves representations¹¹, (ii) the choice of vector space dimension, (iii) the use of a single versus hour-specific representation, and (iv) the calibration window size, on both curves and clearing price forecasting performance. In addition, more complex non-linear models such as feedforward neural networks could be considered to jointly predict the 24-hour score matrix. Finally, the recent transition to 15-minute intervals in European day-ahead markets – resulting in 96 rather than 24 curve pairs to predict – may motivate a functional rather than multivariate modeling perspective to capture intraday dependence. These considerations constitute promising directions for future research.

In Chapter 3, bid outcomes are modeled independently, thereby overlooking potential interdependencies among concurrent bids submitted by different units for the same delivery period. This is a clear shortcoming for market participants managing multiple units and system-level market simulations. Furthermore, the uncertainty associated with the acceptance probability estimates – and, consequently, with the optimal bid prices – is not explicitly quantified. The temporal and geographical scope of the study could also be expanded. Future work could address these limitations by jointly modeling bid outcomes, improving uncertainty quantification, and applying the proposed methodology to other markets, particularly those in which distributed energy resources (DERs) are already integrated.

From a broader perspective, several key aspects of electricity market modeling are not addressed in this PhD thesis.

First, we do not consider the prediction of two fundamental market drivers: demand and VRE generation. The literature on these topics is already extensive (Hong et al., 2020) and at the zonal or country level such forecasts are often readily available, typically provided by TSOs or private financial and weather data providers. From an academic and practical standpoint, these aspects therefore represent a lower priority for the present work.

Second, the analysis focuses exclusively on short-term forecasting horizons. While medium-term (weeks- or months-ahead) and long-term (years-ahead) horizons are less relevant for spot trading, they play a crucial role in the pricing of forward market products and in investment planning. As highlighted by Weron (2014) in the context of EPF, short-term forecasting models may be satisfactorily

¹¹Such as the original form $P(q)$ and/or unconstrained forms (c.f. section 2.2.1).

employed for medium-term horizons, whereas long-term forecasting generally requires fundamentally different modeling approaches. Nevertheless, the results of Ziel and Steinert (2018) suggest that the EPF approach adopted in Chapter 2 can also perform well over longer forecasting horizons.

Third, although the framework proposed in Chapter 2 is applicable to auction-based intraday (ID) markets, this thesis does not address continuous-trading ID markets. These markets differ substantially in structure and share several similarities with financial markets, as supply and demand offers are matched sequentially as they are submitted. As a result, each transaction is associated with its own clearing price, and a single delivery period is characterized by a continuous price trajectory from gate opening to gate closure. In this setting, the problem shifts from forecasting a time series of prices to forecasting a time series of price *trajectories*. A functional time series framework, such as that presented in Chapter 2, may therefore provide an effective modeling approach.

Finally, although this thesis addresses the optimal bidding problem in the context of technical markets, it does not explicitly consider how EPF models can inform bidding strategies in commercial markets. This issue is particularly important for power storage systems, whose profitability – unlike that of generation-only units – depends directly on the timing of electricity purchases (charging) and sales (discharging), and therefore on accurate price forecasts. As emphasized by Yardley and Petropoulos (2021), forecasts in a business context should ideally be evaluated according to their *economic* value rather than solely through generic error metrics. In particular, Serafin and Weron (2025) note that price forecasts that are optimal with respect to classical error measures (e.g., MSE, MAE, MAPE) are not necessarily optimal for trading purposes. Consequently, model selection based on economic validation – such as trading simulations or decision-focused error metrics – may be more appropriate. For example, systematically biased price forecasts may perform poorly according to standard error metrics while still accurately capturing the *shape* of the daily price curve, including peak and valley locations. Such forecasts may nonetheless be valuable for storage systems seeking to optimize daily charge-discharge cycles and exploit intraday price variations. In these cases, shape-matching accuracy measures, such as the Pearson correlation coefficient between actual and predicted price vectors, may be preferred to standard pointwise error metrics.

To conclude, the contributions presented in this PhD thesis demonstrate the value of advanced statistical learning techniques for addressing practical challenges faced by participants, operators, and regulators in electricity wholesale markets. While *accuracy* and *interpretability* are often regarded as opposing characteristics of data-driven models, the present work shows that improvements in one need not come at the expense of the other. By leveraging XAI techniques to explain model predictions and, when possible, by explicitly modeling the mechanisms underlying the formation of the target variable, this thesis

proposes models that are both highly accurate and interpretable.

Moreover, when the phenomenon under study is inherently stochastic – as is the case for electricity markets – point predictions become less informative for decision-making. In this context, a third concept complements the accuracy–interpretability paradigm: *reliability*. Reliable models not only provide accurate forecasts but also appropriately quantify the uncertainty associated with them. Although the methods developed in this thesis successfully address this reliability dimension, a central challenge remains: how to effectively incorporate uncertainty measures into real-world decision-making processes. Addressing this challenge will be essential for power industry stakeholders, who will be called to operate in increasingly uncertain market environments.

Bibliography

- Andersen, F. M., Gunkel, P. A., Jacobsen, H. K., and Kitzing, L. (2021). Residential electricity consumption and household characteristics: An econometric analysis of Danish smart-meter data. *Energy Economics*, 100:105341.
- Andersen, F. M., Larsen, H. V., and Boomsma, T. K. (2013a). Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers. *Energy Conversion and Management*, 68:244–252.
- Andersen, F. M., Larsen, H. V., and Gaardestrup, R. B. (2013b). Long term forecasting of hourly electricity consumption in local areas in Denmark. *Applied Energy*, 110:147–162.
- Andersen, F. M., Larsen, H. V., Juul, N., and Gaardestrup, R. B. (2014). Differentiated long term projections of the hourly electricity consumption in local areas. The case of Denmark West. *Applied Energy*, 135:523–538.
- Andersen, N. N. and Anandha, J. (2025). Transparency in European Electricity Markets and ENTSO-E Transparency Platform. In Estermann, A., Schrade, M., and Anderson, L., editors, *European Electricity Market Coupling: A Practitioner’s Guide*, pages 339–351. Springer Nature Switzerland, Cham.
- Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the Prediction of Stationary Functional Time Series. *Journal of the American Statistical Association*, 110(509):378–392. _eprint: <https://doi.org/10.1080/01621459.2014.909317>.
- Balputnis, K., Broka, Z., Cingels, G., Silis, A., and Junghans, G. (2024). Robust market-based battery energy storage management strategy for operation in European balancing markets. *Journal of Energy Storage*, 102:114082.
- Behm, C., Nolting, L., and Praktijnjo, A. (2020). How to model European electricity load profiles using artificial neural networks. *Applied Energy*, 277:115564.
- Bellinguer, K., Girard, R., Bocquet, A., and Chevalier, A. (2023). ELMAS: a one-year dataset of hourly electrical load profiles from 424 French industrial and tertiary sectors. *Scientific Data*, 10(1):686.

Bibliography

- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Springer Science & Business Media. Google-Books-ID: lqWEH23nytQC.
- Bovera, F., Blaco, A., Rancilio, G., and Delfanti, M. (2019). Assessing the Accuracy of Different Machine Learning Classification Algorithms in Forecasting Results of Italian Ancillary Services Market. In *2019 16th International Conference on the European Energy Market (EEM)*, pages 1–5.
- Bovera, F., Rancilio, G., Falabretti, D., and Merlo, M. (2021). Data-Driven Evaluation of Secondary- and Tertiary-Reserve Needs with High Renewables Penetration: The Italian Case. *Energies*, 14(8).
- Bovera, F., Spiller, M., Zatti, M., Rancilio, G., and Merlo, M. (2023). Development, validation, and testing of advanced mathematical models for the optimization of BESS operation. *Sustainable Energy, Grids and Networks*, 36:101152.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time series analysis: forecasting and control*. Prentice Hall, Englewood Cliffs (N.J.), 3rd ed edition.
- Boyd, K., Eng, K. H., and Page, C. D. (2013). Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals. In Blockeel, H., Kersting, K., Nijssen, S., and Železný, F., editors, *Machine Learning and Knowledge Discovery in Databases*, pages 451–466, Berlin, Heidelberg. Springer.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.
- Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78(1):1–3.
- Băra, A. and Oprea, S.-V. (2024). Machine Learning Algorithms for Power System Sign Classification and a Multivariate Stacked LSTM Model for Predicting the Electricity Imbalance Volume. *International Journal of Computational Intelligence Systems*, 17(1):80.
- Canale, A. and Vantini, S. (2016). Constrained functional time series: Applications to the Italian gas market. *International Journal of Forecasting*, 32(4):1340–1351.
- Cardo-Miota, J., Pérez, E., and Beltran, H. (2023). Deep learning-based forecasting of the automatic Frequency Reserve Restoration band price in the Iberian electricity market. *Sustainable Energy, Grids and Networks*, 35:101110.
- Cerasa, A. and Zani, A. (2025). Enhancing electricity price forecasting accuracy: A novel filtering strategy for improved out-of-sample predictions. *Applied Energy*, 383:125357.
- Chen, Y., Niu, Y., Qu, C., Du, M., and Liu, P. (2024). A pricing strategy based on bi-level stochastic optimization for virtual power plant trading in multi-market: Energy, ancillary services and carbon trading market. *Electric Power Systems Research*, 231:110371.
- Chen, Z. and Cichocki, A. (2004). Nonnegative Matrix Factorization with Temporal Smoothness and/or Spatial Decorrelation Constraints.
- Ciarreta, A., Martinez, B., and Nasirov, S. (2023). Forecasting electricity prices using bid data. *International Journal of Forecasting*, 39(3):1253–1271.

- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.
- Cox, D. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232.
- Dampeyrou, C., Goichon, A., Ghienné, M., Tschannen, V., and Schaack, S. (2024). Un-supervised separation of the thermosensitive contribution in the power consumption at a country scale. *Applied Energy*, 363:123097.
- DeGroot, M. H. and Fienberg, S. E. (1983). The Comparison and Evaluation of Forecasts. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 32(1/2):12–22.
- Department for Energy Security and Net Zero (2024). Review of Electricity Market Arrangements: Options assessment. Technical report, Government of the United Kingdom.
- Diaz-Iglesias, A., Belaunzaran, X., and Florez-Tapia, A. M. (2025). Short-Term Power Demand Forecasting for Diverse Consumer Types to Enhance Grid Planning and Synchronisation. arXiv:2506.04294 [cs].
- Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366):427–431.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263. [_eprint: https://www.tandfonline.com/doi/pdf/10.1080/07350015.1995.10524599](https://www.tandfonline.com/doi/pdf/10.1080/07350015.1995.10524599).
- Diquigiovanni, J., Fontana, M., and Vantini, S. (2024). Distribution-Free Prediction Bands for Multivariate Functional Time Series: an Application to the Italian Gas Market. arXiv:2107.00527 [stat].
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2):407–499.
- Ember (2025). Yearly Electricity Data.
- ENTSO-E (2025). ENTSO-E Transparency Platform.
- European Commission (2019). The European Green Deal. Technical Report COM/2019/640 final, European Commission, Brussels.
- European Union (2017). Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing, Article 2. Legislative Body: COM, ENER.
- Falvo, M., Panella, S., Caprabianca, M., and Quaglia, F. (2021). A Review on Unit Commitment Algorithms for the Italian Electricity Market. *Energies*, 15(1):18.
- Federico, G. and Rahman, D. (2003). Bidding in an Electricity Pay-as-Bid Auction. *Journal of Regulatory Economics*, 24(2):175–211.
- Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, 1(2).

Bibliography

- Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Journal of Nonparametric Statistics*, 16(1-2):111–125. [_eprint: https://doi.org/10.1080/10485250310001622686](https://doi.org/10.1080/10485250310001622686).
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer Series in Statistics. Springer New York.
- Figueiredo, M., Ribeiro, B., and de Almeida, A. (2014). Electrical Signal Source Separation Via Nonnegative Tensor Factorization Using On Site Measurements in a Smart Home. *IEEE Transactions on Instrumentation and Measurement*, 63(2):364–373. Conference Name: IEEE Transactions on Instrumentation and Measurement.
- Flach, P. and Kull, M. (2015). Precision-Recall-Gain Curves: PR Analysis Done Right. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.
- Gerossier, A., Barbier, T., and Girard, R. (2017). A novel method for decomposing electricity feeder load into elementary profiles from customer information. *Applied Energy*, 203:752–760.
- Ghelasi, P. and Ziel, F. (2024). Hierarchical forecasting for aggregated curves with an application to day-ahead electricity price auctions. *International Journal of Forecasting*, 40(2):581–596.
- Ghelasi, P. and Ziel, F. (2025). A data-driven merit order: Learning a fundamental electricity price model. [arXiv:2501.02963 \[stat\]](https://arxiv.org/abs/2501.02963).
- Gillis, N. (2020). *Nonnegative Matrix Factorization*. Data Science. Society for Industrial and Applied Mathematics.
- Giovanelli, C., Sierla, S., Ichise, R., and Vyatkin, V. (2018). Exploiting Artificial Neural Networks for the Prediction of Ancillary Energy Market Prices. *Energies*, 11(7):1906.
- GME (2026). Electricity Market.
- Gneiting, T. and Katzfuss, M. (2014). Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151.
- Gulotta, F., Rossi, A., Bovera, F., Falabretti, D., Galliani, A., Merlo, M., and Rancilio, G. (2020). Opening of the Italian Ancillary Service Market to Distributed Energy Resources: Preliminary Results of UVAM project. In *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*, pages 199–203, Charlotte, NC, USA. IEEE.
- Guo, H., Chen, Q., Zheng, K., Xia, Q., and Kang, C. (2021). Forecast Aggregated Supply Curves in Power Markets Based On LSTM Model. *IEEE Transactions on Power Systems*, 36(6):5767–5779.
- Guo, Y.-T., Li, Q.-Q., and Liang, C.-S. (2024). The rise of nonnegative matrix factorization: Algorithms and applications. *Information Systems*, 123:102379.

- Hart, G. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891.
- Henzi, A., Ziegel, J. F., and Gneiting, T. (2021). Isotonic Distributional Regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(5):963–993.
- Herath, G., Liyanage, M., and Angammana, C. (2023). Assessment of Non-Intrusive Load Monitoring as a Blind Source Separation Problem. In *2023 IEEE PES GTD International Conference and Exposition (GTD)*, pages 156–160.
- Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., and Zareipour, H. (2020). Energy Forecasting: A Review and Outlook. *IEEE Open Access Journal of Power and Energy*, 7:376–388.
- Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, volume 200 of *Springer Series in Statistics*. Springer, New York, NY.
- Hosseinzadeh Aghdam, M. (2022). A novel constrained non-negative matrix factorization method based on users and items pairwise relationship for recommender systems. *Expert Systems with Applications*, 195:116593.
- Hubicka, K., Marcjasz, G., and Weron, R. (2019). A Note on Averaging Day-Ahead Electricity Price Forecasts Across Calibration Windows. *IEEE Transactions on Sustainable Energy*, 10(1):321–323.
- Hyndman, R. J. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd ed)*. OTexts, Melbourne, Australia.
- Hyndman, R. J. and Shahid Ullah, M. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.
- Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, 38(3):199–211.
- Hörmann, S. and Kokoszka, P. (2012). 7 - Functional Time Series. In Subba Rao, T., Subba Rao, S., and Rao, C. R., editors, *Handbook of Statistics*, volume 30 of *Time Series Analysis: Methods and Applications*, pages 157–186. Elsevier.
- IEA (2005). Lessons from Liberalised Electricity Markets. Technical report, International Energy Agency, Paris.
- IEA (2021). Key World Energy Statistics 2021. Technical report, IEA, Paris.
- Jardini, J., Tahan, C., Gouvea, M., Ahn, S., and Figueiredo, F. (2000). Daily load profiles for residential, commercial and industrial low voltage consumers. *IEEE Transactions on Power Delivery*, 15(1):375–380.
- Jia, S. and Qian, Y. (2009). Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):161–173.
- Jin, A. S. and Sanders, K. T. (2024). Characterizing residential sector load curves from smart meter datasets. *Applied Energy*, 366:123316.

Bibliography

- Johnson, R. A. and Wichern, D. W. (2007). *Applied multivariate statistical analysis*. Pearson/Prentice Hall, Upper Saddle River, NJ, 6. ed edition.
- Joskow, P. L. (2008). Lessons Learned From Electricity Market Liberalization. *The Energy Journal*, 29(2_suppl):9–42.
- Kath, C. and Ziel, F. (2021). Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *International Journal of Forecasting*, 37(2):777–799.
- Koechlin, G., Bovera, F., Degli Innocenti, E., Santini, B., Venturi, A., Vazio, S., and Secchi, P. (2026a). A blind source separation framework to monitor sectoral power demand from grid-scale load measurements. *Sustainable Energy, Grids and Networks*, page 102259.
- Koechlin, G., Bovera, F., and Secchi, P. (2024). Longitudinal Study of Dispatchable Power Units Trading Strategies on Electricity Spot Markets in Italy. In *2024 20th International Conference on the European Energy Market (EEM)*, pages 1–6.
- Koechlin, G., Bovera, F., and Secchi, P. (2025a). Deconvolution of the Electrical Load Curve Using Blind Source Separation. In Pollice, A. and Mariani, P., editors, *Methodological and Applied Statistics and Demography IV*, pages 194–199, Cham. Springer Nature Switzerland.
- Koechlin, G., Bovera, F., and Secchi, P. (2025b). Forecasting Electricity Spot Market Merit-Order Curves with Functional Time Series Modeling. In *2025 21st International Conference on the European Energy Market (EEM)*, pages 1–9.
- Koechlin, G., Bovera, F., and Secchi, P. (2025c). Strategic bidding in pay-as-bid power reserve markets: A machine learning approach. *Energy Economics*, 150:108780.
- Koechlin, G., Bovera, F., and Secchi, P. (2026b). Day-Ahead Electricity Price Forecasting Using Merit-Order Curves Time Series. arXiv:2512.17758 [stat].
- Kolter, J., Batra, S., and Ng, A. (2010). Energy Disaggregation via Discriminative Sparse Coding. In *Proceedings of the 24th International Conference on Neural Information Processing Systems - Volume 1*, volume 23. Curran Associates, Inc.
- Kraft, E., Keles, D., and Fichtner, W. (2020). Modeling of frequency containment reserve prices with econometrics and artificial intelligence. *Journal of Forecasting*, 39(8):1179–1197.
- Kraft, J. and Kraft, A. (1978). On the Relationship Between Energy and GNP. *The Journal of Energy and Development*, 3(2):401–403.
- Krarti, M. and Aldubyan, M. (2021). Review analysis of COVID-19 impact on electricity demand for residential buildings. *Renewable and Sustainable Energy Reviews*, 143:110888.
- Kulakov, S. (2020). X-Model: Further Development and Possible Modifications. *Forecasting*, 2(1):20–35. Number: 1.
- Lago, J., Marcjasz, G., De Schutter, B., and Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983.

- Langevin, J., Reyna, J. L., Ebrahimigharehbaghi, S., Sandberg, N., Fennell, P., Nägeli, C., Laverge, J., Delghust, M., Mata, E., Van Hove, M., Webster, J., Federico, F., Jakob, M., and Camarasa, C. (2020). Developing a common approach for classifying building stock energy models. *Renewable and Sustainable Energy Reviews*, 133:110276.
- Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791. Number: 6755.
- Lee, D. D. and Seung, H. S. (2000). Algorithms for non-negative matrix factorization. In *Proceedings of the 13th International Conference on Neural Information Processing Systems*, NIPS’00, pages 535–541, Cambridge, MA, USA. MIT Press.
- Leeuw, J. D., Hornik, K., and Mair, P. (2009). Isotone Optimization in *R* : Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *Journal of Statistical Software*, 32(5).
- Li, J. and Ho, M. S. (2022). Indirect cost of renewable energy: Insights from dispatching. *Energy Economics*, 105:105778.
- Li, Y., Deng, Y., Wang, Y., Jiang, L., and Shahidehpour, M. (2023). Robust bidding strategy for multi-energy virtual power plant in peak-regulation ancillary service market considering uncertainties. *International Journal of Electrical Power & Energy Systems*, 151:109101.
- Li, Z., Alonso, A. M., Elías, A., and Morales, J. M. (2024). Clustering and forecasting of day-ahead electricity supply curves using a market-based distance. *International Journal of Electrical Power & Energy Systems*, 158:109977.
- Li, Z., Alonso, A. M., and Pascual, L. (2025). Predicting electricity supply and demand curves with functional data techniques. *International Journal of Electrical Power & Energy Systems*, 166:110561.
- Lichtendahl, K. C., Grushka-Cockayne, Y., and Winkler, R. L. (2013). Is It Better to Average Probabilities or Quantiles? *Management Science*, 59(7):1594–1611.
- Lin, J. and Magnago, F. H. (2017). *Electricity Markets: Theories and Applications: Theories and Applications*. Wiley, 1 edition.
- Lipiecki, A., Uniejewski, B., and Weron, R. (2024). Postprocessing of point predictions for probabilistic forecasting of day-ahead electricity prices: The benefits of using isotonic distributional regression. *Energy Economics*, 139:107934.
- Lipiecki, A. and Weron, R. (2025). PostForecasts.jl: A Julia package for probabilistic forecasting by postprocessing point predictions. *SoftwareX*, 31:102200.
- Lisi, F., Grossi, L., and Quaglia, F. (2023). Evaluation of Cost-at-Risk related to the procurement of resources in the ancillary services market. The case of the Italian electricity market. *Energy Economics*, 121:106625.
- Liu, H., Wu, Z., Li, X., Cai, D., and Huang, T. S. (2012). Constrained Nonnegative Matrix Factorization for Image Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1299–1311. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

- Liu, X., Xia, W., Wang, B., and Zhang, L. (2011). An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2):757–772.
- Liu, Y., Jin, R., and Yang, L. (2006). Semi-supervised multi-label learning by constrained non-negative matrix factorization. In *Proceedings of the 21st national conference on Artificial intelligence - Volume 1*, AAAI’06, pages 421–426, Boston, Massachusetts. AAAI Press.
- Lokhande, S., Bichpuriya, Y., Kulkarni, A. A., and Sarangan, V. (2023). An optimized trading strategy for an energy storage systems aggregator in an ancillary service market. *Journal of Energy Storage*, 72:108588.
- LSEG Data & Analytics (2025). ECMWF/GFS wind and solar generation forecasts for Italy.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67.
- Lundberg, S. M. and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Maneesha, A. and Swarup, K. S. (2021). Stochastic Optimal Bidding Strategy for Energy and Ancillary Services in Microgrid. *IEEE Transactions on Industry Applications*, 57(6):5698–5705.
- Marcjasz, G., Serafin, T., and Weron, R. (2018). Selection of Calibration Windows for Day-Ahead Electricity Price Forecasting. *Energies*, 11(9):2364. Number: 9.
- Matsumoto, M., Fujimoto, Y., and Hayashi, Y. (2016). Energy Disaggregation Based on Semi-Binary NMF. In Perner, P., editor, *Machine Learning and Data Mining in Pattern Recognition*, pages 401–414, Cham. Springer International Publishing.
- Mazzi, N., Kazempour, J., and Pinson, P. (2018). Price-Taker Offering Strategy in Electricity Pay-as-Bid Markets. *IEEE Transactions on Power Systems*, 33(2):2175–2183.
- McKinsey & Company (2022). Enabling Renewable Energy with Battery Energy Storage Systems.
- Merten, M., Olk, C., Schoeneberger, I., and Sauer, D. U. (2020). Bidding strategy for battery storage systems in the secondary control reserve market. *Applied Energy*, 268:114951.
- Mestre, G., Portela, J., Muñoz San Roque, A., and Alonso, E. (2020). Forecasting hourly supply curves in the Italian Day-Ahead electricity market with a double-seasonal SARMAHX model. *International Journal of Electrical Power & Energy Systems*, 121:106083.
- Michalakopoulos, V., Sarmas, E., Papias, I., Skaloumpakas, P., Marinakis, V., and Doukas, H. (2024). A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs. *Applied Energy*, 361:122943.

- Miyasawa, A., Fujimoto, Y., and Hayashi, Y. (2019). Energy disaggregation based on smart metering data via semi-binary nonnegative matrix factorization. *Energy and Buildings*, 183:547–558.
- Molnar, C. (2020). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Leanpub.
- Mor, G., Cipriano, J., Martirano, G., Pignatelli, F., Lodi, C., Lazzari, F., Grillone, B., and Chemisana, D. (2021). A data-driven method for unsupervised electricity consumption characterisation at the district level and beyond. *Energy Reports*, 7:5667–5684.
- Moscetti, F., Paoletti, S., and Vicino, A. (2014). Analysis and models of electricity prices in the Italian ancillary services market. In *IEEE PES Innovative Smart Grid Technologies, Europe*, pages 1–6, Istanbul, Turkey. IEEE.
- Murphy, A. H. and Winkler, R. L. (1977). Reliability of Subjective Probability Forecasts of Precipitation and Temperature. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26(1):41–47.
- Møller Andersen, F., Gunkel, P. A., Bjerregaard, C., and Jacobsen, H. K. (2024). Changes in hourly electricity consumption profiles and price elasticities in Denmark 2019–2022. *Energy*, 308:132798.
- Niculescu-Mizil, A. and Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning*, ICML '05, pages 625–632, New York, NY, USA. Association for Computing Machinery.
- Nowotarski, J. and Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568.
- Olivares, K. G., Challu, C., Marcjasz, G., Weron, R., and Dubrawski, A. (2023). Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx. *International Journal of Forecasting*, 39(2):884–900.
- O’Connor, C., Bahloul, M., Prestwich, S., and Visentin, A. (2025). A Review of Electricity Price Forecasting Models in the Day-Ahead, Intra-Day, and Balancing Markets. *Energies*, 18(12):3097.
- Paatero, J. V. and Lund, P. D. (2006). A model for generating household electricity load profiles. *International Journal of Energy Research*, 30(5):273–290. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.1136>.
- Pagliarini, G., Bonfiglio, C., and Vocale, P. (2019). Outdoor temperature sensitivity of electricity consumption for space heating and cooling: An application to the city of Milan, North of Italy. *Energy and Buildings*, 204:109512.
- Pauca, V. P., Piper, J., and Plemmons, R. J. (2006). Nonnegative matrix factorization for spectral data analysis. *Linear Algebra and its Applications*, 416(1):29–47.
- Pelagatti, M. (2013). Supply Function Prediction in Electricity Auctions. In Grigoletto, M., Lisi, F., and Petrone, S., editors, *Complex Models and Computational Methods in Statistics*, pages 203–213, Milano. Springer Milan. Series Title: Contributions to Statistics.

- Peng, C., Wong, K.-C., Rockwood, A., Zhang, X., Jiang, J., and Keyes, D. (2012). Multiplicative Algorithms for Constrained Non-negative Matrix Factorization. In *2012 IEEE 12th International Conference on Data Mining*, pages 1068–1073, Brussels, Belgium. IEEE.
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., Ellison, J., Fiszeder, P., Franses, P. H., Frazier, D. T., Gilliland, M., Gönül, M. S., Goodwin, P., Grossi, L., Grushka-Cockayne, Y., Guidolin, M., Guidolin, M., Gunter, U., Guo, X., Guseo, R., Harvey, N., Hendry, D. F., Hollyman, R., Januschowski, T., Jeon, J., Jose, V. R. R., Kang, Y., Koehler, A. B., Kolassa, S., Kourentzes, N., Leva, S., Li, F., Litsiou, K., Makridakis, S., Martin, G. M., Martinez, A. B., Meeran, S., Modis, T., Nikolopoulos, K., Önköl, D., Paccagnini, A., Panagiotelis, A., Panapakidis, I., Pavia, J. M., Pedio, M., Pedregal, D. J., Pinson, P., Ramos, P., Rapach, D. E., Reade, J. J., Rostami-Tabar, B., Rubaszek, M., Sermpinis, G., Shang, H. L., Spiliotis, E., Syntetos, A. A., Talagala, P. D., Talagala, T. S., Tashman, L., Thomakos, D., Thorarinsdottir, T., Todini, E., Trapero Arenas, J. R., Wang, X., Winkler, R. L., Yusupova, A., and Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3):705–871.
- Petrovitch, B. (2026). European Electricity Review 2026. Technical report, Ember.
- Piga, D., Cominola, A., Giuliani, M., Castelletti, A., and Rizzoli, A. E. (2016). Sparse Optimization for Automated Energy End Use Disaggregation. *IEEE Transactions on Control Systems Technology*, 24(3):1044–1051.
- Platt, J. (2000). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in large margin classifiers*, 10.
- Rahimpour, A., Qi, H., Fugate, D., and Kuruganti, T. (2017). Non-Intrusive Energy Disaggregation Using Non-Negative Matrix Factorization With Sum-to-k Constraint. *IEEE Transactions on Power Systems*, 32(6):4430–4441. Conference Name: IEEE Transactions on Power Systems.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, New York, NY.
- Rancilio, G., Bovera, F., and Merlo, M. (2022). Revenue Stacking for BESS: Fast Frequency Regulation and Balancing Market Participation in Italy. *International Transactions on Electrical Energy Systems*, 2022(1):1894003. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2022/1894003>.
- Räsänen, T., Voukantsis, D., Niska, H., Karatzas, K., and Kolehmainen, M. (2010). Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, 87(11):3538–3545.
- Saito, T. and Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3):e0118432.

- Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, pages 166–171, Minneapolis, MN, USA. IEEE.
- Savitzky, A. and Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639. __eprint: <https://doi.org/10.1021/ac60214a047>.
- Schirmer, P. A. and Mporas, I. (2023). Non-Intrusive Load Monitoring: A Review. *IEEE Transactions on Smart Grid*, 14(1):769–784. Conference Name: IEEE Transactions on Smart Grid.
- Semwal, S., Joshi, D., Prasad, R. S., and Raveendhra, D. (2013). The practicability of ICA in home appliances load profile separation using current signature: A preliminary study. In *2013 International Conference on Power, Energy and Control (ICPEC)*, pages 756–759.
- Serafin, T. and Weron, R. (2025). Loss functions in regression models: Impact on profits and risk in day-ahead electricity trading. *Energy Economics*, 148:108596.
- Shah, I. and Lisi, F. (2020). Forecasting of electricity price through a functional prediction of sale and purchase curves. *Journal of Forecasting*, 39(2):242–259. __eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.2624>.
- Shang, H. L. (2013). Functional time series approach for forecasting very short-term electricity demand. *Journal of Applied Statistics*, 40(1):152–168. __eprint: <https://doi.org/10.1080/02664763.2012.740619>.
- Shapley, L. S. (1951). Notes on the N-Person Game — II: The Value of an N-Person Game. Technical report, RAND Corporation.
- Shlens, J. (2014). A Tutorial on Independent Component Analysis. arXiv:1404.2986 [cs].
- Siddiqui, A., Marnay, C., and Khavkin, M. (2001). Spot pricing of electricity and ancillary services in a competitive California market. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pages 9 pp.–.
- Singh, S. and Majumdar, A. (2019). Analysis Co-Sparse Coding for Energy Disaggregation. *IEEE Transactions on Smart Grid*, 10(1):462–470.
- Sinha, N. and Lucheroni, C. (2025). Demand and supply curve forecasting using a monotonic autoencoder for short-term day-ahead electricity market bid curves. *Applied Energy*, 397:126262.
- Solovieva, M. and Vargiolu, T. (2021). Efficient representation of supply and demand curves on day-ahead electricity markets. *The Journal of Energy Markets*.
- S&P Global Ratings (2021). Going With The Flow: The Competitiveness Of Battery Storage Economics In the Power Sector.
- Spiller, M., Rancilio, G., Bovera, F., and Merlo, M. (2022). Opportunities for DERs in Ancillary Services Markets: analysis and deployment strategies in a national framework. In *2022 18th International Conference on the European Energy Market (EEM)*, pages 1–6.

Bibliography

- Swider, D. J. and Weber, C. (2007). Bidding under price uncertainty in multi-unit pay-as-bid procurement auctions for power systems reserve. *European Journal of Operational Research*, 181(3):1297–1308.
- Tang, Q., Guo, H., Zheng, K., and Chen, Q. (2024). Forecasting individual bids in real electricity markets through machine learning framework. *Applied Energy*, 363:123053.
- Teeraratkul, T., O'Neill, D., and Lall, S. (2018). Shape-Based Approach to Household Electric Load Curve Clustering and Prediction. *IEEE Transactions on Smart Grid*, 9(5):5196–5206. Conference Name: IEEE Transactions on Smart Grid.
- Teh, Y. W. (2011). Dirichlet Process. In Sammut, C. and Webb, G. I., editors, *Encyclopedia of Machine Learning*, pages 280–287. Springer US, Boston, MA.
- Terna (2024a). Consumi 20XX. Technical report, Terna S.p.A.
- Terna (2024b). Enel Household consumption Index.
- Terna (2024c). IMCEI Industrial consumption.
- Terna (2024d). IMSER Services consumption.
- Terna (2024e). Total Load.
- Thoma, M. (2004). Electrical energy usage over the business cycle. *Energy Economics*, 26(3):463–485.
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Transparency Market Research (2024). Ancillary Services for Power Market Size, Share, Growth, Trends, and Forecast 2024.
- Uniejewski, B. (2024). Regularization for electricity price forecasting. *Operations Research and Decisions*, 34(3).
- Uniejewski, B., Nowotarski, J., and Weron, R. (2016). Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting. *Energies*, 9(8):621. Number: 8.
- United Nations (2015). Paris Agreement to the United Nations Framework Convention on Climate Change. Technical Report 3156, United Nations.
- van Zoest, V., Lindberg, K., El Gohary, F., and Bartusch, C. (2023). Evaluating the effects of the COVID-19 pandemic on electricity consumption patterns in the residential, public, commercial and industrial sectors in Sweden. *Energy and AI*, 14:100298.
- Viola, L., Mohammadi, S., Dotta, D., Hesamzadeh, M. R., Baldick, R., and Flynn, D. (2024). Ancillary services in power system transition toward a 100% non-fossil future: Market design challenges in the United States and Europe. *Electric Power Systems Research*, 236:110885.
- Vivó, G. and Alonso, A. M. (2024). Prediction of Intraday Electricity Supply Curves. *Applied Sciences*, 14(22):10663.

- Voulis, N., Warnier, M., and Brazier, F. M. T. (2018). Understanding spatio-temporal electricity demand at different urban scales: A data-driven approach. *Applied Energy*, 230:1157–1171.
- Wang, J., Zhong, H., Tang, W., Rajagopal, R., Xia, Q., Kang, C., and Wang, Y. (2017). Optimal bidding strategy for microgrids in joint energy and ancillary service markets considering flexible ramping products. *Applied Energy*, 205:294–303.
- Wang, P., Zareipour, H., and Rosehart, W. (2014). Descriptive models for reserve and regulation prices in competitive electricity markets. *IEEE Transactions on Smart Grid*, 5(1):471–479.
- Wang, P., Zareipour, H., and Rosehart, W. (2018). Modeling Hourly Original Operating Reserve Prices in Electricity Market. In *2018 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5, Portland, OR. IEEE.
- Wang, X., Si, C., Gu, J., Liu, G., Liu, W., Qiu, J., and Zhao, J. (2021). Electricity-consumption data reveals the economic impact and industry recovery during the pandemic. *Scientific Reports*, 11(1):19960.
- Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., and Zhao, Y. (2015). Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, 20(2):117–129.
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081.
- Wytock, M. and Kolter, J. (2014). Contextually Supervised Source Separation with Application to Energy Disaggregation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1). Number: 1.
- Yardley, L. and Petropoulos, F. (2021). Beyond Error Measures to the Utility And Cost of the Forecasts. *Foresight: the International Journal of Applied Forecasting*, 63:36–45.
- Yıldırım, S., Khalafi, M., Güzel, T., Satık, H., and Yılmaz, M. (2023). Supply Curves in Electricity Markets: A Framework for Dynamic Modeling and Monte Carlo Forecasting. *IEEE Transactions on Power Systems*, 38(4):3056–3069. Conference Name: IEEE Transactions on Power Systems.
- Zadrozny, B. and Elkan, C. (2002). Transforming classifier scores into accurate multi-class probability estimates. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '02, pages 694–699, New York, NY, USA. Association for Computing Machinery.
- Zarabie, A. K. and Das, S. (2019). An L0-Norm Constrained Non-Negative Matrix Factorization Algorithm for the Simultaneous Disaggregation of Fixed and Shiftable Loads. arXiv:1908.00142 [cs, eess].
- Zhang, S., Guo, Q., Smyth, R., and Yao, Y. (2022). Extreme temperatures and residential electricity consumption: Evidence from Chinese households. *Energy Economics*, 107:105890.
- Zhu, Y. and Lu, S. (2014). Load profile disaggregation by Blind source separation: A wavelets-assisted independent component analysis approach. In *2014 IEEE PES General Meeting / Conference & Exposition*, pages 1–5.

Bibliography

- Ziel, F. (2016). Forecasting Electricity Spot Prices Using Lasso: On Capturing the Autoregressive Intraday Structure. *IEEE Transactions on Power Systems*, 31(6):4977–4987. Conference Name: IEEE Transactions on Power Systems.
- Ziel, F. and Steinert, R. (2016). Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Economics*, 59:435–454.
- Ziel, F. and Steinert, R. (2018). Probabilistic mid- and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews*, 94:251–266.
- Ziel, F. and Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70:396–420.

APPENDIX $\mathcal{A}$

Supplementary material to Chapter 1

A.1 Modified multiplicative updates for the resolution of LC-NMF

Given $B \in \mathbb{R}_+^{m \times n}$, $A \in \mathbb{R}_+^{K \times g}$, $Y \in \mathbb{R}_+^{m \times g}$, $E \in \mathbb{R}_+^{l \times K}$, $D \in \mathbb{R}_+^{p \times q}$, $Z \in \mathbb{R}_+^{l \times q}$, we propose to solve:

$$(\hat{C}, \hat{S}) = \arg \min_{C \in \mathbb{R}_+^{n \times K}, S \in \mathbb{R}_+^{K \times p}} \mathcal{L}(C, S, \alpha, \beta) \quad (\text{A.1})$$

where

$$\mathcal{L}(C, S, \alpha, \beta) = \|X - CS\|_F^2 + \alpha \|BCA - Y\|_F^2 + \beta \|ESD - Z\|_F^2$$

and α, β are two non-negative penalization parameters.

The case $\alpha = \beta = 0$ corresponds to the ordinary NMF problem:

$$\mathcal{L}(C, S) = \mathcal{L}(C, S, 0, 0) = \|X - CS\|_F^2$$

Using matrix differential calculus, with $'$ denoting the transpose:

$$\nabla_S \mathcal{L}(C, S) = 2C'CS - 2C'X \quad \text{and} \quad \nabla_C \mathcal{L}(C, S) = 2CSS' - 2XS'$$

Separating positive and negative contributions, the gradient $\nabla_S \mathcal{L}$ or $\nabla_C \mathcal{L}$ can be written:

$$\nabla_S \mathcal{L}(C, S) = \nabla_S^+ \mathcal{L}(C, S) - \nabla_S^- \mathcal{L}(C, S)$$

The popular multiplicative updates (MU) algorithm of Lee and Seung Lee and Seung (2000), in case of Frobenius loss, proposes to alternatively update C and S , as follows:

$$S \leftarrow S \cdot \frac{C'X}{C'CS}, \quad C \leftarrow C \cdot \frac{XS'}{CSS'}$$

Where $\cdot$ and $—$ operations are coefficient-wise.

There are several ways to derive these updates, mentioned in the original paper of Lee and Seung and detailed in section 8.2 of Gillis (2020), among which the “gradient ratio heuristic” that shows that these updates can be written:

$$C \leftarrow S \cdot \frac{\nabla_S^- \mathcal{L}(C, S)}{\nabla_S^+ \mathcal{L}(C, S)}, \quad C \leftarrow C \cdot \frac{\nabla_C^- \mathcal{L}(C, S)}{\nabla_C^+ \mathcal{L}(C, S)}$$

We can apply the same idea in the case $\alpha \neq 0$ or $\beta \neq 0$. In this case we have:

$$\nabla_C \mathcal{L}(C, S, \alpha, \beta) = 2CSS' + 2\alpha B'BCAA' - (2XS' + 2\alpha B'YA')$$

$$\nabla_S \mathcal{L}(C, S, \alpha, \beta) = 2C'CS + 2\beta E'ESDD' - (2C'X + 2\beta E'ZD')$$

and we can therefore consider the following updates:

$$S \leftarrow S \cdot \frac{C'X + \beta E'ZD'}{C'CS + \beta E'ESDD'}, \quad C \leftarrow C \cdot \frac{XS' + \alpha B'YA'}{CSS' + \alpha B'BCAA'}$$

With a well-chosen auxiliary function, the proofs in Lee and Seung (2000) can be directly adapted to prove that the previous updates lead to a monotonically decreasing loss $\mathcal{L}$ and a converging update sequence.

APPENDIX *B*

Supplementary material to Chapter 2

B.1 Example of curves and clearing price prediction

B.2 Reliability and resolution of probabilistic forecasts

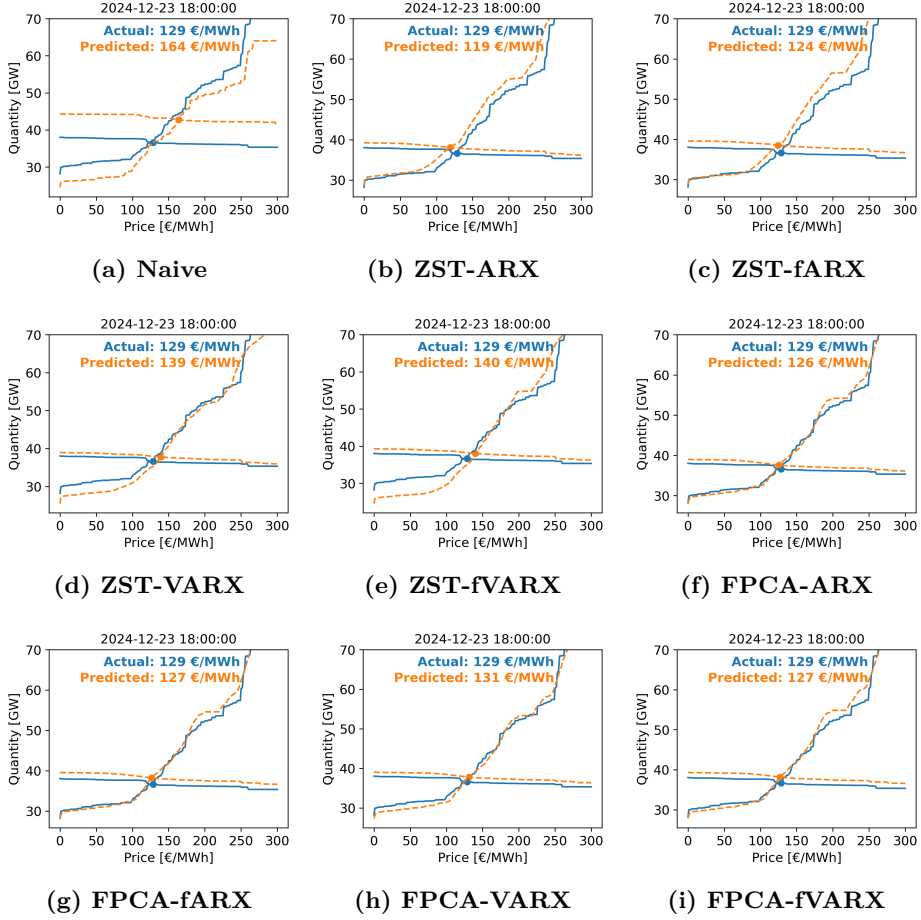


Figure B.1: *Curves and clearing price predictions for the 18:00-19:00 interval of December 23rd, 2024.*

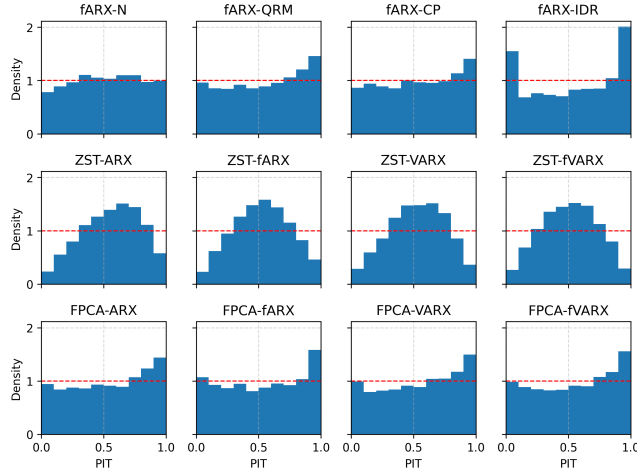


Figure B.2: *Probability Integral Transform (PIT) histograms. The horizontal dashed red line corresponds to a perfect distribution fit.*

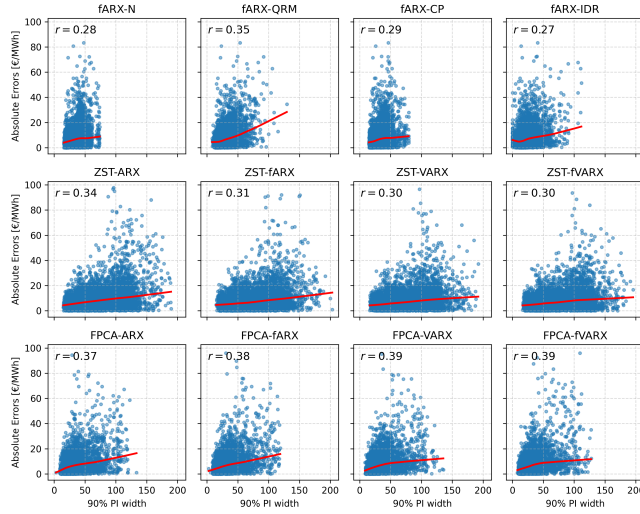


Figure B.3: *Scatter plots of 90% prediction interval widths versus absolute clearing price forecasting errors for all evaluated models. The red trend line is computed using LOWESS (Cleveland, 1979) and the Spearman's correlation coefficient r is reported.*