IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

IEEE POWER & ENERGY SOCIETY SECTION

Received 15 February 2024, accepted 13 May 2024, date of publication 20 May 2024, date of current version 29 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3403422

RESEARCH ARTICLE

Day-Ahead Electricity Price Forecasting in the Contemporary Italian Market

FRANCESCO MORAGLIO AND CARLO S. RAGUSA, (Senior Member, IEEE)
Dipartimento di Energia "Galileo Ferraris," Politecnico di Torino, 10129 Turin, Italy
Corresponding author: Francesco Moraglio (francesco.moraglio@polito.it)

ABSTRACT In competitive electricity markets, prices are determined by the collective behaviour of suppliers and consumers. Hence, these systems rely on the balance between supply and demand, and sudden changes in the underlying conditions can lead to significant price fluctuations. In the face of the recent transformations in Italian electricity markets, which are driven by an increasing share of non-programmable renewables, energy crises, and geopolitical tensions, our study focuses on effective forecasting methodologies. We compare kernel and linear regression for predicting market equilibrium prices, both in point and probabilistic sense. We showcase the potential of both linear and non-linear models when carefully engineering the problem of interest, which involves properly selecting data and variables. The noteworthy outcome is that, while linear and non-linear models may differ in nature, their performance converges closely, attesting to the robustness of our approach in achieving reliable forecasts. We describe data sources and assumptions in exploratory univariate analyses, and the performance of the final multivariate model is evaluated over a test period on September 2023.

INDEX TERMS Electricity supply industry, forecasting, economic forecasting, nonparametric statistics, linear approximation, probability, probability density function.

NOMENCLATURE

AIC	Akaike-Hurvich Information Criterion.
AR	AutoRegression.
CDF	Cumulative Distribution Function.
CI	Computational Intelligence.
EPF	Electricity Price Forecasting.
GME	Gestore del Mercato Elettrico.
IQR	Interquartile Range.
KDE	Kernel Density Estimation.
KR	Kernel Regression.
LR	Linear Regression.
MAPE	Mean Absolute Percentage Error.
MGP	Mercato del Giorno Prima.
ML	Machine Learning.
MPL	Mean Pinball Loss.
MSE	Mean Squared Error.
NECP	National Energy and Climate Plan.
NN	Neural Network.
NPRES	Non-Programmable Renewable Energy Sources.
NWE	Nadaraya-Watson Estimator.
PI	Prediction Interval.
PSE	Parzen-Stone Estimator.
PUN	Prezzo Unico Nazionale.
QR	Quantile Regression.

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Carmen Falvo.

I. INTRODUCTION

The transformation of traditionally centralized and government-regulated energy industries began in the 1990s, with the introduction of deregulation and competitive markets. The electricity sector, in particular, has undergone substantial changes. If in the early stages of an economy, a monopolistic system is necessary to simplify and ensure the development of the grid, once such infrastructure is developed, total or partial privatization can be financially more rewarding. In most Western countries, electricity is now traded under market rules and through various types of contracts, both in wholesale and retail markets. We focus

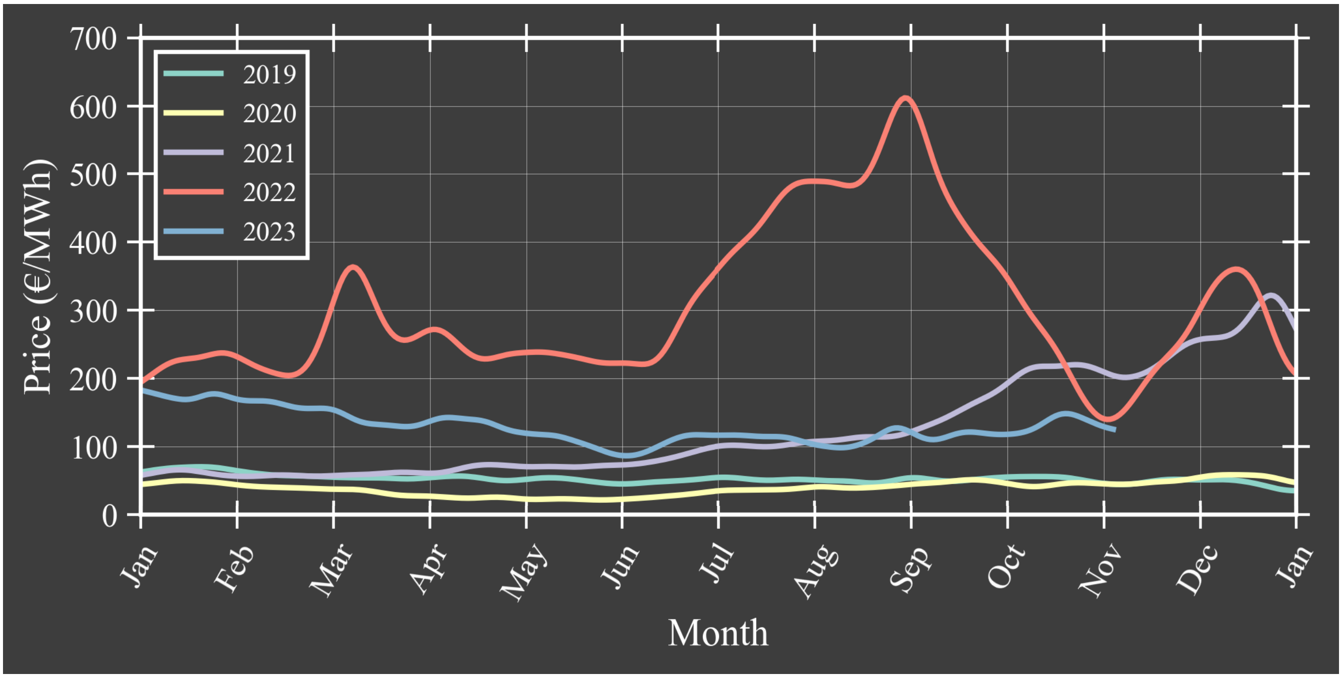

FIGURE 1. **Electricity price evolution in italy (2019-2023). The chart illustrates anomalous trends of the Italian day-ahead electricity price in 2021 (in violet) and, notably, 2022 (in red). The trend of 2023 (in blue) appears to fall halfway between the anomalous trends and the prices observed before 2021. In 2022, the Russian invasion of Ukraine led to a significant conflict and a severe energy crisis, marked by peaks aligning with periods of heightened geopolitical tension. These facts strengthened an uprising trend already present in Italy during the latter months of 2021, attributed to various factors, e.g., water scarcity and the concurrent maintenance of several power plants. The presented trends are extracted using the Nadaraya-Watson estimator, one of the simplest methods belonging to the class of kernel machine learning methods. The technique is also known as kernel smoothing.**

on the Italian context, where the liberalization process began in 1999 with the introduction of the so-called Bersani Law. This law established the unbundling of the electric supply chain, involving the separation of various stages such as electricity production, transmission, and distribution. The aims were to prevent the creation of vertical monopolies, and to promote competitiveness. After a gradual development spanning 25 years, the conclusion of the Italian deregulation process is expected in 2024, with all customers (or nearly all) managed within the framework of free competition, as well as a normative context aligned in flexibility and adequacy with those of other European Union (EU) Member States [6].

In monopolistic systems, prices are set and updated by a regulatory entity, and there is limited interest in developing methods to forecast such prices. By contrast, competitive markets are based on discretising electric demand into distinct load periods. A market equilibrium price is determined for each period, based on supply and demand dynamics: future electricity prices are no longer known a priori. Being able to predict such prices with adequate anticipation and an acceptable margin of error, which varies depending on the context, can be very interesting from a business perspective. In fact, having accurate forecasts enables speculative operations and optimizations of financial planning. For these reasons, researchers and practitioners put their efforts into the development of the discipline known as Electricity Price Forecasting (EPF) [12], [26].

For instance, within the day-ahead electricity market, trading occurs for the 24 hourly intervals of the day immediately following the market closing day. Each participant in the market submits bids indicating the purchase and/or sale of variable energy quantities, specific to an hourly period. At the deadline of the bidding period, an algorithm calculates the market equilibrium price. As described in [21], the process of forecasting day-ahead prices thus amounts to predicting the point of intersection between the aggregated demand and supply curves, for each hourly period in the following day.

In this study, we concentrate on the Italian day-ahead electricity market, referred to as Mercato del Giorno Prima (MGP). This choice is motivated by two main factors. Firstly, day-ahead markets represent the most extensively researched aspect in Electricity Price Forecasting (EPF) literature, constituting 80 % of recent publications within this domain [21]. Additionally, the Italian market stands out as one of the most comprehensively studied electric markets, attributed to its transparency and data accessibility, as the authors of [10] explained.

Over the last three years, the Italian electricity market has experienced significant shifts influenced by the COVID-19 pandemic, the energy crisis, and geopolitical tensions. For example, starting in 2020, there has been a substantial collapse in the liquidity of forward electricity markets, which are now nearly inactive and require an urgent regulatory reform [6]. It is also important to highlight that, despite the principle that free competition should lead to decreased prices, it is not guaranteed to occur in real markets. There are several instances where deregulated markets have undergone rapid price increases, eventually leading to unexpected crises [15].

Specifically, the MGP equilibrium price is termed Prezzo Unico Nazionale (PUN), and visualizing its trend over the last 5 years is helpful in understanding the challenges that have arisen in the industrial sector. Fig. 1 depicts the evolution of PUN daily averages in the period ranging from January 2019 to November 2023. As evident, in 2022, the price reached levels ten times higher than those in 2019 and 2020. This has harmed companies that had to incur debt for energy purchases, aggravated by the simultaneous increase in interest rates by the European Central Bank. An initial motivation for our work lies in the scarcity of research containing recent data analyses. In fact, considering the substantial differences in data trends, it is crucial to pay special attention, regardless of the forecasting approach adopted. In Sec. II, we provide a brief review of recent literature in EPF, introducing the differences proposed in our study.

Letting aside shortages and international tensions, we can state that the recent integration of a substantial share of Non-Programmable Renewable Energy Sources (NPRES) into the Italian generation system has added complexity to EPF. At the end of 2021, Italy's gross final electricity consumption was covered by 36% from renewable sources, of which roughly one-third of NPRES, and two-thirds of programmable sources, mainly geothermal and hydroelectric [6]. The percentage is set to increase, and precisely to double in the next few years. Indeed, investments in NPRES projects, which involve European funding, are regulated in EU Member States by a government document known as the National Energy and Climate Plan (NECP). According to the 2023 version of the Italian NECP, approved by the EU Commission [3], the percentage of consumption to be covered by renewables by 2030 has been raised to 65%. The presence of NPRES like photovoltaic and wind, introduces significant fluctuations in electricity generation, resulting in increased price volatility. This can be understood with the help of Fig. 2, which illustrates the difference between daily price profiles in May for the year 2011, in yellow, and 2021, in violet. The substantial rise in photovoltaic generation capacity, from less than 2% to almost 11% in ten years [6], has led to a reduction in prices during the midday hours when sunlight is typically abundant, as indicated by the light grey area in the figure. Additionally, there is an increased level of volatility, evident in significantly larger InterQuartile Ranges (IQRs) for 2021. Conversely, hours with limited or no solar generation show higher average prices.

As explained in several studies concerning similar markets in Europe [12], [21], traditional point forecasting methods may not be reliable enough: such high levels of volatility render sometimes unfeasible to make precise estimates about future prices. This implies that point predictions lose, at least partially, their industrial applicability. For these reasons, it is more significant to forecast hypothetical intervals in which prices are believed to fluctuate, namely Prediction Intervals (PI). In other words, a probabilistic forecasting approach is needed to complement point forecasts. In this scenario, non-linear Machine Learning (ML) methods are gaining momentum, being presented as promising solutions in a large portion of existing literature [21], [27]. Specifically, ML methods are often reported as being better in EPF than more classical methods based on linear models, attributed to their capability to effectively reproduce the non-linear relationships existing between electricity prices and their predictors. As the aim of this study is to explore the possibilities of creating a reliable EPF system for the contemporary Italian scenario, we present below a comparison of Kernel Regression (KR), versus the most classical and widely used Linear Regression (LR). KR is one the simplest non-linear, non-parametric ML methods [4], and all of the details on the employed models are provided in Sec. III.

In Sec. IV, we describe the data used for prediction, their respective sources, and the assumptions made, along with the results of exploratory univariate analyses. Among the selected input variables, photovoltaic and thermal generation reflect the supply dynamics in the market price formation, while previous-day electricity and natural gas prices are accounted for because of their direct correlation to the variable of interest.

In Sec. V, we define the multivariate models used for the final forecasting and present the results related to predictions over a test period set in September 2023. The model predictions are evaluated both in point and interval senses, with different temporal aggregations, highlighting the advantages derived from the application of the probabilistic approach. In the last Sec. VI, we draw the conclusions of our study.

II. LITERATURE REVIEW

The discipline of EPF falls within the much broader and well-established field of energy forecasting, the area of study that encompasses all techniques and analyses aimed at predicting future values of variables related to the energy sector, ranging from commodity prices to the physical characteristics of distribution networks. The world of energy commodities, indeed, presents numerous challenging situations from both an industrial and academic perspective, and many forecasting experts, primarily from the fields of statistics, econometrics, and engineering, have dedicated their efforts to it. It goes without saying that the enormous financial interests of the energy industry have accelerated the development of EPF, especially with the proliferation of renewable energies, which now constitute a global business. For a recent and detailed review of energy forecasting as a whole field, please see [12].

We must begin by noting that, in this context, electricity takes on the characteristics of an unusual commodity. Unlike oil or natural gas, it is not easily storable or preservable. Once produced, it needs to be transmitted, distributed, and

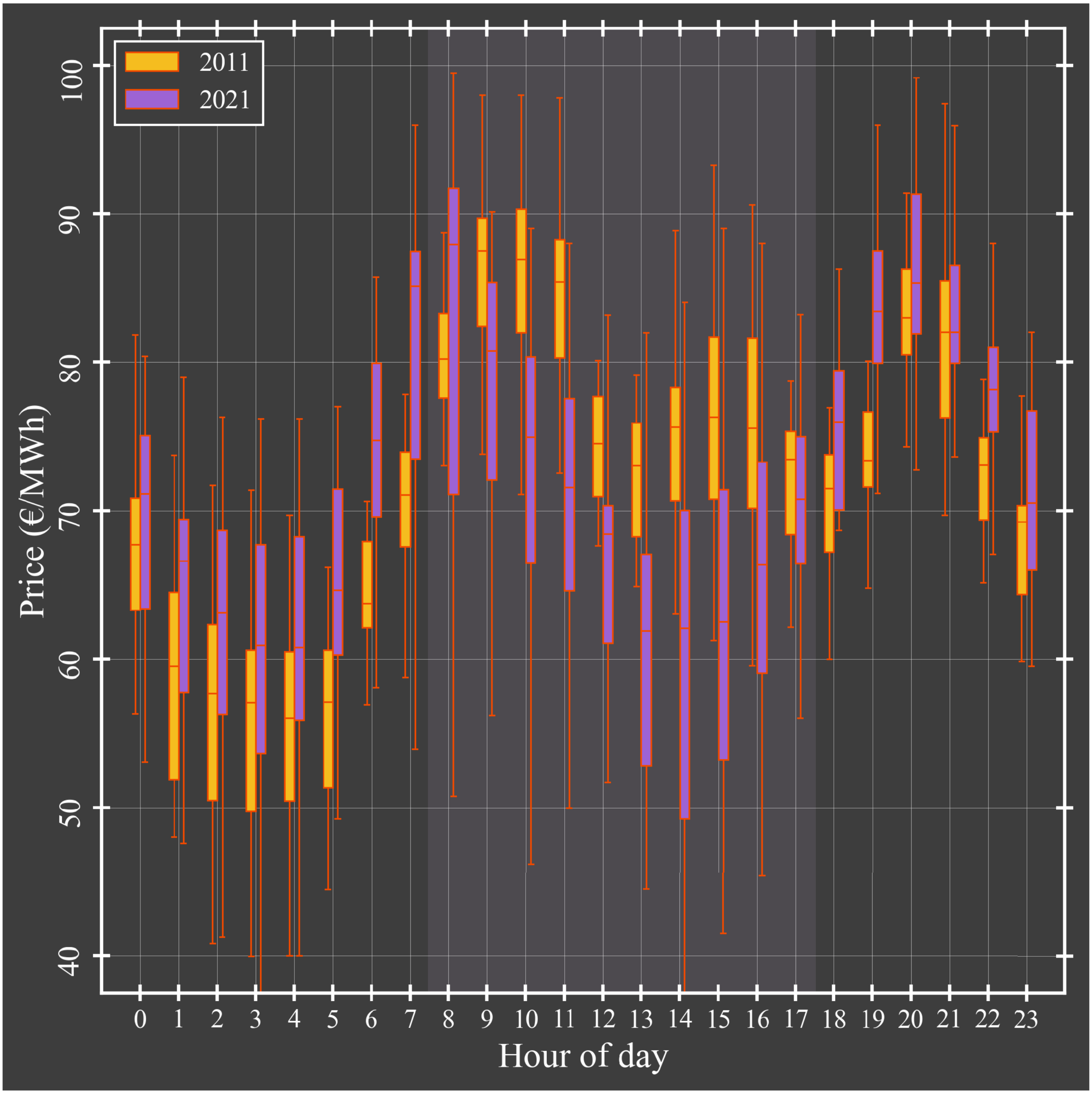

FIGURE 2. Evolution of hourly electricity prices in may (2011 vs 2021). The figure depicts the evolution of daily price profiles on the MGP, the Italian day-ahead electricity market, from 2011 (in turquoise) to 2021 (in purple). It is interesting to note that the percentage of power generated from photovoltaic sources, relative to the total installed generation capacity in Italy, has significantly increased over the years. In 2011, it accounted for less than 2%, while in 2021, it has surpassed 10% [6]. Each hourly period in the y-axis corresponds to two boxplots, each representing the distribution of prices for a specific year. Specifically, the box represents the interquartile range (IQR), with the median marked by a line within the box. Whiskers extend from the box to the minimum and maximum values within 1.5 times the IQR. Outliers, namely values falling outside the whiskers, are here omitted for clarity. The significant increase in photovoltaic generation capacity has caused a decrease in price during the central hours of the day, when irradiation is usually abundant, as is evidenced by the light grey area in the figure. Also, a higher level of volatility can be observed, with significantly larger IQRs for 2021. By contrast, hours with few or no solar generation available exhibit higher average prices. Considering the EU's contemporary energy policies, these tendencies are expected to increase in the near future.

consumed, causing network issues if there are imbalances between demand and supply. Despite the existence of electricity storage systems, these storage solutions are not yet efficient enough for widespread utilization on a large scale. This implies a substantially more challenging handling and management, resulting in time series that are more difficult to forecast. A solid contribution to electricity forecasting can be found in [26], where the singular characteristics of this commodity are exposed. The author explains that forecasters are mainly interested in two main variables related to electricity: price and load. Specifically, electrical load forecasting is a much more established discipline. The intrinsic nature of the time series describing national-scale electric loads allows the construction of accurate point forecasting systems. Load is, in fact, much more stable and exhibits strong regularity patterns, which can be effectively captured by conventional statistical methods. It is also important to note that load forecasting historically predates EPF, both as an industrial necessity and as an academic discipline. This is because accurate load forecasts are not only useful for market purposes but also, and especially, for the management of the electric grid. Notably, the initial methods employed for EPF were indeed those statistical techniques previously developed for load forecasting.

The substantial interest in price forecasting arises with the introduction of deregulated markets, and the first systematic review of EPF, with an emphasis on point forecasting and day-ahead markets, is relatively recent [27]. Before introducing further distinctions, it is important to specify that day-ahead EPF, and therefore our work, falls under the category of short-term forecasting. This involves forecasting horizons ranging from a few hours to a few days. On an industrial level, short-term forecasting is valuable for daily supplies and day-to-day speculative trading. Mid- and long-term forecasting prove useful in other applications, such as risk management and investment planning, respectively, but this goes beyond the scope of our article.

The author of [27] notes that in the decade preceding his work, research in EPF witnessed significant growth, with a plethora of diverse solutions proposed. However, he highlights a general lack of coherence in the field, primarily due to the differences among various markets analyzed and the lack of shared criteria for choosing and evaluating the approach to modelling the problem. Despite efforts made in this direction over the last ten years, unfortunately, a standard has not yet been achieved. This situation is primarily due to existing differences between countries and the diverse backgrounds of researchers, who are divided between engineers and statisticians, each group with its own practices and customs.

As anticipated above, our solution is based on kernel methods, which undergo the class of the so-called Computational Intelligence (CI) or nonlinear statical algorithms. By contrast, the most established class is that of linear statistical methods, namely linear models. It is reasonable to consider them as benchmark models because the superiority of CI over such methods is still an open question [10]. While CI methods can capture complexity and nonlinear relationships, it is not guaranteed that, in practice, they provide better forecasts than regression models. Among CI methods, the most popular solutions include various types of Neural Networks (NN) [21], [27], including multi-layer perceptrons, recurrent NNs and convolutional NNs. Our choice to use kernel methods instead of NNs depends on several factors. Firstly, kernel methods are conceptually simpler, easy to implement, and generally less data-intensive. Furthermore, recent literature shows that, in the Italian context, nonparametric kernel methods perform better than NNs [14], while the superiority of kernel methods to autoregressive linear models is still debated [10].

In particular, the sample efficiency of NNs is a well-known and actively studied problem. Deep learning models, in particular, require the design of certain data preprocessing techniques to achieve optimal performance. To cite some recent examples, in [28], it is highlighted how spikes in price series are highly pathological for model training, and an oversampling system is devised to correct prices. In [18], a generative model is used for data augmentation, creating synthetic data with characteristics similar to historical data. Another recent work explicitly addressing the issue of sample efficiency in deep NNs is [22]. In this work, it is explained that the problem of data insufficiency is particularly pronounced in the field of EPF, mainly due to the low-to-moderate sampling frequency (hourly data) and because market rules and operator preferences usually change very quickly, rendering old data often useless for predictions. In this case as well, the proposed solution relies on data augmentation through the generation of synthetic data. Without delving into the topic, as it is beyond the scope of this article, we limit ourselves to expressing scepticism regarding data augmentation for EPF applications.

An example that differs from the previous ones but is quite similar to the proposed work is [7], where the employed model is a variant of the General Regression Neural Network, also a non-parametric kernel method. However, in this work, a rather intricate optimization mechanism is proposed, and, most importantly, the analysis focuses on American data, which is quite different from the Italian context and predates 2018. Another important difference is that the authors of [7] only perform point forecasting, while ours includes both point forecasting and probabilistic forecasting.

Other approaches include, but are not limited to, simulation models, both statistical and multi-agent-based, fundamental models that explicitly describe demand and supply dynamics, and similar-day methods. A recent example of a simulation-based approach is [20], where probabilistic forecasts are generated by aggregating simulations conducted using well-known econometric models (e.g. GARCH-type). The markets considered are the German and Austrian markets, and, once again, the data under investigation predates 2018. The fundamental approach has been recently applied in [5], where bid data from the Spanish day-ahead electricity market are interpolated with simple statistical models to reconstruct the aggregated demand and supply curves. At that point, the price forecast is determined by the intersection of the curves and evaluated using various methods. On the contrary, multi-agent models simulate on a large scale the collective behaviour of many market operators, using it to predict prices.

Similar-day methods involve searching in a historical database for the most similar day (or days) to the current day, considering factors such as the day of the week, load, weather and other indicators, and using them as forecasts. The employed kernel-based method could also be considered a similar-day approach, as certain kernel functions can be treated as affinity measures. In a broader context, it is important to remember that there are no strict distinctions between methods, and often the approach is hybrid.

Moving on, another important degree of freedom in EPF concerns the choice of variables used as predictors: even in this case, there is no common rule. Despite the superiority of multivariate models with respect to univariate models is still an open question [9], the inclusion of exogenous variables, i.e., variables different from the price series itself (the endogenous variable), is predominant. Variables related to load and/or supply are often added, as well as weather variables (e.g. temperature, wind speed, irradiation) or prices of other commodities such as fuels.

It is now crucial to note that the choice is highly dependent on the market under consideration, meaning that from state to state, the most significant exogenous variables change. For example, different European countries feature substantially different generation mixes, thus directly influencing price formation. There are also extreme cases, such as a study on the English market [19], which suggests the limited usefulness of exogenous variables for that specific market. The study [11] examines another factor believed to influence market price formation, namely the effect of the integration of European electricity markets. The authors focus on the Dutch case, but more generally, they demonstrate how the performance of various models, both statistical and CI, varies across different markets. In our case, cross-border exchanges and foreign prices were not included as they did not lead to improvements in forecasts.

As explained in Sec. V, our choice depends on both fundamental considerations and a series of empirical analyses, and what we obtained is consistent with the literature on Italian day-ahead EPF. Specifically, the authors of [11] take into account load forecasts and natural gas prices, while in [14] the feature set also includes calendar variables. While we adopt supply proxies, namely solar and thermal generation forecasts, we did not include neither the widespread load forecasts, nor calendar variables, as no performance improvement was observed.

The only widely used variable that we did not consider here is the weather variable, primarily because weather forecasts are typically available to companies on payment and are not easily accessible to researchers. However, recent research in [23] clarifies that short-term EPF models, including ours, usually incorporate weather variables indirectly. For instance, instead of directly considering solar irradiance, they might use photovoltaic generation, or replace wind speed with wind generation, and so on. In the same study, it is explained that pure numerical weather predictions are useful in generating few-day-ahead forecasts, because estimates for NPRES generation are often available only for the next day.

Regarding the choice of temporal lags for the price, despite we restricted our attention to the model of order one, they empirically proved effective at capturing the intrinsic daily seasonality of the data [10]. Explicit decomposition of the series into trend, seasonality, and residual components has not been performed, as this practice is more characteristic of statistical approaches. However, there are several exceptions, even in recent works, where researchers combine CI models with seasonal decomposition. For example, the authors of [18] hybridize a generative NN with seasonal decomposition, in addition to the frequency analysis carried out with Variational Mode Decomposition for each component.

Even more recent is the interest and systematic study of probabilistic forecasting methods, which, as mentioned in the previous section, arises from the need in markets destabilized by less controllable generation sources. Despite the increasing interest in this approach, probabilistic EPF is still an underdeveloped topic. To date, the most detailed review in probabilistic EPF is [21]. The authors describe the existing approaches to the problems, which can significantly vary in nature, as is for point forecasting methods. The generation of probabilistic forecasts can be accomplished through various approaches, with the two main ones being Prediction Intervals (PI), representing ranges of values associated with a given probability, and forecasting the entire distribution of the variable of interest (which can encompass all PIs as special cases). As we will see later, the proposed method falls into the second category but will be subsequently employed to PIs, which are the most common form of probabilistic forecasting. This is because PIs are directly interpretable and easier to evaluate. More precisely, assessing the performance of probabilistic forecasting models is a broad and open topic, and interested readers are referred to [21] and [25].

Notably, in [21], it is also reported that the largest share of probabilistic EPF works rely on NNs, as was for the very first publication on the topic [8]. In this regard, our work seeks to differentiate itself by preferring a comparison of simple KR methods and linear models. Specifically, the method employed for distribution estimation belongs to the same family of estimators [17], [24] as the method used for KR point forecasting, facilitating understanding and implementation.

III. METHODOLOGY

Adopting for a moment a statistical view, the development of a comprehensive EPF system involves solving two regression problems: mean regression, to output point forecasts, and Quantile Regression (QR), to construct flexible distribution estimates, in a formally rigorous manner and without Gaussianity assumptions. Our decision to adopt kernel methods is based, on one hand, on the desire to leverage the advantageous properties of CI methods mentioned earlier and, on the other hand, on the choice to avoid more complex approaches, such

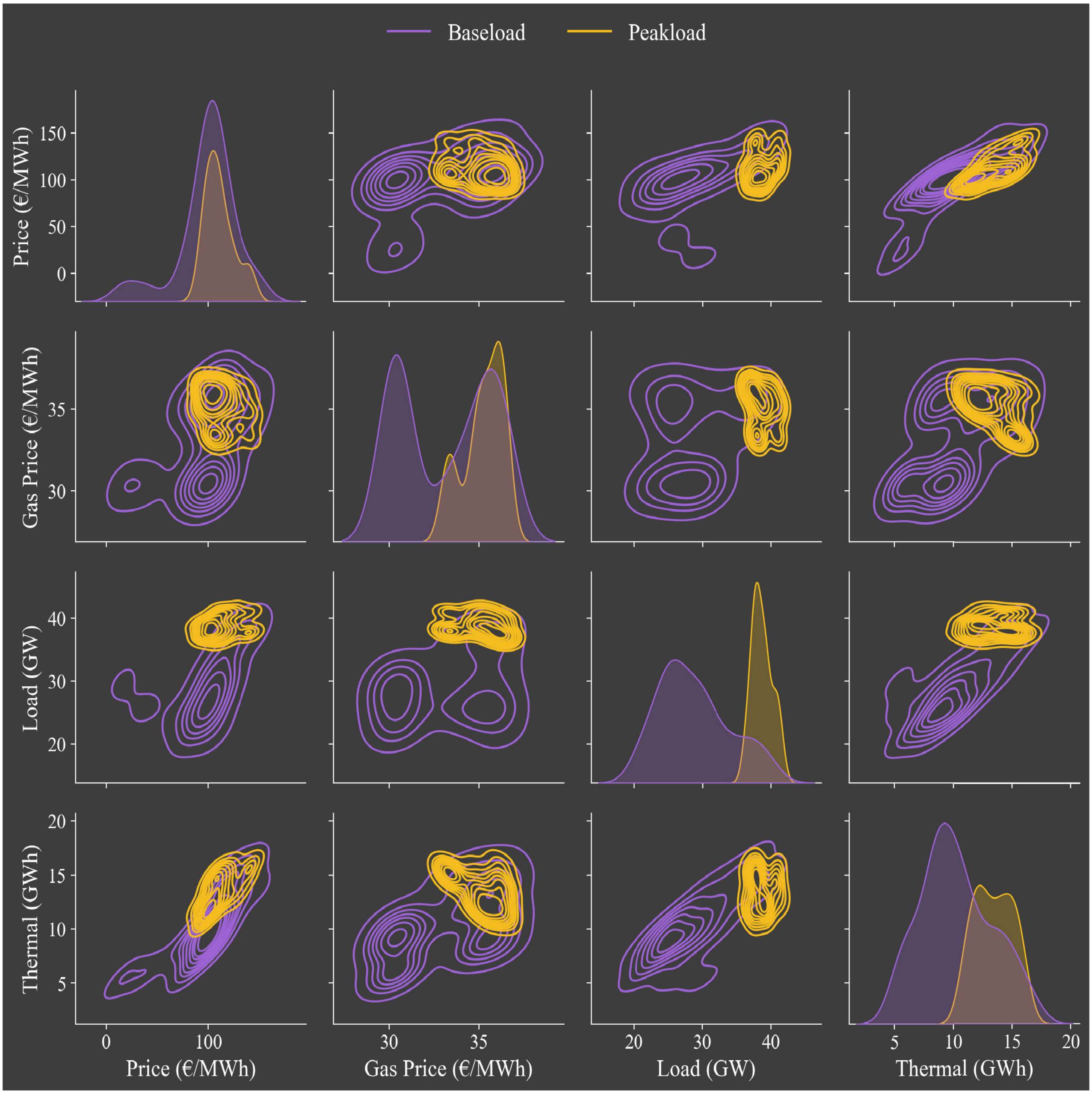

FIGURE 3. Disparities between peakload and baseload (may 2023). The figure highlights the distinction between periods categorized as "peakload" and "baseload", using May 2023 data as a reference. In Italy, the peak period spans from 8:00 AM to 8:00 PM, Monday to Friday, while the off-peak period includes the remaining hours and the entire weekend. The chart presents a matrix of graphs with identical variables on both the vertical and horizontal axes, namely electrical energy price (Price), natural gas price (Gas Price), national load (Load), and thermal generation (Thermal). Baseload data is represented in purple, while peakload data is depicted in yellow. Along the main diagonal, Kernel Density Estimation (KDE) plots for each considered variable are showcased, reconstructed using a Gaussian kernel. In the off-diagonal entries (i.e. entries corresponding to different variables), contour plots illustrate the two-dimensional KDE applied to the respective variable pairs, providing a visualization of the joint probability distribution. The graph explains the rationale behind the terminology "peakload" versus "baseload": focusing on the 'Load' variable, the data for the baseload period (in purple) exhibit a lower mean and higher variance compared to the data for the peakload period (in yellow). This distinction is highly significant in the industrial world, and many wholesale contracts are based on it: also, notice the bimodality of natural gas prices.

as those based on the NN family, which might be excessively challenging to implement, without featuring clear benefits. As better explained below, choosing kernel methods enables the construction of methods that are conceptually very similar for both the mean and QR forecasting.

As an additional note, it is worth mentioning that the QR problem has been extensively explored by the author of [16], who extended linear models for mean regression problems to predict quantiles.

A. THE NADARAYA-WATSON ESTIMATOR

The selected model for point forecasting is the Nadaraya-Watson Estimator (NWE), also known as the KR estimator [4]. It is one of the simplest models belonging to

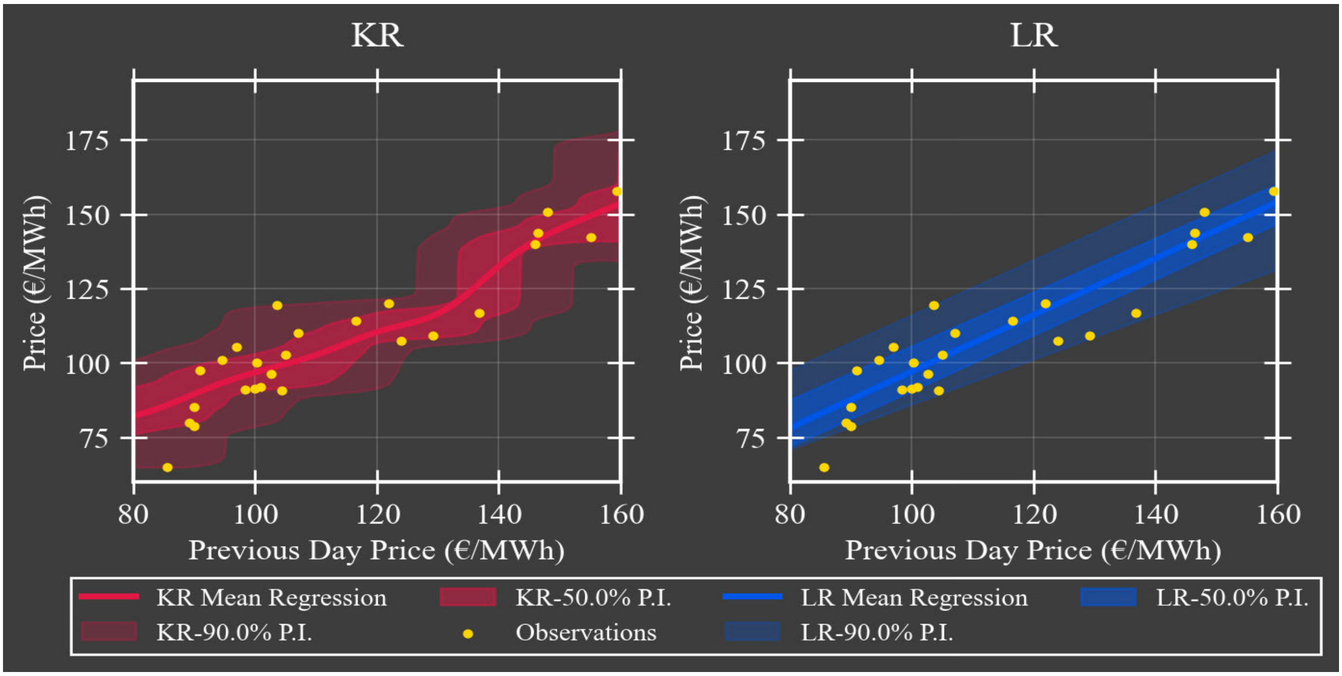

FIGURE 4. **Univariate regression analysis for electricity price at wednesdays, 12.00 PM, with respect to Previous Day Price, at the same hour. The figure presents a visual comparison of Linear AutoRegression of order one (AR(1) and Kernel Regression (KR) models. On the right, results from LR are displayed in blue, while on the left, KR outcomes are showcased in red. This analysis focuses on the prediction of prices for Thursdays at noon, spanning from January to August 2023. Both models exhibit mean regression and central Prediction Intervals (PIs) at 50% and 90%. The initial set of PIs encompasses the 25th to 75th percentiles, while the subsequent set ranges from the 5th to the 95th percentiles. Despite the inherent differences between these two approaches, the results of univariate regressions appear remarkably similar. This similarity is explained by the evident linear correlation between prices on day d and the preceding day $d-1$. The univariate linear model, commonly referred to as AR(1), is recommended as a baseline due to its simplicity and satisfactory performance when compared to more complicated models.**

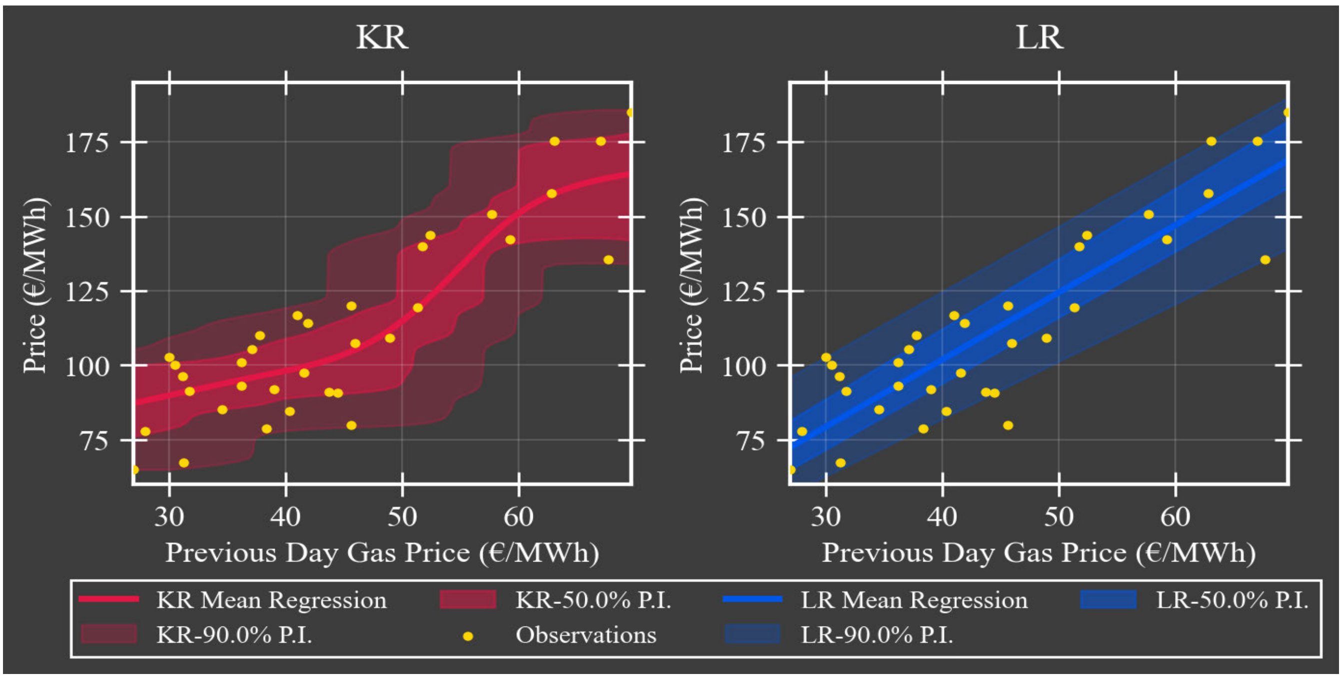

FIGURE 5. **Univariate regression analysis for electricity price at wednesdays, 12.00 PM, with respect to Natural Gas Price (previous day average). In the figure above, Kernel Regression (KR) results, depicted in red, are compared to Linear Regression (LR) in blue for the problem of predicting electricity prices using the previous day's natural gas prices. Prediction Intervals (PIs) at 50% and 90% accompany mean regression. The data span from January to August 2023, included. KR depicts an apparent non-linear trend, noticeable in both mean predictions and quantiles. The acceleration in electricity price variations seems to occur notably for MGP-GAS prices surpassing e50. It is important to notice that despite the significant growth in photovoltaic and wind generation systems, Italy predominantly relies on thermoelectric generation through natural gas power plants for electricity production.**

the class of kernel methods (e.g. Support Vector Machines, Gaussian Processes), and it is intended to predict the expected value of a random variable. More precisely, we study the relationship between one dependent random variable Y and one predictor variable X. Initially, we want to estimate the expectation, or mean, of the response conditioned on a given value x taken by X. In other words, our goal is predicting

$$\mu(x) = \mathbb{E}[Y|X = x]. \tag{1}$$

Given training data $\{(x_i, y_i)\}_{i=1}^n$, smooth estimates are provided by averaging the observed values of the response variable y_i, weighted by kernel functions centred around each data point x_i:

$$\hat{\mu}(x) = \frac{\sum_{i=1}^n K_h(x, x_i) y_i}{\sum_{j=1}^n K_h(x, x_j)}, \tag{2}$$

where $K_h(x, x_i)$ is a kernel function with bandwidth parameter $h > 0$. Its purpose is to measure the similarity between x, the point at which the estimation is made, and x_i, the points or situations known and used as training data. Specifically, we choose the Radial Basis Function (RBF), also known as the Gaussian kernel, the most popular choice in nonlinear modelling. For arbitrary input $a, b \in \mathbb{R}$, the RBF is a function of the euclidean distance between a and b:

$$K_h(a, b) = \exp\left[-\frac{(a-b)^2}{2h^2}\right]. \tag{3}$$

It is important to highlight that the bandwidth parameter h determines the behaviour of the model and, therefore, the goodness of fit with respect to data. For this reason, we must focus on choosing an optimal value of h, a topic extensively studied in literature [17] and known as the bandwidth selection problem for the NWE. While a common approach is based on minimizing the cross-validated Mean Squared Error (MSE), this method can be misleading, especially when dealing with limited amounts of noisy data. We address this concern by opting for the Akaike-Hurvich Information Criterion (AIC) [13], which can be stated as follows:

$$\begin{aligned}
\min_h \quad & \mathrm{AIC}(h) = \log(\hat{S}^2) + \frac{1 + \frac{\mathrm{Tr}(H)}{n}}{1 - \frac{\mathrm{Tr}(H)+2}{n}} \\
\text{where} \quad & \hat{S}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_h(x_i))^2 \quad \text{(training MSE)} \\
& H = \left(K_h(x_i, x_j)\right)_{i,j=1}^n \quad (n \times n \quad \text{kernel matrix}),
\end{aligned} \tag{4}$$

where $\mathrm{Tr}(H)$ denotes the trace of matrix H. This simple optimization problem can be solved using any algorithm; in our case, we choose Differential Evolution.

B. THE PARZEN-STONE ESTIMATOR

The NWE allows estimating the expected value (or mean) of a random variable. This corresponds, in the context of forecasting, to producing a point estimate of the variable of interest, i.e., a single numerical value. Reflecting on the meaning of such an estimate, we can state that it is inherently incorrect, as a single point in a continuous spectrum of values has a probability of realization equal to zero. In our case, we are interested in predicting the price of electrical energy for the next day, and the high volatility observed can make point estimation less meaningful, as introduced in the previous sections. Rather, we are interested in estimating the full probability distribution of the next day's price, and this can be achieved as follows.

Let us first recall that, for any target variable Y and any realization x of the input variable X, the conditional Cumulative Distribution Function (CDF) is defined as

$$F(y|x) = \mathbb{P}(Y \leq y|X = x), \quad y \in \mathbb{R}. \tag{5}$$

It is interesting to note that the CDF of Y given $X = x$ can be written as the conditional expectation of a random variable, namely

$$F(y|x) = \mathbb{E}[I(Y \leq y)|X = x], \tag{6}$$

where $I(A)$ is the indicator function, such that $I(A) = 1$ if A is true, otherwise $I(A) = 0$. This motivates the estimation of the CDF $F(y|x)$ by means of the NWE, as introduced in eq. (2):

$$\tilde{F}(y|x) = \frac{\sum_{i=1}^n K_h(x, x_i) I(y_i \leq y)}{\sum_{j=1}^n K_h(x, x_j)}. \tag{7}$$

As was extensively discussed by the authors of [17], the estimator in the equation above is limited by the fact it does not smooth the dependent variable Y. To deal with this issue, the estimator $\tilde{F}$ can be easily extended through any smooth kernel cumulative distribution $G_{h_0}(y)$:

$$\hat{F}(y|x) = \frac{\sum_{i=1}^n K_h(x, x_i) G_{h_0}(y - y_i)}{\sum_{j=1}^n K_h(x, x_j)}, \tag{8}$$

where $h_0 > 0$ is the bandwidth associated with the response variable Y. In our case, G is the standard Gaussian CDF:

$$G_{h_0}(y) = \frac{1}{\sqrt{2\pi}h_0}\int_{-\infty}^{y} \exp -\frac{u^2}{2h_0^2} du. \tag{9}$$

Throughout this work, we will refer to the estimator $\hat{F}$ eq. (8) as the Parzen-Stone Estimator (PSE), as it belongs to a broad class of methods for nonparametric regression discussed in detail in [24], all of which rely on the established Parzen window theory.

IV. DATA AND EXPLORATORY ANALYSIS

In this section, we will list the variables considered in the study, describing their sources and explaining why they were chosen for the final regression model. Specifically, the final set of predictors included:

- previous day PUN prices P^{d-1}, expressed in €/MWh;
- previous day natural gas prices G^{d-1}, in €/MWh
- solar generation forecasts $\hat{S}^d$, in GWh;
- thermal generation forecasts $\hat{T}^d$, in GWh.

We will also present, for both selected and non-selected variables, a univariate regression analysis for exploratory purposes, providing initial motivations for their inclusion as regressors. In other words, the PUN P represents the

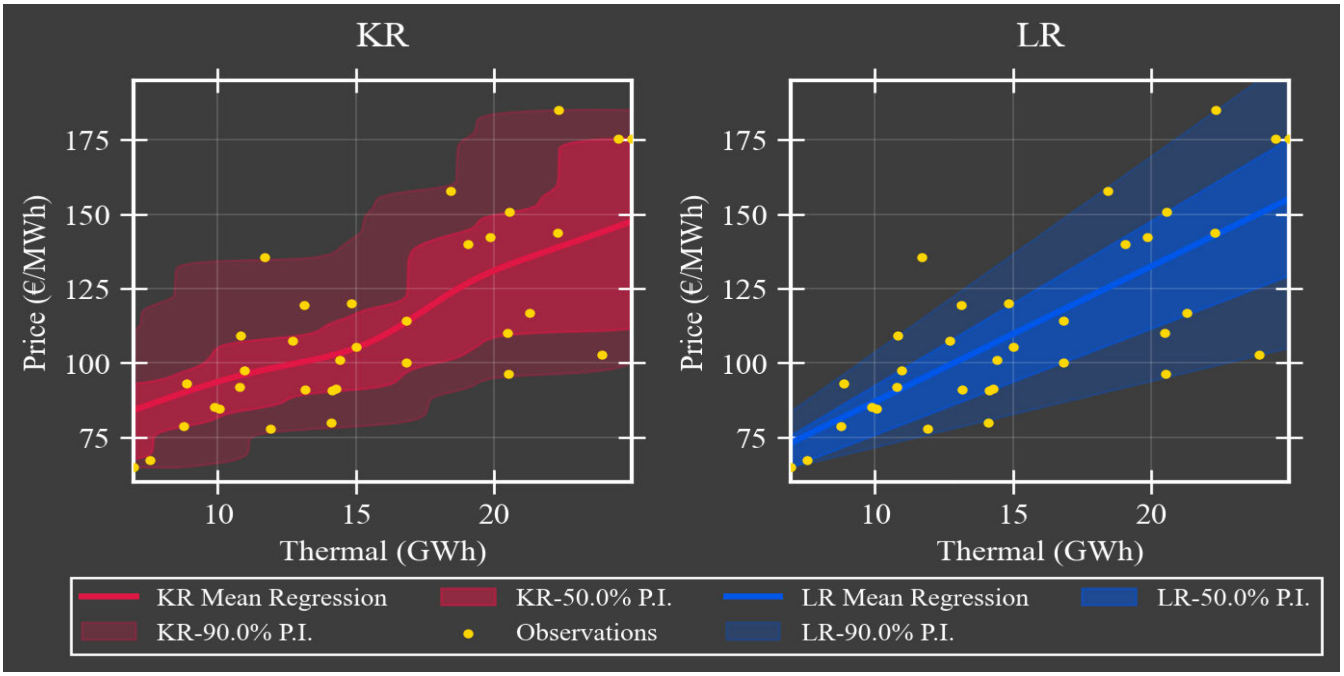

FIGURE 6. **Univariate regression analysis for electricity price at wednesdays, 12.00 PM, with respect to Thermoelectric Generation, at the same hour. The figure illustrates the core influence of thermal generation, constituting 40% of Italy's electricity production, on the fluctuation of electricity prices. Specifically, the price volatility increases significantly during periods of high generation. These periods often signify a shortage of generation from other sources, making even small variations in thermal supply notably influential on prices. The phenomenon is captured both by KR in red and LR in blue, and even for this variable, the results are not too different. As was in the previous figures, the data cover the first eight months of 2023.**

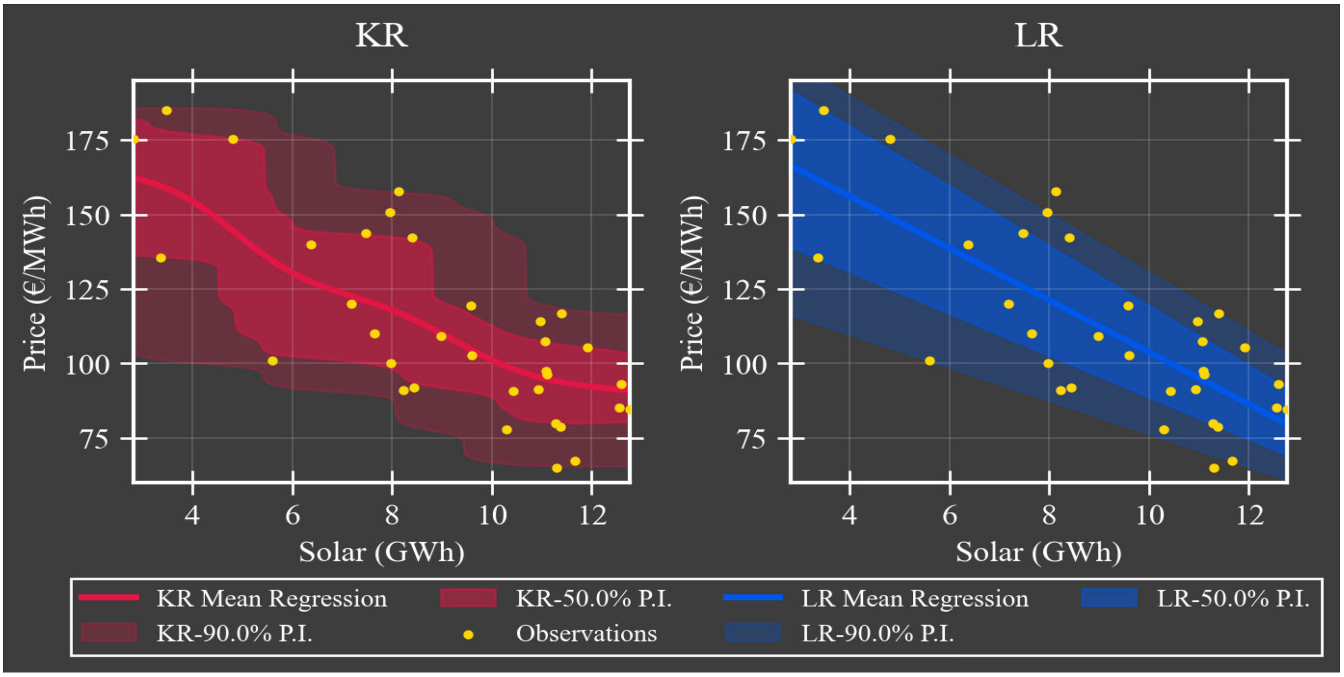

FIGURE 7. **Univariate regression analysis for electricity price at wednesdays, 12.00 PM, with respect to Photovoltaic Generation, at the same hour. With photovoltaic systems now constituting 12% of Italy's total installed capacity [6], the trend is opposite to that of thermal generation. During high supply periods, prices tend to be low and exhibit low volatility. This phenomenon is captured by both KR (in red) and LR (in blue), and the results for this variable are not significantly different between the two models. Also in this case the data range from January to August 2023.**

endogenous variable, while all other variables which can be eventually included as regressors are termed exogenous.

As explained in Sec. II, the selection of variables used as predictors is an open problem, or at least a problem that

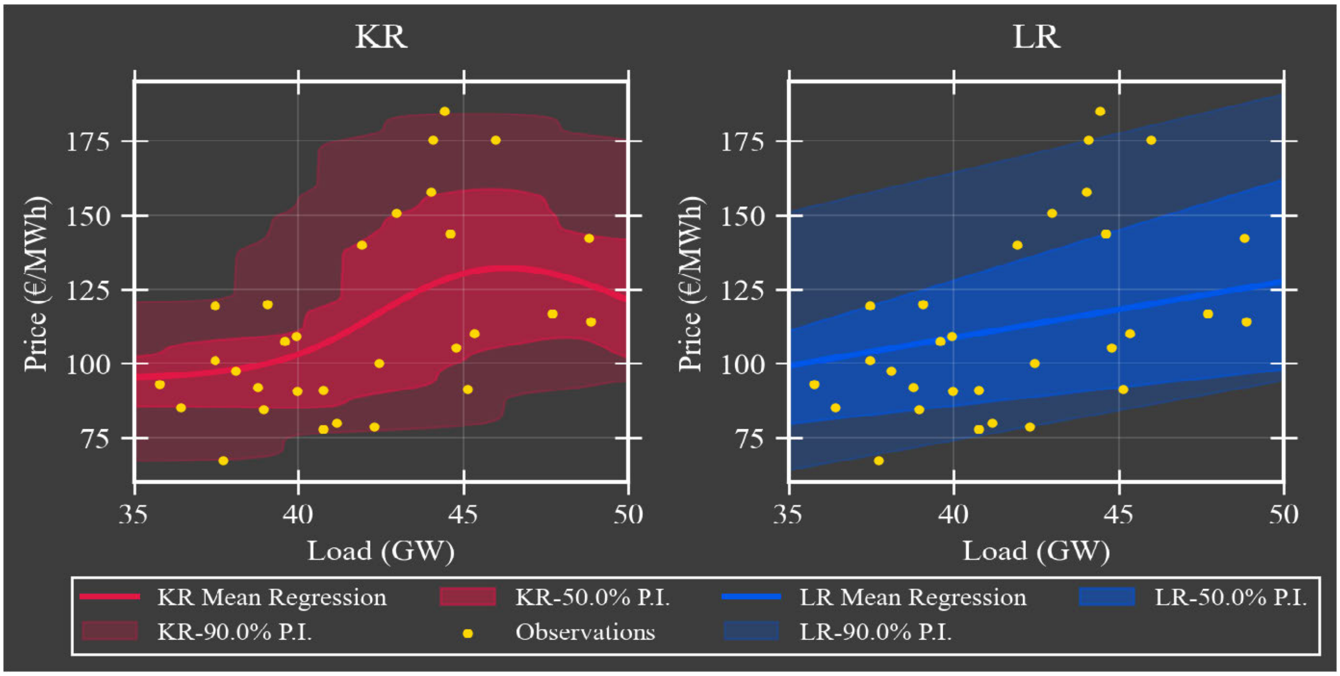

FIGURE 8. Univariate regression analysis for electricity price at wednesdays, 12.00 PM, with respect to National Electric Load, at the same hour. In this case, the univariate analysis focuses on the price against the national electrical load, a variable that likely represents one of the most commonly used predictors aside from the past price values. Despite much of the literature converging on the use of the load, we excluded it because, as shown in the figure, it does not exhibit a clear correlation either with the average prices or their distribution. This holds true for both KR (on the left in red) and LR (on the right in blue). However, we need to clarify that this result holds under the conditions of the contemporary Italian market, and the correlation of price with load was more significant in past years. In fact, amidst the drastic transformation of electricity prices, the Italian load has remained relatively stable over the last five years [6], with exceptions during lockdown periods.

does not admit general solutions, i.e., solutions that are independent of the market under consideration and the historical moment in analysis. In our case, we began with a selection of variables derived from fundamental reasoning, which was then refined through empirical analysis of the examined data. Since the market price is determined by the intersection of demand and supply, we started with variables that could act as proxies for demand and supply. In one case, identifying the variable is straightforward: demand is represented by the national electrical load. On the other hand, as the source of energy supplied to the market is highly variable, we considered thermal and photovoltaic generation, which are highly relevant in the contemporary Italian context. Last but not least, the price of natural gas was included because it is nowadays the primary fossil fuel for electricity generation in Italy. On the contrary, the load was excluded from the multivariate regression model, as it did not improve the final performance, neither in point nor in probabilistic sense.

Given that a secondary objective of this work is to assess the effectiveness of nonlinear prediction models in EPF, we present, for comparison, the results achievable with linear methods in the univariate analysis. We anticipate that there are no significant differences in the final performance, despite the modelling apperars qualitatively better when using nonlinear tools. The analyses presented here refer to data ranging from January to August 2023.

A. DAY-AHEAD ELECTRICITY PRICE

Let us start with the target variable, the MGP equilibrium price, whose past values are also included as regressors. More precisely, let us number days from Mondays to Sundays, namely $d = 0, \ldots, 6$, with the convention that if $d = 0$ (Monday), the $d - 1 = 0$ (Sunday). We are predicting the price at hour h on day d, using as inputs the prices at the same hour h on the previous day $d - 1$.

From now on and throughout this section, for explanatory purposes, we set the goal to predict the price for Thursday (d=3) at noon (h=12). Also, the data presented in the analysis cover the period from January to August 2023, namely the first eight months of the year. Consider Fig. 4: the results of KR are on the left, depicted in red, while the results of LR are on the right, represented in blue. For both models, we report the mean regression and the central PIs at 50% and 90%. The first set of PIs has limits at the 25th and 75th percentiles, while the second set has limits at the 5th and 95th percentiles. Despite the two approaches are substantially different in nature, the results of the univariate regressions are very similar. This similarity arises from the evident linear correlation between the price on the day d and the price on the previous day $d-1$. In the linear case, this univariate model is often referred to in the literature as an AutoRegressive (AR) model of order one, denoted as AR(1). Given its extreme simplicity and the relatively good results it achieves, it is

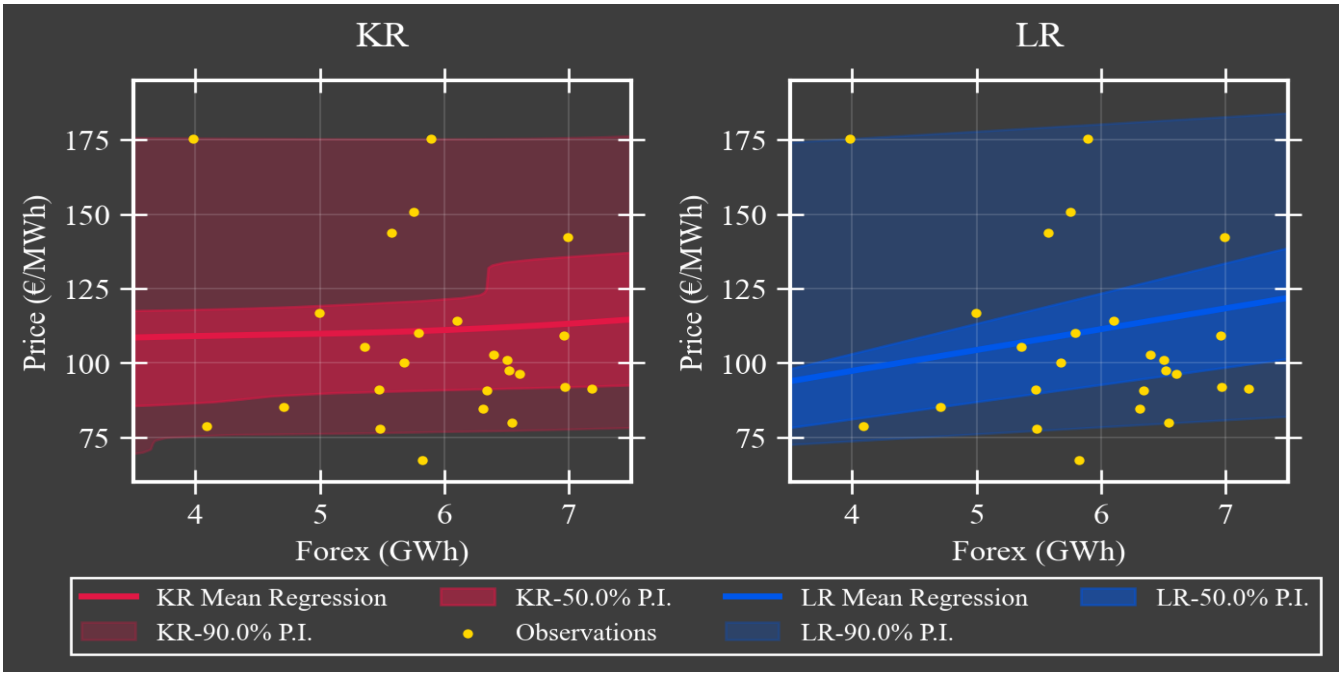

FIGURE 9. **Univariate regression analysis for electricity price at wednesdays, 12.00 PM, with respect to Net Foreign Exchange, at the same hour. The Forex variable represents the net cross-border exchanges, providing information into the interconnectedness of electricity markets. To date, Italy is electrically linked with France, Switzerland, Austria, Slovenia, Montenegro, and Greece through 25 interconnection lines [2]. When it is said, even in popular discourse, that Italy "imports energy", it means that variable Forex is often positive, meaning that it is more cost-effective to import from abroad. This is especially true in countries like France and Switzerland, where nuclear energy is abundant. In the examined data, which spans from January to August 2023, there are no significant correlations with the energy price, and consequently, no model can derive benefits from this variable.**

advisable to use it as a baseline when employing more complicated models.

Although [27] emphasized the importance of incorporating the price two days prior ($d-2$) and one week prior ($d-7$) in regression models for short-term EPF, we found no empirical justification for their inclusion. The use of prices at lags 2 and 7 should allow the model to capture better the daily and weekly seasonality of the PUN series. Still, in our multivariate models, including linear ones, their addition did not enhance final performance.

Price data are freely available for download on the website of the Gestore dei Mercati Energetici (GME) [1], the Italian, state-owned corporation that manages almost all national commodity exchanges, including the MGP and the day-ahead natural gas market (MGP-GAS, see below). Data are regularly updated, on a daily basis.

B. NATURAL GAS DAY-AHEAD PRICE

Despite the strongly increasing amount of photovoltaic and wind generation systems, the main method for electricity production in Italy is thermoelectric generation with natural gas power plants. Accounting for 40% of total electricity generation, thermal generation can be considered the backbone of the Italian generation systems since the early 2000s. It is useless to write that the Italian electricity price is strongly correlated with the price of natural gas, as is shown in Fig. 5. In this case, KR in red on the left captures what appears to be a non-linear trend, both for the mean prediction and the quantiles: the change in electricity prices appears to become faster for gas prices exceeding €50. Additionally, KR predicts wider PIs compared to LR.

In this study, we refer to the spot price of gas in the Italian day-ahead market, MGP-GAS, on the day prior to prediction $d-1$; denote it by G^{d-1}. Notice that while the electricity market data have hourly resolution, gas prices are available as daily averages. This is because the day-ahead gas market is not structured into hourly periods. The MGP-GAS market, marked by increasing volumes over the years, is highly liquid. Consequently, MGP-GAS prices partially reflect the costs associated with gas-based electricity generation. Also these data have been obtained through the GME website [1], where the interested reader can find further detail about the MGP-GAS functioning, not included here for brevity. It is important to note that Italy does not have active nuclear power plants; if it did, the impact of gas prices on electricity prices might be less significant.

C. THERMAL GENERATION FORECAST

The thermal generation variable T is included for similar reasons to the previous one. However, it is important to note that only a portion of gas employed in thermal generation is procured in the day-ahead market, MGP-GAS, and actual generation levels may vary. Other factors influencing thermal supply variations include plant-specific technical operating limits, as well as scheduled or unexpected maintenance periods, and failures. Consider Fig. 6: thermal generation, representing 40% of Italy's electricity production [6], has a substantial effect on electricity price

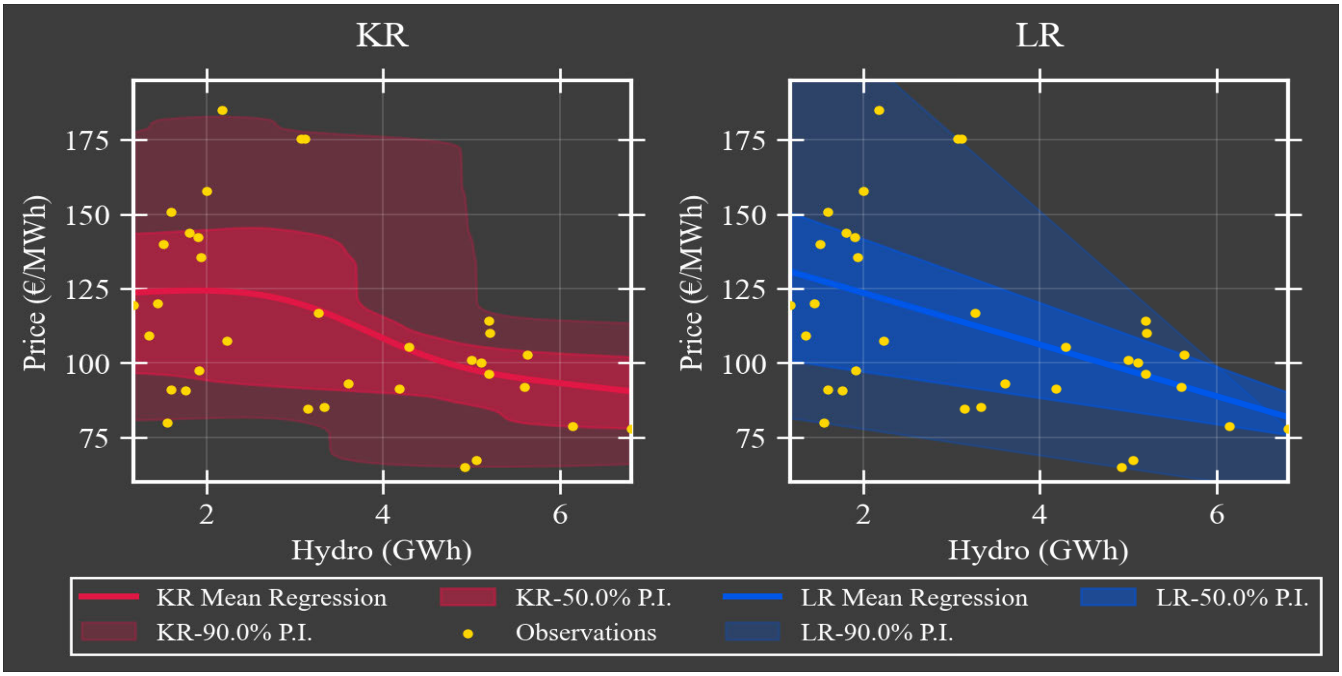

FIGURE 10. Univariate regression analysis for electricity price at wednesdays, 12.00 PM, with respect to hydroelectric generation, at the same hour. Hydro generation, depicted in blue for LR and red for KR, exhibits a behaviour partly similar to solar generation, being bearish and less volatile at high generation levels. However, KR provides more realistic modelling in this case. Despite this, the variable was excluded from the final model as it did not contribute significant improvements. Likely, the drought in recent years, affecting countries in the Mediterranean region, encourages energy producers to invest in alternative sources. In other markets, the impact of this variable could be more significant, as observed, for instance, in the Nordic European electricity markets.

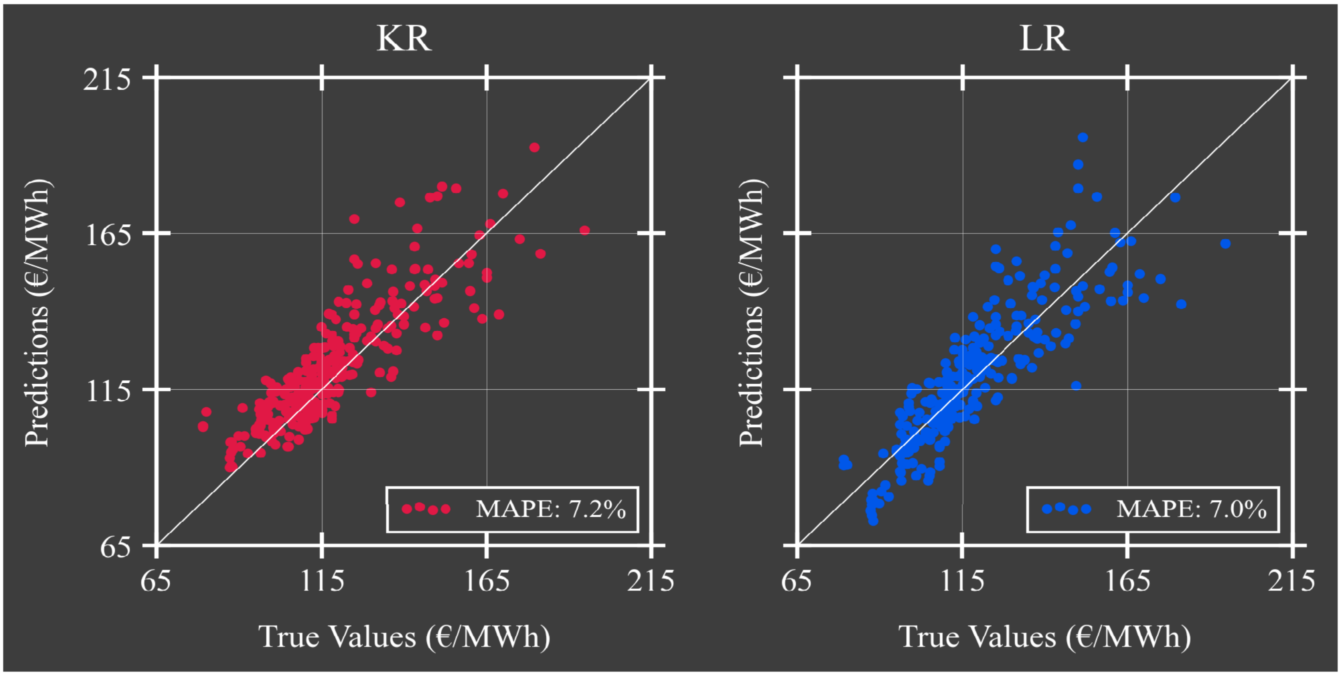

FIGURE 11. Multivariate regression results for september 2023 test period: baseload. The figure illustrates predictive performance over the baseload hours of the test period, set in September 2023. The performance achieved by the two models is very similar. The LR methods produce biased predictions for lower price levels, with a tendency to overestimate. This is not a critical issue from an industrial standpoint, as lower price levels are always associated with inferior operating risks.

fluctuations. The results demonstrate that price volatility intensifies notably during high thermal generation periods, indicating a scarcity of generation from alternative sources. This influence is depicted by both KR in red and LR in blue, revealing relatively similar outcomes for this variable as well.

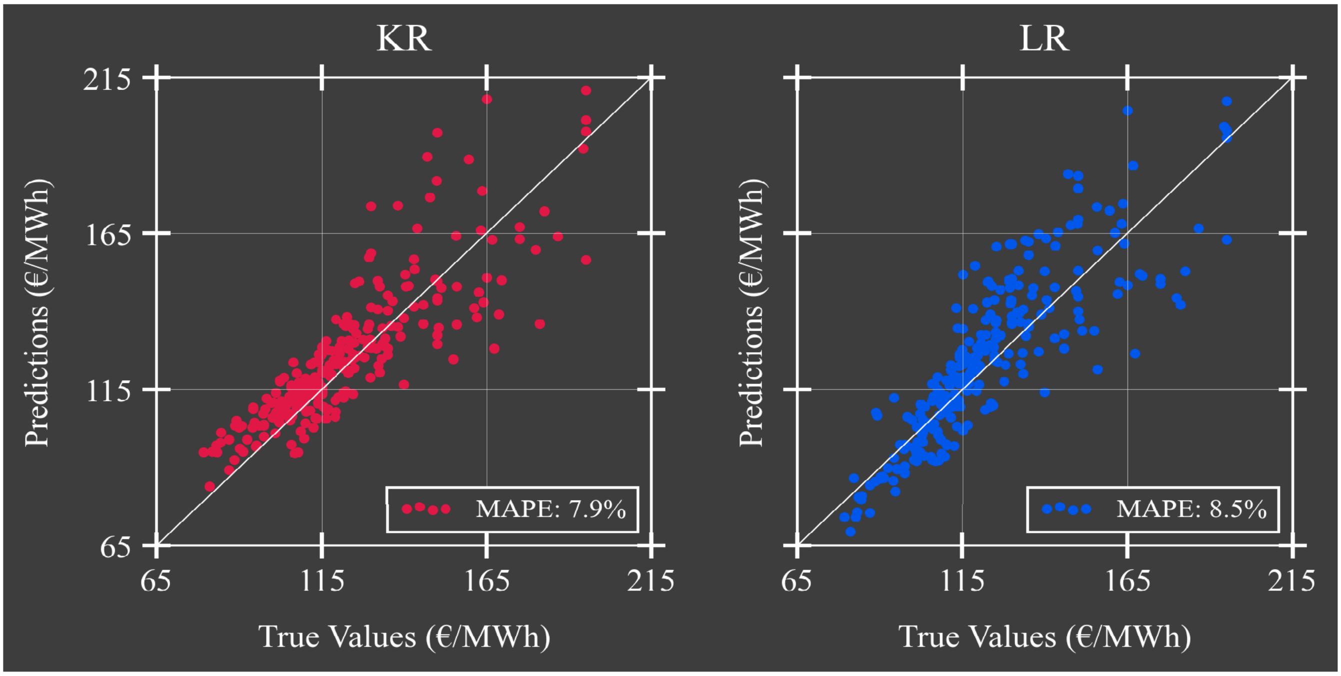

FIGURE 12. Multivariate regression results for september 2023 test period: peakload. The figure illustrates the predictive performance during the peakload hours of the test period in September 2023. Notably, the predictive accuracy of both models does not degrade significantly from baseload to peakload. This is particularly advantageous, as accurate predictions are crucial for managing the higher risks and volatility associated with peakload periods.

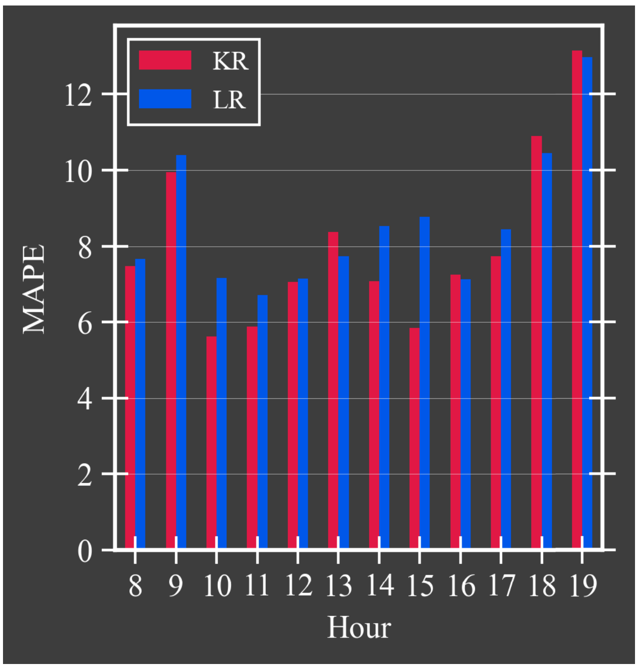

FIGURE 13. Point forecasting results. error by hour of day: peakload. The figure illustrates the predictive performance during the peakload hours of the test period in September 2023. The predictive accuracy of both models degrades in the late evening, when load oscillations are more frequent.

In this case, the data comes from Terna's website [2], the state-owned company serving as the transmission system operator for the Italian electric grid. The data, which has an hourly frequency, is updated and available daily. It is important to note that in this study, lacking generation forecasts, actual data from day d were used, as measured by Terna. Hence, we will denote the (exact) thermal generation forecast as T^d. This assumption seemed legitimate, as short-term thermal generation can be predicted with low errors, and several companies provide these values to utility companies. The same reasoning applies to photovoltaic generation, as explained below.

D. PHOTOVOLTAIC GENERATION FORECAST

This variable reflects the amount of electricity produced by photovoltaic installations throughout the country. As an NPRES, photovoltaic generation S can fluctuate throughout the day and is typically higher during periods of abundant sunlight and null at night. It has been included because in the Mediterranean area of Europe, photovoltaics is one of the most promising renewable generation systems. As such, significant investments have been and continue to be, contributing to a constant growth in its installed capacity.

The univariate analysis of the PUN with respect to this variable is presented in Fig. 7. When solar supply is abundant, electricity prices tend to be lower with reduced volatility. This behaviour is effectively captured by both KR, depicted in red, and LR, shown in blue. Notably, the results exhibit no significant differences between the two models.

The fact that photovoltaic generation is easily predictable in the short term justifies the assumption of including it as a perfect forecast in the regression model, as was for thermal generation. The data, which also in this case have hourly resolution, can be found at Terna's website [2].

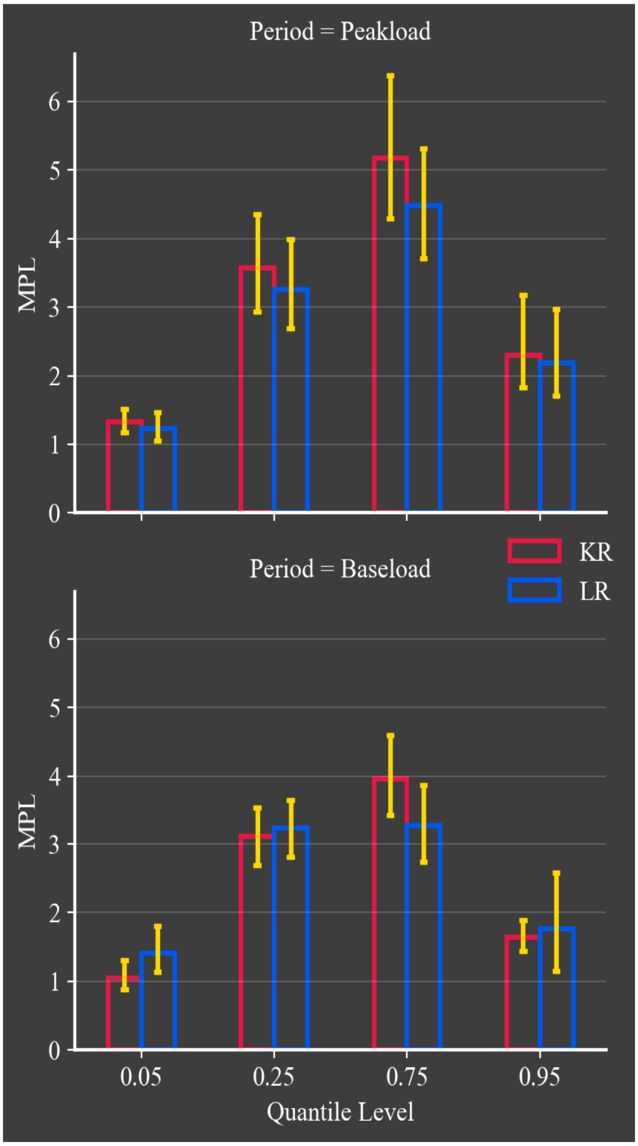

FIGURE 14. **Probabilistic forecasting performance for september 2023: MPL across quantiles for baseload and peakload periods. The bar plots depict Mean Pinball Loss (MPL), with results for Kernel Regression (KR) in red and Linear Regression (LR) in blue. Yellow lines above the bars indicate estimation errors. The quantiles posing greater predictive challenges are always found at the central levels (25th and 75th percentiles), with the 95th percentile generally being more elusive than the 5th percentile. This aligns with expectations, given the asymmetric distribution of electricity prices, where higher levels exhibit heightened volatility. Also in this case, linear and nonlinear models exhibit very similar performance.**

E. VARIABLES EXCLUDED FROM REGRESSION

Now let us briefly discuss some of the variables that were analyzed but excluded from the final regression model. The first variable to mention is the electrical load, which has already been discussed in Sec. II and is likely one of the most widely used exogenous variables in EPF models. Despite the widespread use of the load variable in existing literature, we chose to exclude it from our model. As illustrated in Fig. 8, no strong correlation is observed either with average prices or their distribution and the values of national demand. This observation holds for both KR (depicted on the left in red) and LR (on the right in blue). It is important to note that this result is based on the current conditions of the Italian market, and the correlation between price and load was more pronounced in previous years. In the face of significant changes in electricity prices, the Italian load has remained relatively stable over the past five years [6], with exceptions during lockdown periods of the COVID-19 pandemic. Given the fundamental importance of electricity in any developed country, where many activities rely on it, the demand for electrical energy remains essentially constant despite significant price fluctuations. Interestingly, this characteristic is commonly referred to by economists as the electricity market being "inelastic".

Another variable linked to the competitive market dynamics, particularly to supply, is international, or foreign, exchange (Forex), as shown in Fig. 9. The positive values of the Forex variable imply that importing energy is often more cost-effective. The examined data spanning from January to August 2023 shows no significant correlations with the energy price. Consequently, no model can derive benefits from this variable. The influence of the FOREX variable may be diminished due to the magnitude of the European energy crisis stemming from the Russian-Ukrainian conflict. This event has had significant repercussions not only on Italy but on almost all EU countries.

Furthermore, we mention the variable related to hydroelectric generation (Hydro), which exhibits a limited impact on the Italian market. The results are depicted in Figure 10. It is important to say that both hydroelectric generation and foreign exchange data are made available by Terna [2].

V. RESULTS

In electricity markets, it is common to divide the hourly periods of the week into two groups, peak and off-peak, based on an average load that is generally higher or lower. Therefore, the expressions peakload and baseload are used interchangeably, as well as simply "peak" and "base". In Italy, the peak period includes all hours from 8:00 AM to 8:00 PM, Monday to Friday. The remaining hours and the entire weekend constitute the off-peak period. Since many forward contracts are based on this division, and the day-ahead market serves to make load adjustments compared to what was obtained in advance through forwards, it is interesting to follow this division to assess the performance of models. Particularly in industrial EPF, the predictive performance on the peak is more interesting because higher prices generally form due to higher demand.

Following this distinction, we begin by presenting the results for point forecasting. The measure of prediction accuracy chosen is the Mean Absolute Percentage Error (MAPE). As done previously, values related to KR in the figures will be depicted in red, while those for LR models will be in blue. The predictive accuracy of both linear and non-linear models is presented for baseload in Fig. 11 and for

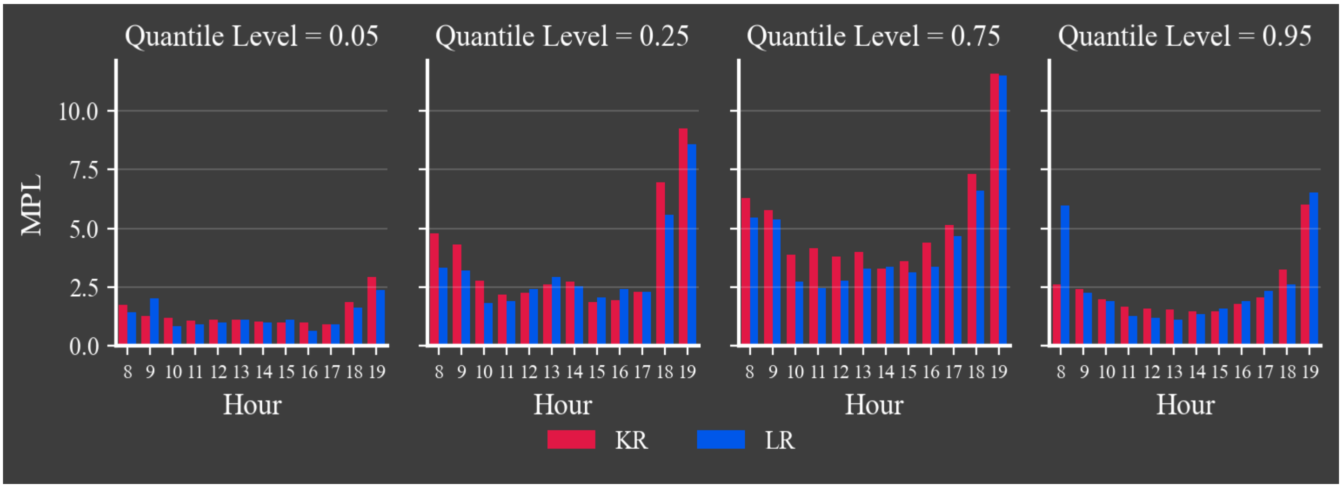

FIGURE 15. **Probabilistic forecasting performance for september 2023: MPL across quantiles and hour of day, for peakload periods The figure illustrates the predictive performance during the peakload hours of the test period in september 2023. Also in this case, the predictive accuracy of both models degrades in late evening hours, for all quantile levels.**

peakload in Fig. 12, covering all days from Monday to Friday ($d = 0, \ldots, 4$) and all hours of the day ($h = 0, \ldots, 23$). As an initial general observation, the models demonstrate very similar accuracy in both cases, around 7% for baseload and approximately 8% during peakload periods. This result can be considered satisfactory, as it aligns with the accuracy of forecasting models reported in the literature and tested on Italian market data from previous periods (pre-2020). For baseload, while KR produces unbiased estimates for each response variable value, LR tends to overestimate for high price levels and underestimate for lower levels. Still, from an industrial perspective, this is not a critical concern, as the final error metric does not vary much. For the peakload hours, the regression reveals a slightly lower accuracy for both models compared to baseload, as was expected. Importantly, almost no evident residuals are observed. This low difference in accuracy is particularly advantageous, given the critical role of precise predictions in managing the increased risks and volatility associated with peakload periods. It is worth noting that the estimation of MAPE may be slightly biased, especially as the variable average levels increase, causing the MAPE to decrease and vice versa. This consideration underscores why many authors opt to sacrifice the interpretability of MAPE for more rigorous error metrics.

Moving on to the more recent and interesting topic of probabilistic forecasts, we can start by looking at the Mean Pinball Loss (MPL) for each quantile level considered: $\alpha = 0.05, 0.25, 0.75, 0.95$. The results are shown in Fig. 14, segmented for baseload and peakload periods. Let us begin by noting that, in both cases, the quantiles that are more challenging to predict are those corresponding to the central levels (25th and 75th percentiles), while the highest quantile (95th percentile) is generally more difficult to predict than the lowest quantile (5th percentile). This latter observation was expected, as the distribution of electricity prices is asymmetric, with higher price levels exhibiting greater volatility. In this case, as well, the performance of KR and LR is nearly equivalent.

VI. CONCLUSION

In this study, we aimed to develop a robust EPF system for the contemporary Italian electricity market. In particular, the request for reliability and robustness is deeply linked to recent changes in international electricity markets. While these changes embrace the future with substantial shares of NPRES, they also frequently deal with challenges arising from geopolitical tensions. The literal unpredictability of prices observed during 2022 has significantly impacted utilities, especially considering that in the Italian context, they often have a mixed public-private capital structure. Consequently, both public interests and entrepreneurship have been affected.

The literature often presents CI models, or nonlinear models, as ideal solutions for forecasting problems, which, like that of EPF in the contemporary Italian market, lack similar cases in the past. Since employing data-driven techniques implies that only by trying these techniques on data, and conducting a heuristic evaluation, can their effectiveness be assessed, we tested the performance of KR versus LR, being the former one of the simplest approaches in the world of ML methods.

We attempted to make both methods interpretable by selecting variables based on fundamental reasoning and individually evaluating their suitability as predictors. In constructing multivariate models, we preferred including a relatively small number of variables, as we did not observe improvements in using models with more inputs.

The models underwent detailed testing for both point and probabilistic forecasting, utilizing PIs with different assigned probabilities and various temporal aggregations. Notably, the results reveal that, despite the distinct nature of linear and non-linear models, their performance closely converges, highlighting the correctness of our approach to developing reliable EPF systems.

In essence, and to emphasize, this study aims to highlight how traditional techniques remain suitable for contemporary tasks, provided they are correctly applied and preferably avoid resorting to black-box methods. Approaches involving indiscriminate use of numerous variables, often characterized by complicated structures, are not reliable nor interpretable, contrasting the effectiveness of more conventional methods. It is also worth noting that KR, as was stated multiple times, is one of the simplest CI methods. A more in-depth study, which we may conduct in the near future, should include the evaluation of more advanced methods.

REFERENCES

[1] *Gestore Dei Mercati Energetici*. Accessed: Apr. 15, 2024. [Online]. Available: https://www.mercatoelettrico.org
[2] *Terna Download Center*. Accessed: Apr. 15, 2024. [Online]. Available: https://www.terna.it/en/electric-system/transparency-report/download-center
[3] (2023). *Piano Nazionale Integrato Per L'energia e il Clima (The Italian National Energy and Climate Plan)*. Accessed: May 15, 2024. [Online]. Available: https://commission.europa.eu/energy-climate-change-environment/implementation-eu-countries/energy-and-climate-governance-and-reporting/national-energy-and-climate-plans_en
[4] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. New York, NY, USA: Springer, 2007.
[5] A. Ciarreta, B. Martinez, and S. Nasirov, "Forecasting electricity prices using bid data," *Int. J. Forecasting*, vol. 39, no. 3, pp. 1253–1271, Jul. 2023.
[6] Energy & Strategy (Politecnico di Milano). (2023). *Electricity Market Report*. Accessed: May 15, 2024. [Online]. Available: https://www.terna.it/en/electric-system/transparency-report/download-center
[7] E. E. Elattar, S. K. Elsayed, and T. A. Farrag, "Hybrid local general regression neural network and harmony search algorithm for electricity price forecasting," *IEEE Access*, vol. 9, pp. 2044–2054, 2021.
[8] P. S. Georgilakis, "Market clearing price forecasting in deregulated electricity markets using adaptively trained neural networks," in *Advances in Artificial Intelligence*, G. Antoniou, G. Potamias, C. Spyropoulos, and D. Plexousakis, Eds. Berlin, Germany: Springer, 2006, pp. 56–66.
[9] A. Gianfreda, F. Ravazzolo, and L. Rossini, "Comparing the forecasting performances of linear models for electricity prices with high RES penetration," *Int. J. Forecasting*, vol. 36, no. 3, pp. 974–986, Jul. 2020.
[10] S. Golia, L. Grossi, and M. Pelagatti, "Machine learning models and intra-daily market information for the prediction of Italian electricity prices," *Forecasting*, vol. 5, no. 1, pp. 81–101, Dec. 2022.
[11] T. V. D. Heijden, J. Lago, P. Palensky, and E. Abraham, "Electricity price forecasting in European day ahead markets: A greedy consideration of market integration," *IEEE Access*, vol. 9, pp. 119954–119966, 2021.
[12] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access J. Power Energy*, vol. 7, pp. 376–388, 2020.
[13] C. M. Hurvich, J. S. Simonoff, and C.-L. Tsai, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 60, no. 2, pp. 271–293, Jul. 1998.
[14] M. H. Imani, E. Bompard, P. Colella, and T. Huang, "Forecasting electricity price in different time horizons: An application to the Italian electricity market," *IEEE Trans. Ind. Appl.*, vol. 57, no. 6, pp. 5726–5736, Nov. 2021.
[15] P. L. Joskow, "California's electricity crisis," Nat. Bureau Econ. Res., Cambridge, MA, USA, NBER Working Papers 8442, Aug. 2001.
[16] R. Koenker, "Quantile regression," in *Econometric Society Monographs*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[17] Q. Li and J. S. Racine, *Nonparametric Econometrics: Theory and Practice*, vol. 1. Princeton, NJ, USA: Princeton Univ. Press, 2006.
[18] X. Lu, J. Qiu, G. Lei, and J. Zhu, "An interval prediction method for day-ahead electricity price in wholesale market considering weather factors," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 2558–2569, Mar. 2024.
[19] K. Maciejowska and R. Weron, "Short- and mid-term forecasting of baseload electricity prices in the U.K.: The impact of intra-day price relationships and market fundamentals," Hugo Steinhaus Center, Wrocław Univ. Technol., Wrocław, Poland, HSC Res. Rep. HSC/15/04, Jan. 2015.
[20] P. Muniain and F. Ziel, "Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices," *Int. J. Forecasting*, vol. 36, no. 4, pp. 1193–1210, Oct. 2020.
[21] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1548–1568, Jan. 2018.
[22] G. Ruan, J. Wang, H. Zhong, Q. Xia, and C. Kang, "Improving sample efficiency of deep learning models in electricity market," *IEEE Trans. Power Syst.*, vol. 38, no. 5, pp. 4761–4773, Oct. 2023.
[23] R. Sgarlato and F. Ziel, "The role of weather predictions in electricity price forecasting beyond the day-ahead horizon," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2500–2511, May 2023.
[24] Charles J. Stone, "Consistent nonparametric regression," *Ann. Statist.*, vol. 5, no. 4, p. 595– 620, 1977.
[25] J. W. Taylor, "Evaluating quantile-bounded and expectile-bounded interval forecasts," *Int. J. Forecasting*, vol. 37, no. 2, pp. 800–811, Apr. 2021.
[26] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. Wrocław, Poland: Wrocław Univ. Technology, 2006.
[27] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *Int. J. Forecasting*, vol. 30, no. 4, pp. 1030–1081, Oct. 2014.
[28] C. Zhang, Y. Fu, and L. Gong, "Short-term electricity price forecast using frequency analysis and price spikes oversampling," *IEEE Trans. Power Syst.*, vol. 38, no. 5, pp. 4739–4751, Nov. 2023.

FRANCESCO MORAGLIO was born in Ceva, Italy. He received the Diploma degree (Hons.) from Liceo Scientifico Leonardo Cocito, Alba, in 2015, and the Bachelor of Science (B.S.) and Master of Science (M.S.) degrees in pure mathematics under the supervision of Dr. Stefano Barbero, with mentoring provided by Prof. Lea Terracini. He is currently pursuing the Ph.D. degree in engineering with the Department of Energy "Galileo Ferraris," Politecnico di Torino. During his university years, he was concurrently involved in the construction entrepreneurship sector. After completing the degree, he ventured into the energy and utilities industry. His research interests include renewable energies, machine learning, and contemporary history.

CARLO S. RAGUSA (Senior Member, IEEE) was born in Catania, Italy. He received the M.Sc. degree in electrical engineering from the University of Catania, Italy, in 1993, and the Ph.D. degree in electrical engineering from the Politecnico di Torino, Turin, Italy, in 1997. He has been a Faculty Member of the Politecnico di Torino, since 1998, where he is currently a Full Professor in electrical engineering with the Department of Energy Alileo Ferraris. His research interests include the experimental characterization and modeling of soft magnetic materials for electrotechnical applications and the analysis and optimization of electromagnetic devices. He has been the Chair of the IEC/TC 68 Magnetic Alloys and Steels, since 2020. He has been a member of the International Organizing Committee of the Soft Magnetic Materials Conference, since 2015, and the Chairperson of the Steering Committee of the International Workshop on 1&2 Dimensional Magnetic Measurement and Testing, since 2016.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement