

Review

Recent Advances in Probabilistic Electricity Price Forecasting: A Review of Methods and Evaluation Metrics

Grzegorz Dudek ^{1,2,3,*} , Paweł Piotrowski ⁴ , Marcin Kopyt ⁴ and Dariusz Baczyński ⁴ 

¹ Faculty of Electrical Engineering, Czestochowa University of Technology, Al. AK 17, 42-200 Czestochowa, Poland

² Faculty of Mathematics and Computer Science, University of Łódź, Banacha 22, 90-238 Łódź, Poland

³ CAMINO—Centre for Data Analysis, Modeling and Computational Sciences, University of Łódź, Narutowicza 68, 90-136 Łódź, Poland

⁴ Electrical Power Engineering Institute, Warsaw University of Technology, Koszykowa 75 Street, 00-662 Warsaw, Poland; pawel.piotrowski@pw.edu.pl (P.P.); marcin.kopyt@pw.edu.pl (M.K.); dariusz.baczynski@pw.edu.pl (D.B.)

* Correspondence: grzegorz.dudek@pcz.pl

Abstract

Electricity price forecasting has traditionally relied on point predictions, which provide a single expected value for each future delivery period. However, the high volatility, spikes, heavy tails, negative prices and regime changes observed in modern electricity markets make point forecasts insufficient for many trading, bidding, storage, scheduling and risk-management decisions. This review focuses on probabilistic electricity price forecasting (PEPF), which represents uncertainty through quantiles, prediction intervals, predictive densities and multivariate scenarios. It systematizes recent developments in post-processing methods, Quantile Regression Averaging, conformal prediction, Bayesian and heteroscedastic models, distributional neural networks, copula-based approaches, normalizing flows, generative models and scenario generation. Particular attention is paid to probabilistic forecast evaluation, including proper scoring rules, calibration diagnostics and statistical testing. The review highlights the transition from marginal uncertainty quantification toward coherent multivariate and decision-oriented forecasting, and identifies open challenges related to calibration, dependence modeling, benchmark design, economic value assessment and reproducibility.

Keywords: electricity price forecasting; probabilistic forecasting; prediction intervals; quantile regression; conformal prediction; normalizing flows; scenario generation; energy markets



Academic Editor: Mohsin Jamil

Received: 30 June 2026

Revised: 23 July 2026

Accepted: 26 July 2026

Published: 28 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Short-term electricity price forecasting is a central task in modern electricity markets. Day-ahead, intraday and balancing prices influence bidding strategies, generation scheduling, demand response, storage operation, portfolio allocation, risk management and market participation decisions. As the penetration of renewable energy sources increases and electricity systems become more dependent on weather-driven production, price uncertainty becomes more pronounced. Electricity prices may react sharply to unexpected changes in residual load, fuel prices, network constraints, market liquidity and strategic behavior of market participants. Consequently, electricity price time series often exhibit strong seasonality, nonlinear dynamics, volatility clustering, price spikes, negative-price episodes and heavy-tailed distributions.

A large part of the electricity price forecasting literature has focused on deterministic or point forecasts. Point forecasts are easy to interpret, directly usable in many engineering applications and naturally evaluated by metrics such as mean absolute error (MAE) or root mean square error (RMSE). Nevertheless, many real market decisions are not determined by the expected price alone. A forecast with a small average error may still be economically inadequate if it underestimates tail risk, fails to represent the probability of a price spike, or ignores dependence across delivery hours. For market participants, two situations with the same expected price may require different actions if one is associated with narrow uncertainty and the other with substantial upside or downside risk.

Probabilistic electricity price forecasting addresses this limitation by estimating the predictive distribution of future prices rather than only a single value. A probabilistic forecast may be represented by selected quantiles, prediction intervals, full cumulative distribution functions, density forecasts or scenario trajectories. This richer representation is particularly important in electricity markets, where uncertainty is structural rather than incidental. The development of probabilistic forecasting has therefore become one of the most important directions in electricity price forecasting, as emphasized by earlier reviews and methodological contributions [1,2].

The motivation for the present review is also related to a gap in the recent review literature. The last widely cited, dedicated and published review of probabilistic electricity price forecasting was provided by Nowotarski and Weron in 2018 [1]. Since then, PEPF has developed rapidly, while more recent reviews have usually focused on point forecasting, benchmarking, deep learning, or broader EPF trends, treating probabilistic forecasting only as one component of the wider field [2–4]. A very recent preprint by O'Connor et al. reviews probabilistic price forecasting across day-ahead, intra-day, and balancing markets [5], confirming the timeliness of the topic. In contrast, the present review places stronger emphasis on probabilistic forecast evaluation, including proper scoring rules, calibration diagnostics, interval and quantile evaluation, multivariate scores, statistical testing, economic value, and forecast-to-decision links.

The present article reviews recent advances in probabilistic electricity price forecasting, focusing on studies published between 2020 and May 2026. The literature search was aimed at identifying recent publications on probabilistic forecasting of electricity prices, with particular attention paid to papers published in high-impact journals in the fields of energy, forecasting, applied economics, operations research, and artificial intelligence. The main search criterion was the presence of the terms “probabilistic”, “forecasting”, “electricity”, and “prices” in the title, keywords, or main bibliographic description of the publication. In total, 66 candidate publications related to recent PEPF developments were identified and screened. Six papers were excluded because they did not satisfy the adopted relevance or quality criteria, for example because they focused primarily on point forecasting, were only indirectly related to electricity prices, lacked sufficient methodological detail, or did not provide a clear connection with short-term electricity markets. The final synthesis was therefore based on 60 recent PEPF publications, complemented by foundational review, methodological, and evaluation references where necessary. The largest groups of retained studies came from journals such as *Applied Energy*, *Energy Economics*, the *International Journal of Forecasting*, *Energies*, *IEEE Access*, *Energy and AI*, *Electric Power Systems Research*, the *Journal of Forecasting*, and *IEEE Transactions on Power Systems*, with additional contributions from related energy, forecasting, operations research, econometrics, and artificial intelligence outlets.

Google Scholar was used as the primary search tool because of its broad interdisciplinary coverage, including journal articles, conference papers, preprints, and early-access publications. Empirical comparisons also indicate that Google Scholar provides broader

coverage than more selective bibliographic databases such as Web of Science or Scopus. For this reason, it was considered suitable as a discovery-oriented search tool for a rapidly developing and interdisciplinary research field.

ChatGPT 5.5 was used only for clerical support in organizing bibliographic information and for suggesting provisional methodological labels. It was not used to retrieve authoritative bibliographic records, assess scientific quality, verify factual claims, or make inclusion or exclusion decisions. All candidate studies and all extracted information were checked by the authors against the original publications.

The selection procedure combined journal-level and article-level criteria. Journal impact and reputation were considered when prioritizing sources, but each publication was assessed individually in terms of its relevance to probabilistic electricity price forecasting, methodological contribution, and connection with short-term electricity markets. The review focuses primarily on day-ahead and intraday markets, where uncertainty is both operationally important and methodologically challenging.

The paper is organized as follows. Section 2 explains the motivation and decision value of probabilistic price forecasts. Section 3 discusses the main applications of PEPF in electricity markets. Section 4 systematizes probabilistic forecasting approaches. Section 5 reviews major families of probabilistic forecasting methods and models. Section 6 presents evaluation metrics, calibration diagnostics and statistical testing procedures. Section 7 discusses methodological trends, practical implications, evaluation challenges, and open research directions. Section 8 concludes the paper.

2. Motivation and Advantages of Probabilistic Electricity Price Forecasting

Electricity prices are inherently uncertain, and this uncertainty is structural rather than incidental. Unlike many other commodities, electricity is non-storable at scale in the usual market sense, demand and supply must be balanced continuously, and prices react sharply to unexpected changes in load, renewable generation, outages, fuel prices, weather conditions, and bidding behavior. These characteristics lead to volatility, spikes, asymmetry, heavy tails, and regime changes. Consequently, the conditional distribution of future prices is often at least as important as its conditional mean. Probabilistic electricity price forecasting (PEPF) has therefore emerged as a natural extension of deterministic forecasting, shifting the focus from single-point predictions to quantiles, intervals, densities, and scenario-based representations of uncertainty [1].

The methodological motivation for PEPF follows from the fact that many electricity-market decisions are nonlinear in forecast errors. A point forecast may be accurate on average and still be economically inadequate if it fails to represent tail risk, negative-price episodes, scarcity events, or asymmetric upside and downside risks. For this reason, recent studies increasingly model predictive distributions rather than only expected values. Examples include conformal prediction, regularized quantile regression averaging, distributional neural networks, isotonic distributional regression as a post-processing layer, Bayesian posterior predictive distributions for intraday markets, and scenario generation based on normalizing flows [6–11].

The main advantage of probabilistic forecasts is that they provide decision-relevant information about uncertainty. While deterministic forecasts may be sufficient for simple benchmarking tasks or deterministic optimization models, they do not indicate whether a forecast is associated with low or high uncertainty, whether extreme prices are likely, or whether risks are concentrated in one tail of the distribution. By contrast, quantiles, prediction intervals, predictive densities, and scenarios allow market participants to distinguish between situations with similar expected prices but substantially different risk

profiles [7]. This distinction is particularly important in electricity markets, where the economic consequences of underestimating a price spike can be much larger than those of a similarly sized overestimate.

Probabilistic forecasts are also more consistent with modern forecast evaluation principles. Point forecasts are typically assessed using metrics such as MAE or RMSE, which measure closeness to a single realized value. However, when forecasts are used for hedging, bidding, portfolio allocation, storage operation, or risk control, evaluation should account for both calibration and sharpness of the predictive distribution. Accordingly, recent PEPF studies rely on scoring rules and diagnostics such as the pinball loss, empirical coverage, interval scores, and the continuous ranked probability score (CRPS), which are specifically designed for probabilistic predictions [6]. This aligns the evaluation criterion more closely with the practical use of the forecast.

Another important motivation is tail-risk management. Electricity markets are exposed to extreme price events, negative prices, and volatility clustering, especially under high renewable penetration and changing market conditions. Probabilistic forecasts can be directly linked to Value-at-Risk, Conditional Value-at-Risk, and other portfolio risk measures, making them useful for risk-aware trading and operational decision-making. Recent studies show that probabilistic forecasts can improve portfolio performance and trading strategies when evaluated under risk-adjusted criteria [8,12].

Finally, PEPF provides a natural framework for representing dependence across hours, products, and markets. Electricity prices are not isolated hourly random variables: market participants often require a joint distribution across the 24 h of the day, or across day-ahead and intraday products, because optimization and trading decisions depend on the shape of the entire price path. Recent work has therefore moved beyond univariate prediction intervals toward multivariate distributions and full scenario generation. For example, ref. [13] converts point forecasts into multivariate probabilistic forecasts for day-ahead prices, while normalizing-flow-based methods generate coherent daily scenarios that preserve cross-hour dependence [11]. Deterministic forecasts, even when accurate hour by hour, do not by themselves provide such internally consistent joint structures.

3. Applications of Probabilistic Price Forecasts

Probabilistic electricity price forecasts are used most directly in trading and bidding [14]. Traders on day-ahead and intraday markets face decisions under uncertainty, and expected-price signals alone are not sufficient when profit depends on the entire distribution of possible market outcomes. Recent work shows that probabilistic forecasts can be embedded in market strategies that optimize reward–risk trade-offs, including Mean-CVaR formulations and trading rules based on prediction bands or forecast distributions [12]. The economic value of the forecast is therefore not limited to better statistical fit; it is realized through better trading decisions.

Another important application is short-term risk management. Utilities, retailers, and portfolio managers use probabilistic forecasts to evaluate risk–return trade-offs and optimize market participation decisions. For instance, ref. [15] develops dynamic strategies for choosing between electricity markets on the basis of probabilistic forecasts of profit and risk measures. The study explicitly uses day-ahead PEPF to estimate risk–return trade-offs in market allocation decisions. Similarly, ref. [16] applies probabilistic price forecasts to construct dynamic diversification strategies based on volatility-type risk measures for small electricity producers and buyers trading in wholesale electricity markets. These examples clearly illustrate why deterministic forecasts are insufficient: risk-aware allocation requires an estimated distribution of possible outcomes, not merely a mean forecast.

Probabilistic forecasting is also crucial in intraday market operation and continuous trading. Intraday prices are often even more volatile and information-sensitive than day-ahead prices, especially in systems with high renewable penetration. Recent papers on intraday markets argue that probabilistic approaches are needed precisely because they can represent large uncertainty and changing market conditions [10]. Bayesian posterior predictive distributions and multivariate path forecasts are especially useful here because they support rolling operational decisions as uncertainty resolves closer to delivery [17,18].

Finally, probabilistic forecasts serve as inputs to stochastic optimization models, including unit commitment, dispatch, storage management, and renewable participation strategies [19]. Probabilistic forecasting provides either predictive densities or scenario sets that can be passed into stochastic optimization models. The article on normalizing flows [11] illustrates this logic clearly: it generates full-day price scenarios conditioned on external factors, thereby creating decision inputs for downstream optimization under uncertainty. In this sense, PEPF is not only a forecasting task but also an enabling layer for optimization and operations research models. This connection between forecasting and decision-making is further emphasized by Gao and Dowling [20], who incorporate probabilistic forecasts into stochastic bidding and self-scheduling optimization models, demonstrating the direct value of probabilistic information in operational decision-making under uncertainty. Similarly, Maciejowska et al. show that hybrid Factor-QRA probabilistic forecasts can improve the performance of battery trading strategies, highlighting the economic benefits of accurate uncertainty quantification in energy market applications [21].

Nevertheless, the value of probabilistic forecasts depends on the decision context. Beykirch et al. theoretically analyze the role of probabilistic forecasts in electricity market bidding and scheduling under uncertainty and show that not all operational decision problems require full multivariate predictive distributions [22]. In many applications, marginal predictive distributions or even expected values may be sufficient, whereas storage optimization, portfolio management, and bidding strategies typically require full joint probabilistic representations. Their results highlight that the appropriate level of probabilistic sophistication should be aligned with the requirements of the downstream optimization problem rather than assumed a priori.

4. Probabilistic Forecasting Approaches

Following the classification introduced in [23], probabilistic forecasting approaches can be broadly grouped into three main categories:

- (i) input scenario simulation,
- (ii) post-processing of point forecasts, and
- (iii) models that generate probabilistic forecasts directly.

These architectural categories describe where uncertainty enters the forecasting pipeline and are not intended to form a mutually exclusive taxonomy of individual algorithms. The first class of methods constructs probabilistic forecasts indirectly by propagating uncertainty from input variables through deterministic models. In this framework, multiple scenarios of exogenous variables (e.g., load, renewable generation, or weather conditions) are generated and used as inputs to a forecasting model, resulting in an ensemble of point forecasts. The distribution of these outputs is then interpreted as a probabilistic forecast. In the context of electricity price forecasting, this approach is particularly relevant when probabilistic forecasts of fundamental drivers, such as wind or solar generation, are available. For example, ref. [11] proposes a scenario generation approach based on conditional normalizing flows, i.e., deep generative models built on invertible neural networks capable of representing complex, non-Gaussian distributions. The model learns the conditional distribution of day-ahead electricity prices given forecasts of exogenous

variables, such as load and renewable generation, and generates coherent price scenarios that reflect the underlying uncertainty.

The second class of methods generates probabilistic forecasts by post-processing existing point forecasts. Typically, a deterministic model is first used to produce point predictions, after which a statistical procedure is applied to characterize forecast uncertainty, either through the distribution of residuals or directly through predictive intervals and densities. A prominent example is conformal prediction, which provides distribution-free prediction intervals with formal coverage guarantees and has been successfully applied to both day-ahead and intraday electricity markets [7].

Another important example is Quantile Regression Averaging (QRA), a widely used post-processing approach introduced by Nowotarski and Weron [24] and subsequently extended by Uniejewski et al. [6]. QRA constructs probabilistic forecasts by combining multiple point forecasts into conditional quantile estimates, thereby transforming deterministic predictions into predictive distributions. More recent developments include distributional post-processing techniques such as Isotonic Distributional Regression (IDR), which improves the calibration and sharpness of predictive distributions derived from point forecasts [9].

The third category consists of models that generate probabilistic forecasts directly, without relying on intermediate point predictions. These models estimate the conditional distribution of electricity prices as a function of explanatory variables. Examples include distributional neural networks, which directly output parameters of the predictive distribution [8]. Bayesian models represent another important class, where predictive distributions are obtained from posterior distributions of model parameters. For instance, ref. [10] propose a Bayesian hierarchical model for PEPF of intraday electricity prices.

Recent advances also include deep learning architectures such as DeepAR, transformers, and normalizing flows, which can model complex, non-Gaussian and multivariate distributions. For example, ref. [11] use conditional normalizing flows to generate realistic joint scenarios of day-ahead electricity prices, preserving temporal dependencies across hours.

Independently of the construction mechanism, probabilistic forecasting methods can also be classified as parametric, semi-parametric, or non-parametric [23]. Parametric approaches assume a specific functional form for the predictive distribution and estimate its parameters. The simplest case involves Gaussian distributions characterized by mean and variance. However, the heavy-tailed and asymmetric nature of electricity prices often necessitates the use of more flexible distributions, such as the Student's *t* and Johnson's SU distributions. Parametric approaches are commonly combined with machine learning models, including neural networks that output distribution parameters, as demonstrated in recent work on distributional forecasting [8].

Non-parametric approaches, on the other hand, make minimal assumptions about the shape of the predictive distribution. These include kernel density estimation, empirical distributions derived from ensembles, and machine learning methods that directly approximate the distribution. A prominent example is quantile regression, which estimates conditional quantiles without specifying the full distribution. QRA and related methods have been widely used in electricity price forecasting due to their robustness and interpretability [6]. Conformal prediction methods also fall into this category, as they provide distribution-free prediction intervals with theoretical coverage guarantees [7]. Semi-parametric approaches combine the interpretability of parametric models with the flexibility of non-parametric components, for example by using parametric volatility models with non-parametric calibration or by combining point forecasts through quantile-regression post-processing.

Overall, the choice between parametric, semi-parametric, and non-parametric approaches depends on the trade-off between model flexibility, interpretability, and computational complexity. In electricity price forecasting, where data exhibit strong nonlinearities and extreme events, semi-parametric and non-parametric methods have gained increasing popularity due to their ability to capture complex distributional features without restrictive assumptions.

5. Probabilistic Forecasting Methods and Models

While Section 4 classified probabilistic forecasting approaches according to the architecture used to generate probabilistic information, this section reviews the main methodological families used to implement these approaches.

5.1. Classical Statistical Methods

Classical statistical approaches remain important in PEPF despite the growing popularity of machine learning and deep learning models. Their usefulness stems from interpretability, relatively low computational complexity and strong theoretical foundations.

A first important group consists of heteroscedastic and volatility-based models. Since electricity prices exhibit strong time-varying volatility, models that explicitly describe conditional variance are particularly useful. Bernardi and Lisi propose a non-parametric heteroscedastic additive regression framework for point and interval forecasting of zonal electricity prices and demand in the Italian market [25]. Similarly, ARX-GARCH-type models (autoregressive models with exogenous variables combined with generalized autoregressive conditional heteroscedasticity models) are used in [16] to jointly forecast day-ahead and intraday or balancing prices in Germany and Poland. The results show that modeling conditional heteroscedasticity improves probabilistic forecasts, especially when combined with variance-stabilizing transformations. Another ARX-based example is provided by Muniaín and Ziel, who develop a bivariate jump-diffusion framework for probabilistic forecasting of peak and off-peak electricity prices [26]. Their model captures dependence between peak and off-peak prices through correlated jump processes, making it suitable for modeling both regular price dynamics and extreme price movements.

Another important category is formed by Bayesian and posterior predictive models. Nickelsen and Müller develop a Bayesian hierarchical model for German intraday electricity prices using autoregressive structures, exogenous variables, Orthogonal Matching Pursuit for feature selection, and posterior predictive distributions [10]. This approach explicitly accounts for parameter uncertainty and is particularly useful in intraday markets, where uncertainty is high and market conditions change rapidly. Classical statistical methods are also used for multivariate dependence reconstruction and scenario generation. Grothe et al. apply the Schaake shuffle methodology with copula techniques to reconstruct realistic cross-hour dependence structures in day-ahead electricity price scenarios [13]. Their method first estimates univariate forecast distributions for individual hours and then uses historical dependence patterns to construct coherent multivariate forecasts.

A further group of studies focuses on trajectory and path forecasting for intraday markets. Narajewski and Ziel propose an ensemble simulation framework for realistic intraday electricity price trajectories using GAMLSS models (generalized additive model for location, scale and shape), LASSO regression, and heavy-tailed distributions [27]. In their framework, price differences are modeled with a zero-inflated distribution, represented as a mixture of a Dirac distribution and a Student's *t* distribution. This allows the model to capture both frequent small price changes and occasional extreme movements. Serafin et al. extend this idea toward trading applications by combining LASSO-based forecasting, quantile regression, Gaussian copulas, and simultaneous prediction bands [18].

Other statistical and semi-parametric methods also contribute to PEPF. Moraglio and Ragusa demonstrate the usefulness of kernel and linear regression methods for quantile forecasting in contemporary electricity markets [28]. Gao and Dowling incorporate probabilistic forecasts generated through autoregressive Gaussian process regression into stochastic bidding and self-scheduling optimization models [20]. Haben et al. develop a probabilistic forecasting framework based on the X-model, which estimates future supply and demand curves using linear regression and LASSO regularization [29]. Probabilistic forecasts are then obtained by bootstrapping residual time series, allowing the model to represent uncertainty around future market-clearing prices.

5.2. Conformal Prediction and Calibrated Probabilistic Forecasting

Conformal prediction methods have recently gained substantial attention because they provide prediction intervals with formal coverage guarantees under weak assumptions. Kath and Ziel adapt conformal prediction to day-ahead and intraday electricity price forecasting and demonstrate that conformal intervals can simultaneously achieve reliability and sharpness even under changing point forecast inputs, while remaining competitive with QRA approaches [7].

O'Connor et al. extend conformal prediction toward balancing and day-ahead electricity markets through Ensemble Batch Prediction Intervals (EnbPI) and Sequential Predictive Conformal Inference (SPCI) [30]. EnbPI leverages ensemble learning techniques and bootstrap sampling to provide reliable prediction intervals, whereas SPCI extends the EnbPI framework to non-exchangeable time-series data. In particular, SPCI incorporates quantile regression forests to estimate conditional quantiles of residuals and to better adapt to the temporal dependencies and heteroscedasticity commonly observed in electricity price time series.

Dutot et al. further develop this adaptive perspective for French electricity spot prices under highly non-stationary market conditions [31]. They propose OSSCP-horizon, an online sequential conformal strategy designed to improve interval reliability when the distribution of prices changes over time. Their results also show that combining conformalization with online aggregation of experts can substantially improve the robustness and sharpness of probabilistic forecasts during turbulent market periods.

Brusaferri et al. [32] combine conformal inference with online neural-network ensembles and adversarial recalibration procedures in order to improve calibration under distribution shifts. In their framework, prediction bands corresponding to different coverage levels are dynamically adjusted by tracking the associated quantiles of conformity-score sequences and incorporating the running sum of miscoverage events to support stable long-run calibration.

Lipiecki et al. additionally show that conformal prediction can be effectively combined with QRA and Isotonic Distributional Regression (IDR) post-processing frameworks [9]. Combining the IDR-generated predictive distributions with those obtained from the generally better-performing QRA and conformal prediction schemes significantly improves forecasting accuracy: the resulting ensemble framework outperforms state-of-the-art distributional deep neural networks.

5.3. Quantile Regression Averaging and Related Methods

One of the most established methodological streams is based on quantile regression for the aggregation of point forecasts. Quantile Regression Averaging (QRA), originally introduced by Nowotarski and Weron [24], remains one of the most influential post-processing methods for transforming point forecasts into predictive distributions [1]. In QRA, multiple point forecasts are combined through quantile regression to estimate conditional quantiles

of future electricity prices. A variant of QRA, the Quantile Regression Machine (QRM), also models the distribution of electricity spot prices using the same pool of point forecasts, but relies on a single averaged forecast for each time point [33].

Subsequent developments introduced regularized and smoothed variants of QRA. Uniejewski and Weron propose LASSO-regularized QRA (LQRA), which automatically selects informative predictors and improves predictive sharpness and trading performance relative to classical QRA [6]. Cornell et al. extend the QRA framework by incorporating not only point forecasts of electricity prices but also quantile estimates obtained from multiple probabilistic forecasting models [34]. Maciejowska et al. further extend the quantile aggregation paradigm through a hybrid Factor-QRA framework combining principal component analysis with QRA for probabilistic forecasting and battery trading optimization [21]. Uniejewski later introduces Smoothing QRA (SQRA), replacing classical quantile regression with smoothing quantile regression in order to improve calibration and economic performance during turbulent periods such as the COVID-19 pandemic and the European energy crisis [35]. Jiang et al. demonstrate that QRA may be adversely affected by nuisance variables and propose a nonconvex regularized QRA framework to improve probabilistic forecasting performance in high-dimensional settings [36]. Zakrzewski et al. additionally systematize the growing family of QRA-based methods through the ReModels package, which implements QRA, QRM, FQRA, LQRA, and SQRA models for PEPF [37].

Alternative distributional summaries have also been investigated. Janczura proposes Expectile Regression Averaging (ERA), which replaces quantile regression with expectile regression for the aggregation of point forecasts [38]. Since expectiles are closely related to coherent risk measures, the proposed framework is particularly suitable for risk management applications, including the estimation of Expectile Value-at-Risk (EVaR). Recent studies further extend quantile-regression frameworks toward non-crossing and outlier-adaptive probabilistic forecasting. Osone and Kodaira provide a recent example of direct quantile-regression-based PEPF for the U.K. electricity market, emphasizing model simplicity and the use of general predictors under practical bidding-deadline constraints [39]. Lipiecki et al. compare QRA, Conformal Prediction (CP), and Isotonic Distributional Regression (IDR) as post-processing techniques for transforming point forecasts into predictive distributions [9]. Chen et al. propose an outlier-adaptive non-crossing quantile regression framework that combines Local Outlier Factor filtering, Variational Mode Decomposition, recurrent neural networks, and conformal calibration mechanisms [40].

5.4. Probabilistic Forecast Combination and Ensemble Learning

Forecast combination and probabilistic ensemble learning remain central themes in the electricity price forecasting literature. Berrisch and Ziel propose the CRPS-learning framework and further extend it to multivariate probabilistic forecast aggregation [41]. The main innovation of their approach is the explicit consideration of dependencies between quantiles and marginal distributions through a smoothing procedure that enables online learning and adaptive updating of combination weights. Using the CRPS-learning framework, Nitka and Weron investigate whether combining predictive distributions through CRPS minimization improves economic performance in electricity market bidding [42]. Their results indicate that optimizing combination weights using CRPS-learning consistently improves probabilistic forecast accuracy, while trading profits generally follow improvements in forecasting performance.

Xu et al. propose a probabilistic forecasting framework based on quantile combination and multi-objective optimization [43]. In their approach, individual quantile forecasts are generated by recurrent neural networks and subsequently combined to obtain the final predictive density. The combination weights are optimized using the Tuna Swarm

Optimization (TSO) algorithm, allowing the framework to balance multiple forecasting objectives simultaneously and improve the quality of the resulting probabilistic forecasts.

5.5. Neural Networks and Deep Learning

Neural networks and deep learning methods have become increasingly important in probabilistic electricity price forecasting. These approaches extend conventional neural forecasting architectures with probabilistic output layers, quantile-based loss functions, density estimation procedures, distributional assumptions, copula structures, or generative components. Their main advantage is the ability to capture nonlinear relationships, temporal dependencies, and complex interactions between electricity prices and exogenous variables, while producing probabilistic outputs.

A major methodological advance is represented by distributional neural networks, which estimate the parameters of full predictive distributions directly rather than producing only point forecasts or selected quantiles. Marcjasz et al. propose distributional neural networks based on multilayer perceptrons with probability layers that estimate the parameters of normal and Johnson's SU distributions [8]. This framework is particularly relevant for electricity prices because Johnson's SU distribution can capture skewness and heavy tails. A similar distributional neural-network approach is used in [12], where the model outputs the parameters of Johnson's SU distributions and extends distributional forecasting toward multivariate electricity price forecasting and trading applications by modeling dependencies between different European electricity exchanges. Distributional neural networks are also used in imbalance price forecasting. In [44], imbalance price distributions are modeled using LASSO regression, GAMLSS, and probabilistic multilayer perceptrons. The best performance is obtained by averaging the cumulative distribution functions generated by GAMLSS and neural network models.

Another important group consists of neural interval forecasting models. These approaches generate prediction intervals directly, often by combining neural architectures with lower–upper bound estimation or quantile regression. Chaweewat and Singh propose a ResNet-based interval forecasting framework that combines lower–upper bound estimation, quantile regression, and mean–variance estimation [45]. Yang et al. develop a multi-step interval forecasting framework based on adaptive Variational Mode Decomposition, extreme learning machines (ELMs), and optimized lower–upper bound estimation [46]. Zhang and Fu also employ ELMs for prediction interval construction by reformulating the optimal prediction interval problem so that ELMs estimate the upper and lower bounds directly, while accounting for both reliability and sharpness requirements [47]. Ghimire et al. further extend this direction through a hybrid deep learning framework combining Variational Mode Decomposition, maximal overlap wavelet transforms, convolutional LSTM networks, and online sequential ELMs for point and interval forecasting [48]. Jiang et al. develop the L-NBeatsX model for prediction interval estimation by integrating N-BEATSx and LassoNet with residual connections and adaptive feature selection, thereby improving prediction stability and interpretability [49].

A further research stream combines neural networks with copula models and state-space structures. Wang et al. introduce PrEF, a copula-augmented neural state-space model that combines nonlinear neural state-space inference with copula-based dependence modeling [50]. Their framework explicitly captures uncertainty dependencies, multistep error accumulation, and interactions between electricity prices and external explanatory variables. Another example is provided in [51], where deep distributional time-series models based on echo state networks, copula processes, and Bayesian regularization are proposed for intraday electricity prices. In this study, an ensemble of echo state networks with different parameter settings is used to construct probabilistic forecasts.

Other neural architectures estimate predictive densities directly through mixture models or density-learning mechanisms. He et al. introduce an end-to-end probabilistic forecasting framework based on convolutional neural networks and label distribution learning forests that simultaneously generates point forecasts, prediction intervals, and probability density functions [52]. Brusaferrri et al. propose a parametric probabilistic forecasting framework for day-ahead electricity prices based on Gaussian mixture distributions, where the mixture parameters are generated by a recurrent neural network equipped with an L1-norm-based feature selection mechanism [53]. Similarly, Afrasiabi et al. develop a deep Gabor convolutional mixture network for probability density forecasting using a parametric mixture-distribution approach [54].

Recurrent and sequence-based deep learning models are also widely used in probabilistic electricity price forecasting. Trebbien et al. extend probabilistic deep learning toward volatility-aware forecasting using long short-term memory architectures (LSTM) that jointly predict electricity price levels and stochastic volatility dynamics under post-crisis market conditions [55]. In this framework, the model generates the parameters of a predictive distribution under a Gaussian assumption. Another LSTM-based probabilistic framework, focused on conditional quantile prediction, is proposed in [56]. The authors investigate spatio-temporal probabilistic forecasting for distribution locational marginal prices using a deep bidirectional LSTM architecture. To capture spatial dependencies, they train a single shared model in which each nodal price series is forecasted individually while simultaneously exploiting information from neighboring nodes. Liu et al. integrate SHAP-based feature selection (Shapley additive explanations), LSTNet quantile regression, and kernel density estimation to generate predictive densities and prediction intervals [57].

Normalizing flows and deep generative models represent one of the most advanced recent developments in PEPF. Cramer et al. propose a multivariate probabilistic intraday forecasting framework based on conditional normalizing flows capable of modeling nonlinear dependencies between intraday electricity products [58]. Normalizing flows are able to model complex probability distributions with strong inter-dimensional dependencies and high multimodality by transforming a simple base density through an invertible neural network under the change-of-variables formula. Hilger et al. further extend conditional normalizing flows toward coherent multivariate scenario generation for day-ahead electricity prices under changing market regimes [11]. The proposed approach generates full-day price scenarios conditioned on explanatory variables such as residual load forecasts, thereby preserving realistic temporal dependence structures across hours.

Generative adversarial networks (GANs) have also emerged as a promising approach for probabilistic scenario generation. Walter and Wagner develop a conditional generative adversarial network combined with convolutional neural networks and bidirectional LSTM layers (TSS-CGAN) for generating realistic day-ahead electricity price scenarios conditioned on load and renewable generation forecasts [59]. Lu et al. similarly introduce a Conditional Time-Series GAN (CTSGAN) framework for probabilistic electricity price scenario generation and forecasting interval estimation [60]. A key innovation of their approach is the incorporation of additional conditioning variables related to electricity prices, including electricity demand, weather conditions, temperature, and economic indicators, into the GAN architecture. This allows the model to generate scenarios that better reflect the influence of external market drivers. Chen et al. further propose a generative neural-network-based path forecasting framework for intraday electricity markets that directly learns multivariate trajectory distributions through scoring-rule optimization [17].

Finally, recent work increasingly studies global, multivariate, transformer-based, and market-microstructure-aware deep learning models. Mashlakov et al. evaluate global deep learning architectures such as DeepAR, DeepTCN, DSANet, and deep factor models for

multivariate probabilistic energy forecasting, including electricity price forecasting [61]. Yu et al. develop the OrderFusion framework, which models order-book microstructure directly through interaction-aware probabilistic forecasting [62]. OrderFusion is an end-to-end and parameter-efficient neural model that consistently outperforms multiple baselines, generalizes across markets with both high and low liquidity, and addresses the quantile-crossing problem. Zhang et al. propose a hybrid deep-learning-based probabilistic forecasting framework combining feature preprocessing, deep neural networks, error compensation, and quantile-regression-based uncertainty estimation [63]. Cantillo-Luna et al. develop a probabilistic transformer neural network architecture with Time2Vec embeddings for intraday electricity price forecasting [64]. Another transformer-based approach, the Temporal Fusion Transformer (TFT), is used in [65], where clustering outcomes from the LASSO model are incorporated by TFT to generate probabilistic forecasts. These studies show how deep neural networks can be combined with interpretable feature selection, temporal representation learning, and quantile-based uncertainty estimation.

5.6. General Methodological Trends

Overall, the reviewed literature demonstrates that PEPF has evolved from relatively simple residual-based interval estimation toward highly sophisticated probabilistic learning systems combining quantile regression, conformal calibration, Bayesian inference, deep neural networks, copula models, normalizing flows, generative adversarial networks, isotonic distributional regression, order-book representation learning, and multivariate scenario generation.

The dominant research trend is therefore toward increasingly flexible semi-parametric and non-parametric frameworks capable of representing heavy tails, spikes, asymmetry, volatility clustering, regime changes, and complex dependence structures while simultaneously providing coherent probabilistic information for downstream trading, bidding, optimization, and operational decision-making.

Probabilistic graphical models may represent a promising complementary direction in PEPF, although they have been less frequently explored in recent studies. Bayesian Networks and Dynamic Bayesian Networks can explicitly model conditional dependencies between electricity prices and market drivers such as demand, renewable generation, fuel prices, outages, network constraints, and calendar effects. Their main advantage is interpretability, since the dependence structure is directly encoded rather than hidden inside a black-box model. This makes them potentially useful for causal reasoning, structural scenario analysis, stress testing, and communication with market participants.

Dynamic Bayesian Networks are particularly relevant when relationships between market variables evolve over time. In a related direction, Markov regime-switching models can represent abrupt transitions between normal, volatile, and spike-prone price regimes, which is important in markets affected by negative prices, scarcity events, and sudden structural changes. Although these approaches are less prominent in recent PEPF research than quantile regression, conformal prediction, distributional neural networks, and generative models, they remain useful when interpretability, structural dependence modelling, and regime identification are important.

Table 1 summarizes the main methodological families reviewed in this section and synthesizes their forecast representations and principal findings. The table is intended to complement the narrative review by highlighting broad relative patterns across recent PEPF studies rather than providing a paper-by-paper inventory or a universal ranking of methods.

Table 1. Summary of the main methodological families reviewed in Section 5.

Methodological Family	Cited Studies	Forecast Representation	Main Findings: Strengths and Limitations
Classical statistical methods	[10,13,16,18,20,25–29]	Prediction intervals; posterior predictive distributions; parametric and semi-parametric predictive distributions; intraday price trajectories; multivariate scenarios; bootstrapped residual distributions	Classical statistical methods remain important because they are transparent, relatively parsimonious, computationally efficient, and easy to interpret. Heteroscedastic and ARX-GARCH-type models explicitly represent time-varying volatility, while Bayesian models account for parameter uncertainty and are useful in intraday markets. Schaake-shuffle, copula-based, GAMLSS, Gaussian-process, kernel-regression, and bootstrap-based approaches extend statistical modelling toward trajectories, multivariate dependence, and uncertainty around market-clearing prices. Their main limitation is lower flexibility than modern deep or generative models when price dynamics are highly nonlinear, strongly non-Gaussian, or structurally unstable.
Conformal prediction and calibrated probabilistic forecasting	[7,9,30–32]	Prediction intervals; conformalized intervals; calibrated predictive distributions obtained by post-processing	Conformal prediction is attractive because it provides prediction intervals with formal coverage guarantees under weak assumptions and can be applied as a post-processing layer to different underlying models. Adaptive and online conformal variants improve reliability under non-stationary market conditions and distribution shifts. Combining conformal prediction with QRA or IDR can improve probabilistic forecast quality. The main limitation is that conformal methods primarily control coverage and may require careful calibration-window selection; good coverage alone does not guarantee sharpness, tail accuracy, or realistic multivariate dependence. QRA and its extensions are among the most established post-processing approaches in PEPF. They transform point forecasts or other predictive inputs into conditional quantiles and remain strong, interpretable, and practically useful benchmarks. Regularized, smoothed, factor-based, nonconvex, and software-supported variants address predictor selection, nuisance variables, calibration, quantile crossing, and robustness. Expectile regression extends this family toward risk-measure-oriented forecasting. The main limitation is that QRA-type methods usually focus on marginal quantiles and depend on the quality and diversity of the input forecasts; without additional modelling, they do not fully represent cross-hour dependence or full scenario paths.
Quantile Regression Averaging and related methods	[6,9,21,33–40]	Conditional quantiles; quantile grids approximating predictive distributions; expectiles; non-crossing quantile forecasts	

Table 1. Cont.

Methodological Family	Cited Studies	Forecast Representation	Main Findings: Strengths and Limitations
Probabilistic forecast combination and ensemble learning	[41–43]	Combined predictive distributions; combined quantile forecasts; multivariate probabilistic forecasts	Forecast combination methods exploit diversity among component forecasts and can improve probabilistic accuracy. CRPS-learning provides a principled way to combine predictive distributions using a proper scoring rule, and its multivariate extension accounts for dependencies between quantiles and marginal distributions. Quantile-combination frameworks with multi-objective optimization can improve the quality of predictive densities. Their main advantage is adaptability and the ability to integrate several probabilistic models; their limitations include dependence on the availability and quality of component forecasts, calibration-window choices, optimization complexity, and the fact that better statistical scores do not always translate directly into better economic outcomes.
Neural networks and deep learning	[8,11,12,17, 44–65]	Parametric predictive distributions; prediction intervals; conditional quantiles; predictive densities; mixture distributions; copula-augmented distributions; multivariate scenarios; generated price paths; order-book-based probabilistic forecasts	Neural and deep learning methods substantially extend PEPF by modelling nonlinear relationships, temporal dependencies, high-dimensional inputs, skewness, heavy tails, cross-market or cross-hour dependence, and full scenario paths. Distributional neural networks estimate distribution parameters directly; neural interval and quantile models generate bounds or quantiles; mixture-density, copula-augmented, recurrent, LSTM, transformer, normalizing-flow, GAN, and order-book-aware architectures provide increasingly rich probabilistic representations. Their main advantages are flexibility and the ability to model complex nonlinear and multivariate structures. Their main limitations are higher computational cost, sensitivity to architecture and hyperparameter choices, reduced interpretability, risk of miscalibration, and greater reproducibility challenges compared with simpler statistical or post-processing methods.

6. Probabilistic Forecast Evaluation Metrics and Statistical Testing

Reliable evaluation is a central element of PEPF, because the purpose of a probabilistic forecast is not only to be close to the realized price, but also to quantify uncertainty in a calibrated and informative way. For this reason, the evaluation framework used for point forecasts, which is typically based on MAE or RMSE, is insufficient. Probabilistic forecasts should be assessed with scoring rules and diagnostic procedures that verify whether the predictive distribution is both statistically reliable and practically useful. This is particularly important in electricity markets, where prices are volatile, asymmetric, heavy-tailed and strongly dependent across delivery hours. Recent studies on PEPF use quantile scores, interval scores, the continuous ranked probability score (CRPS), multivariate scores and formal statistical tests to compare competing forecasting methods [6–9,11].

Let $y_{t,h}$ denote the observed electricity price for day t and delivery hour h , and let $F_{t,h}$ be the corresponding predictive cumulative distribution function (CDF). If the forecast is issued for the whole day, it is often more appropriate to evaluate the vector $\mathbf{y}_t = (y_{t,1}, \dots, y_{t,24})$ jointly rather than each hour separately, because many trading, storage and portfolio decisions depend on the full daily price path. Therefore, evaluation should distinguish between marginal forecast quality, which concerns the distribution of each individual hour, and multivariate forecast quality, which additionally concerns the dependence structure across hours or products [11,13,41].

6.1. Quantile-Based Evaluation

A common representation of probabilistic price forecasts is a set of conditional quantiles $q_{t,h}^{(\tau)}$, where $\tau \in (0, 1)$. Quantile forecasts are usually evaluated by the pinball loss, also called the quantile score. For a single observation, it is defined as

$$\text{PL}_\tau(q, y) = (y - q)(\tau - \mathbb{1}\{y < q\}), \quad (1)$$

where q is the predicted τ -quantile and $\mathbb{1}\{\cdot\}$ is the indicator function. Lower values indicate better performance. For day-ahead electricity prices, the score is usually averaged over all T days, delivery hours and evaluated probability levels:

$$\overline{\text{PL}} = \frac{1}{THK} \sum_{t=1}^T \sum_{h=1}^H \sum_{k=1}^K \text{PL}_{\tau_k}(q_{t,h}^{(\tau_k)}, y_{t,h}), \quad (2)$$

where $H = 24$ for hourly day-ahead markets and K is the number of evaluated quantiles.

The pinball loss is strictly consistent for quantiles, meaning that it is minimized, in expectation, by the true conditional quantile. It is therefore the natural criterion for comparing quantile-based PEPF models. A recent application of this score to the evaluation of regularized QRA is provided in [6].

Quantile forecasts should also be checked for calibration. A nominal τ -quantile should be exceeded by the realized price with probability approximately $1 - \tau$. Hence, empirical quantile coverage can be measured as

$$\hat{C}_\tau = \frac{1}{TH} \sum_{t=1}^T \sum_{h=1}^H \mathbb{1}\{y_{t,h} \leq q_{t,h}^{(\tau)}\}. \quad (3)$$

A well-calibrated model should satisfy $\hat{C}_\tau \approx \tau$. However, calibration alone is not sufficient: very wide predictive distributions may be well calibrated but not useful. Evaluation should therefore combine calibration with sharpness, i.e., the concentration of the predictive distribution around plausible outcomes.

6.2. Prediction Interval Scores

Prediction intervals are another common way to present probabilistic forecasts. Let $[L_{t,h}^{(\alpha)}, U_{t,h}^{(\alpha)}]$ denote a central prediction interval with nominal coverage $1 - \alpha$. Its empirical coverage, usually referred to as the prediction interval coverage probability (PICP), is defined as

$$\text{PICP}_\alpha = \frac{1}{TH} \sum_{t=1}^T \sum_{h=1}^H \mathbb{1}\{L_{t,h}^{(\alpha)} \leq y_{t,h} \leq U_{t,h}^{(\alpha)}\}. \quad (4)$$

When coverage deviations are reported explicitly, they can be summarized by the average coverage error (ACE) [10,36], defined as the difference, or absolute difference, between empirical coverage and nominal coverage, with values close to zero indicating better calibration.

A calibrated interval should have $\text{PICP}_\alpha \approx 1 - \alpha$. Nevertheless, interval coverage should not be interpreted separately from interval width. A model can achieve nominal coverage by producing excessively wide intervals, which is undesirable for market decisions. Therefore, coverage is often complemented by measures of interval sharpness. The prediction interval average width (PIAW) is

$$\text{PIAW}_\alpha = \frac{1}{TH} \sum_{t=1}^T \sum_{h=1}^H (U_{t,h}^{(\alpha)} - L_{t,h}^{(\alpha)}). \quad (5)$$

where lower values indicate sharper intervals.

PIAW and its normalized version, PINAW, reward narrow intervals regardless of whether the empirical coverage is adequate. Therefore, they are best interpreted jointly with PICP and, preferably, with a proper interval score [36,43,46].

A proper way to jointly evaluate the calibration and sharpness of a central prediction interval is the interval score, also known as the Winkler score [48]:

$$\text{IS}_\alpha(L, U; y) = (U - L) + \frac{2}{\alpha}(L - y)\mathbb{1}\{y < L\} + \frac{2}{\alpha}(y - U)\mathbb{1}\{y > U\}. \quad (6)$$

This score rewards narrow intervals but penalizes observations that fall outside the interval, with larger penalties for more severe violations. It is therefore consistent with the calibration-sharpness principle.

A complementary diagnostic measure is the accumulated width deviation (AWD) [46]:

$$\text{AWD}_\alpha(L, U; y) = \frac{L - y}{U - L}\mathbb{1}\{y < L\} + \frac{y - U}{U - L}\mathbb{1}\{y > U\}. \quad (7)$$

In electricity price forecasting, interval-based evaluation is particularly relevant for conformal prediction methods, because they are designed to construct intervals with finite-sample coverage properties under suitable assumptions [7]. In high-volatility periods, such as the European energy crisis, reporting both coverage and interval score is important because pointwise errors alone can hide substantial changes in uncertainty.

6.3. Distributional Scores

When a full predictive cumulative distribution $F_{t,h}$ is available, it should be evaluated directly using proper scoring rules. The continuous ranked probability score (CRPS) is one of the most widely used measures in probabilistic forecasting. For a predictive cumulative distribution function F and observation y , it can be written as

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(z) - \mathbb{1}\{y \leq z\})^2 dz. \quad (8)$$

An equivalent representation is

$$\text{CRPS}(F, y) = \mathbb{E}_F |X - y| - \frac{1}{2} \mathbb{E}_F |X - X'|, \quad (9)$$

where X and X' are independent random variables drawn from F .

For empirical applications where the forecast is represented by a finite number of Q equidistant quantiles, the CRPS can be efficiently approximated via the average pinball loss across those quantiles [8,9]:

$$\hat{\text{CRPS}}(F, y) = \frac{2}{|Q|} \sum_{\tau \in Q} \text{PL}_\tau(q, y). \quad (10)$$

The CRPS generalizes the absolute error to probabilistic forecasts and simultaneously accounts for calibration and sharpness. It is widely used in recent PEPF studies, including distributional neural networks, Bayesian probabilistic models and post-processing approaches [8–10]. Since lower CRPS values indicate better distributional accuracy, it is also a natural loss function for forecast combination and online learning. Berrisch and Ziel, for example, extend CRPS-based learning to multivariate probabilistic forecasts of day-ahead electricity prices [41].

6.4. Multivariate Evaluation and Scenario Forecasts

Many recent PEPF models generate full daily scenarios rather than independent hourly distributions. This is motivated by the fact that electricity-market decisions often depend on temporal dependence, spreads, ramps and aggregated costs over several hours. If a model produces scenarios $\mathbf{x}_1, \dots, \mathbf{x}_M \in \mathbb{R}^H$ for the full price vector $\mathbf{y}$, then marginal scores computed hour by hour do not fully assess whether the dependence structure is realistic. For this purpose, multivariate proper scores are required.

The energy score is a common multivariate extension of CRPS [11,13,58]:

$$ES(F, \mathbf{y}) = \mathbb{E}_F \|\mathbf{X} - \mathbf{y}\| - \frac{1}{2} \mathbb{E}_F \|\mathbf{X} - \mathbf{X}'\|, \quad (11)$$

where $\|\cdot\|$ is usually the Euclidean norm. For scenario forecasts, it can be estimated as

$$\hat{ES} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{x}_m - \mathbf{y}\| - \frac{1}{2M^2} \sum_{m=1}^M \sum_{j=1}^M \|\mathbf{x}_m - \mathbf{x}_j\|. \quad (12)$$

Another useful measure is the variogram score, which is more sensitive to dependence between components of the forecast vector [11,58]:

$$VS_p(F, \mathbf{y}) = \sum_{i=1}^H \sum_{j=1}^H w_{ij} \left(|y_i - y_j|^p - \mathbb{E}_F |X_i - X_j|^p \right)^2, \quad (13)$$

where $p > 0$ and w_{ij} are non-negative weights.

These multivariate scores are especially relevant for models such as the Schaake shuffle, normalizing flows and multivariate distributional learning, which aim to preserve cross-hour dependence and generate coherent daily scenarios [11,58]. In practice, it is advisable to report both marginal and multivariate scores, because a model may have accurate hourly distributions but poor dependence structure, or conversely realistic trajectories with less accurate marginal calibration.

6.5. Calibration Diagnostics

In addition to numerical scores, probabilistic forecasts should be evaluated graphically and diagnostically. If $F_{t,h}$ is a CDF, the probability integral transform (PIT) is defined as $u_{t,h} = F_{t,h}(y_{t,h})$. For a calibrated forecast, the PIT values should be approximately and uniformly distributed on $[0, 1]$. Deviations from uniformity indicate systematic underdispersion, overdispersion, skewness errors or biased forecasts. For quantile forecasts, analogous diagnostics are based on empirical exceedance frequencies, reliability diagrams and coverage plots. In electricity price forecasting, these diagnostics are important because average scores may hide systematic failures during specific hours, price regimes or high-renewable periods. A model may appear competitive on average but be poorly calibrated during evening ramps, weekends, negative-price episodes or scarcity-price events.

Formal backtesting procedures may also be used as a compact complement to PIT histograms and coverage plots. Marcjasz et al. [8] mention the Kupiec, Christoffersen and

Berkowitz tests: the Kupiec test checks unconditional coverage of interval violations, the Christoffersen test additionally examines whether violations are independent over time, and the Berkowitz test assesses full-distribution calibration using normally transformed PIT values. In PEPF, these tests help detect incorrect coverage, clustered exceedances and distributional miscalibration during volatile or extreme-price periods.

Calibration diagnostics should be interpreted together with sharpness measures. A forecast distribution that is very wide may pass coverage checks but provide little decision value. Conversely, a sharp but miscalibrated distribution may underestimate risk. This trade-off is especially relevant in recent post-processing studies, where the objective is to improve the calibration of point-forecast-based distributions without making them unnecessarily diffuse [9].

6.6. Statistical Testing of Probabilistic Forecasts

Forecast comparisons should not rely only on differences in average scores. Since electricity prices are serially dependent and forecasts are typically evaluated over multiple delivery hours, apparent improvements may be driven by sampling variability. Formal statistical tests are therefore needed to assess whether score differences are significant.

Let $S(F_{t,h}, y_{t,h})$ denote a proper scoring rule, such as the pinball loss or CRPS. For two competing forecasts $F^{(1)}$ and $F^{(2)}$, define the loss differential

$$d_{t,h} = S(F_{t,h}^{(1)}, y_{t,h}) - S(F_{t,h}^{(2)}, y_{t,h}). \quad (14)$$

If lower scores are better, negative values of $d_{t,h}$ favor model 1, while positive values favor model 2. A common testing problem is $H_0 : \mathbb{E}(d_{t,h}) = 0$, against the alternative that one model has lower expected loss.

The Diebold-Mariano test [66] can be applied to the sequence of loss differentials, usually after aggregating over hours or using a heteroscedasticity- and autocorrelation-consistent variance estimator:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/T}}, \quad (15)$$

where $\bar{d}$ is the average loss differential and $\hat{\sigma}_d^2$ is a long-run variance estimate.

In this setting, the interpretation of the test depends on the score used: a significant difference in pinball loss indicates different quantile accuracy, a significant difference in CRPS indicates different marginal distributional accuracy, and a significant difference in the energy or variogram score indicates different multivariate scenario quality [13,48].

A more general approach is the Giacomini-White conditional predictive ability test [67], which evaluates whether the expected loss differential is zero conditionally on an information set. If $\mathbf{z}_t$ is a vector of instruments, such as lagged prices, forecasted load, renewable generation or hour-specific indicators, the test is based on

$$\mathbf{g}_t = \mathbf{z}_t d_t, \quad (16)$$

and the null hypothesis is $H_0 : \mathbb{E}(\mathbf{g}_t) = \mathbf{0}$. The corresponding statistic is commonly written as

$$GW = T \bar{\mathbf{g}}' \hat{\Omega}^{-1} \bar{\mathbf{g}}, \quad (17)$$

where $\hat{\Omega}$ is a consistent covariance estimator.

Under standard conditions, the statistic is asymptotically chi-square distributed with degrees of freedom equal to the dimension of $\mathbf{z}_t$. This test is useful in electricity markets because a forecasting model may dominate only under specific conditions, for example during high wind generation, peak demand hours or volatile intraday trading periods.

When many models are compared simultaneously, pairwise testing can lead to multiple-testing problems. In such cases, researchers may use model confidence set procedures [68], bootstrap-based rankings or adjusted significance levels. This is particularly relevant in recent PEPF studies, where large pools of models, calibration windows, quantile grids or ensemble combinations are compared. For multivariate scenario forecasts, testing should be based on scores that reflect the intended use of the forecast. If the forecast is used for storage arbitrage, portfolio allocation or risk management, statistical scores should ideally be complemented by economic evaluation, because a statistically superior probabilistic forecast does not always imply higher decision value.

6.7. Metric Selection, Interpretation, and Limitations

The choice of evaluation metric should be aligned with the type of probabilistic forecast and with the intended decision context. Since no single score captures all aspects of probabilistic forecast quality, evaluation should combine complementary measures rather than rely on one metric only.

The following practical guidelines can be formulated:

- **Quantile forecasts** should primarily be evaluated using the pinball loss or the CRPS approximated from quantiles. The selected quantile grid should be reported explicitly, particularly for tail-risk assessment, because sparse grids may miss relevant distributional features, whereas denser grids provide more detailed information at the cost of greater evaluation complexity.
- **Prediction intervals** should be assessed using both coverage and sharpness information. Interval scores are preferable to reporting coverage or width alone because they jointly penalize excessive width and observations falling outside the interval.
- **Full marginal predictive distributions** are naturally evaluated using CRPS. It generalizes the absolute error to probabilistic forecasts and jointly reflects calibration and sharpness. However, CRPS is scale-dependent and does not assess cross-hour or cross-market dependence. Therefore, comparisons across markets with different price levels or volatility should be interpreted carefully or complemented by normalized or relative scores.
- **Multivariate scenarios** require multivariate scores, such as the energy score or variogram score. These metrics are more appropriate when the forecast is used for decisions depending on the full price path, but their results may depend on the chosen norm, order, or component weights.
- **Calibration diagnostics** should complement proper scoring rules because average scores may hide systematic miscalibration in specific hours, regimes, or tail events.
- **Statistical testing** should be used when comparing competing models, especially in studies involving many methods, markets, horizons, or quantile levels.
- **Decision-oriented evaluation** is necessary when forecasts are used for bidding, trading, storage operation, portfolio allocation, or risk management, because statistical improvements do not always translate into higher economic value.

Several additional metrics may be useful in more specific decision contexts. When full predictive densities are available, the log score, or logarithmic penalty, provides a strictly proper scoring rule that strongly penalizes forecasts assigning very low probability to events that actually occur. It is therefore particularly relevant for tail-sensitive applications, although it may be unstable when predictive densities are poorly estimated in the tails. For binary probabilistic events, such as whether the electricity price exceeds a critical threshold, the Brier Score is a natural evaluation metric. It is useful when market participants are interested not in the full predictive distribution, but in the probability of events such as price spikes, scarcity prices, or negative prices. When probabilistic forecasts are converted

into risk measures such as VaR or CVaR, evaluation may also include exceedance-based backtesting, VaR exceedance ratios, conditional coverage diagnostics, and asymmetric loss functions. These tools provide an additional link between statistical forecast evaluation and the economic or risk-management value of probabilistic forecasts.

7. Discussion

7.1. Main Methodological Trends

The reviewed literature shows that probabilistic electricity price forecasting has evolved from a niche extension of deterministic forecasting into a central research direction. This development is driven by increasing electricity-market volatility, growing renewable generation, the expansion of intraday trading, and the need for risk-aware decision-making. The main methodological shift is from marginal interval estimation toward full predictive distributions and multivariate scenario generation.

A key observation is that post-processing of point forecasts remains highly competitive. Methods such as QRA, conformal prediction, and isotonic distributional regression show that strong probabilistic forecasts can be obtained by calibrating and combining deterministic predictions [6,7,9]. This is practically important because many organizations already operate deterministic forecasting systems. Post-processing therefore offers a relatively low-cost path toward uncertainty-aware forecasting, especially when reliable point forecasts are available, interpretability is important, and the main goal is to improve calibration, interval coverage, or predictive distribution quality. These methods are also attractive in operational settings because they usually require lower computational effort than fully trained deep probabilistic models and can often be integrated into existing forecasting pipelines.

Forecast combination is an important part of this trend, but its effectiveness depends on the type of probabilistic output being combined. Quantile-based combination methods, such as QRA and its extensions, are most suitable when the objective is to combine point forecasts or quantile forecasts into calibrated marginal quantiles. CRPS-based learning is more appropriate when full predictive distributions are available and the combination can be optimized directly with respect to a proper scoring rule. Multivariate ensemble methods are needed when the main goal is not only marginal calibration, but also the preservation of dependence across hours, products, or markets. Thus, there is no single universally best ensemble strategy; the choice depends on the forecast representation, the diversity of component models, the calibration window, the market regime, and the intended decision use.

At the same time, direct probabilistic models are becoming increasingly sophisticated. Distributional neural networks, normalizing flows, GANs, and transformer-based architectures can represent nonlinear, non-Gaussian, and multivariate predictive distributions that are difficult to capture using classical methods [8,11,58,59]. These models are particularly useful when the forecasting problem requires flexible modelling of heavy tails, asymmetric distributions, cross-hour dependence, or coherent price scenarios. However, their practical value depends on whether improvements in forecast quality or scenario realism justify the additional costs of data preparation, model training, hyperparameter tuning, recalibration, and specialized computational resources.

7.2. Conditional Usefulness of Forecasting Approaches

The reviewed studies do not support a universal quantitative ranking of probabilistic forecasting methods, because they differ substantially in terms of markets, horizons, input variables, evaluation periods, benchmarks, calibration procedures, and scoring rules. Nevertheless, several qualitative conclusions can be drawn. Heteroscedastic, ARX-GARCH-type, and Bayesian models remain useful in markets characterized by volatility

clustering and time-varying uncertainty, especially when transparency is required. Adaptive conformal methods and robust post-processing are particularly relevant under regime changes, price spikes, and negative-price episodes, where static predictive intervals may become miscalibrated. In contrast, multivariate generative methods, including copulas, Schaake-shuffle-type approaches, normalizing flows, and GANs, are most appropriate when downstream decisions depend on the joint evolution of prices across hours, products, or markets.

Thus, method selection should depend not only on average statistical accuracy, but also on the market regime, forecast representation, importance of tail events, intended decision use, and operational constraints such as training time, inference time, update frequency, and available computational infrastructure. Recent comparative evidence also suggests that the way uncertainty is represented may be as important as the choice of the forecasting model itself. For example, Lebedev et al. show that statistical and deep learning models differ not only in predictive accuracy, but also in how they capture data uncertainty, model uncertainty, calibration uncertainty, and domain-shift uncertainty; in their study, simpler XGBoost- and LEAR-based models with appropriate uncertainty layers remain highly competitive with more complex deep learning approaches [69].

Another factor affecting performance is sensitivity to modelling and evaluation choices. Reported results may depend on the length of the rolling training window, the calibration window used for post-processing or conformal prediction, the number and placement of quantiles, the nominal coverage levels of prediction intervals, and the architectural parameters of neural or generative models. For example, QRA, conformal prediction, and isotonic distributional regression may perform differently depending on whether the calibration window reflects the current market regime. Similarly, quantile-based scores may be affected by the density of the evaluated quantile grid, especially in the tails, while distributional neural networks, normalizing flows, GANs, and transformers are sensitive to model depth, hidden dimensions, regularization, training procedures, and hyperparameter tuning. Future studies should therefore report these design choices transparently and, where possible, include sensitivity analyses with respect to window length, quantile grids, coverage levels, and key architectural parameters.

7.3. Practical Adoption and Industry Relevance

A further distinction should be made between methods that are close to operational practice and those that remain mainly academic or experimental. The actual use of PEPF methods in industry is difficult to assess precisely, because many forecasting systems used by utilities, traders, retailers, aggregators, and market-service providers are proprietary. Nevertheless, the reviewed literature allows several qualitative observations.

Post-processing approaches, including QRA, conformal prediction, isotonic distributional regression, and residual-based interval construction, appear particularly close to practical implementation. They can be added to existing deterministic forecasting systems, require relatively limited changes in operational workflows, and are easier to explain to decision-makers than many black-box generative models. Classical statistical, heteroscedastic, and Bayesian models are also attractive when transparency, maintainability, and robustness are important.

Methods explicitly linked to trading, bidding, battery operation, portfolio allocation, or stochastic optimization are also closer to real-world deployment because they translate probabilistic forecasts into operational decisions. In this context, practical relevance depends not only on statistical performance, but also on whether the forecast improves decision quality under realistic market constraints. By contrast, normalizing flows, GANs, transformer-based architectures, and other deep generative models are currently more

prominent in academic research and advanced experimental settings. They are promising for multivariate scenario generation and complex dependence modelling, but their industrial adoption may be limited by computational cost, training complexity, interpretability, reproducibility, and the need for specialized expertise.

Practical maturity should therefore be considered alongside methodological novelty. A less sophisticated but transparent, stable, and easily deployable method may be more useful in routine market operation than a highly complex model that offers better scenario realism but is costly to maintain or difficult to validate. In future industrial applications, federated learning may also become relevant as a privacy-preserving framework that allows market participants or institutions to jointly train probabilistic forecasting models without sharing commercially sensitive raw data. However, this direction remains largely prospective in PEPF and would require further research on calibration, robustness, communication costs, and validation under heterogeneous market conditions.

7.4. Multivariate Dependence and Scenario Realism

A major open challenge concerns multivariate dependence and scenario realism. Many market decisions depend on the joint behavior of prices across consecutive hours, products, or markets rather than on isolated marginal distributions. Probabilistic forecasts should therefore preserve cross-hour dependence, price spikes, negative-price episodes, ramping patterns, and temporal clustering of volatility. Normalizing flows, copula-based models, Schaafe-shuffle-type approaches, and multivariate scoring-rule-based learning provide promising tools in this direction [11,13,41,58].

Nevertheless, scenario realism remains difficult to evaluate. A set of scenarios may achieve a good average score while still producing unrealistic temporal paths or failing to reproduce extreme market states. Future research should therefore combine multivariate scores with diagnostics focused on path coherence, tail dependence, spike timing, and the reproduction of market-relevant temporal structures.

7.5. Evaluation, Statistical Testing, and Economic Value

Probabilistic forecasts cannot be assessed by point-error metrics alone. Proper scoring rules, calibration diagnostics, and multivariate scores are needed to evaluate predictive distributions. However, there is no single universal metric. Pinball loss is appropriate for quantiles, interval scores for prediction intervals, CRPS for marginal distributions, and energy or variogram scores for scenarios. The choice of metric should therefore be aligned with the forecast representation and the intended decision use. Formal statistical testing of score differences should also become more common, especially when many competing models are compared. In addition, empirical studies should report computational requirements, because a statistically superior model may be less useful in practice if it is too costly to train, update, or deploy in real time.

The relationship between statistical accuracy and economic value requires particular attention. Probabilistic forecasts do not influence profits or operational risk directly; they first have to be transformed into market or operational decisions. This transformation depends on the application. Quantiles may be used to construct prediction-band trading rules, risk limits, or asymmetric bidding rules. Predictive densities may be converted into Value-at-Risk, Conditional Value-at-Risk, expected utility, or other risk-adjusted criteria. Multivariate scenarios may serve as inputs to stochastic programming, storage optimization, bidding models, or portfolio-allocation frameworks.

As a result, an improvement in CRPS, pinball loss, or interval score does not automatically imply higher trading profit or lower operational risk. Hirsch and Ziel reinforce this point by showing that popular quantile-based battery trading strategies may produce

misleading forecast rankings, because they are not proper scoring rules and may ignore the intertemporal dependence structure of electricity prices [70]. They argue that economic backtesting should be interpreted as an additional layer of decision-quality assessment rather than as a substitute for proper probabilistic forecast evaluation. Conversely, a forecast with a slightly worse average statistical score may be more valuable if it better represents tail events, temporal dependence, or downside risk. The final economic outcome depends not only on the quality of the predictive distribution, but also on the downstream decision model, transaction costs, market constraints, risk preferences, timing of market participation, and computational feasibility. Future studies should therefore combine statistical evaluation with decision-oriented assessment and report the full forecast-to-decision chain, including how predictive distributions are converted into trading, bidding, storage, or portfolio decisions and how the resulting profits, losses, and risk-adjusted performance are evaluated.

7.6. Calibration Under Regime Changes and Intraday Microstructure

Calibration under regime changes remains a key challenge. Electricity markets can change rapidly because of fuel-price shocks, policy interventions, weather extremes, geopolitical events, changes in market design, or structural shifts in renewable penetration. Probabilistic forecasts must therefore remain calibrated not only during stable historical periods, but also under changing market conditions. Online recalibration, adaptive conformal methods, rolling estimation, regime-aware models, and robust post-processing are important directions for future work [9,32]. This issue is closely related to real-time distribution drift, where the underlying price distribution changes as new information, liquidity conditions, or market shocks appear. This issue is especially relevant in markets characterized by negative prices, scarcity events, and rapid changes in intraday liquidity.

Probabilistic forecasting in intraday order-book and market-microstructure settings is another underdeveloped area. Most PEPF studies still focus on day-ahead prices or hourly intraday prices, whereas modern intraday markets are increasingly continuous, high-frequency, and liquidity-dependent. In such settings, uncertainty is shaped not only by fundamentals such as load, renewable generation, and fuel prices, but also by order-book depth, bid-ask spreads, trading intensity, liquidity shocks, and strategic bidding behavior. Probabilistic models for these environments should account for high-frequency information, irregularly spaced transactions, changing liquidity, and the joint distribution of prices and volumes. Computational efficiency is particularly important here, because forecasts may need to be recomputed many times within a trading day as new orders, trades, and liquidity signals arrive. Future research should therefore combine probabilistic forecasting with drift-detection mechanisms and fast recalibration procedures suitable for real-time market operation.

7.7. Interpretability, Reproducibility, and Reporting Standards

Interpretability is becoming increasingly important as deep generative models enter PEPF. GANs, normalizing flows, and related architectures can generate complex predictive distributions and realistic-looking scenarios, but their internal mechanisms are often difficult to understand. This lack of transparency may reduce trust and hinder adoption, especially when forecasts are used in high-stakes bidding, hedging, or risk-management decisions. Interpretability should therefore concern not only expected prices, but also probabilistic outputs, including quantiles, prediction intervals, tail probabilities, distribution parameters, and generated scenario paths. Future work should combine probabilistic deep learning with explainability tools, sensitivity analysis, feature-attribution methods, and

diagnostics that clarify why a model assigns high probability to certain price paths or extreme events.

Reproducibility also remains a persistent limitation. Many studies use different markets, forecasting horizons, training and calibration windows, input variables, benchmarks, quantile grids, and evaluation metrics, which limits direct comparison between methods. The field would benefit from open benchmark datasets, standardized probabilistic evaluation protocols, transparent reporting of input data, window choices, quantile levels, hyperparameters, and formal statistical testing of score differences. Open code and publicly available experimental pipelines are particularly important because many recent PEPF methods involve complex architectures, nontrivial preprocessing, extensive hyperparameter tuning, and sensitivity to calibration design.

Reproducibility should also include transparent reporting of computational resources, training and inference times, hardware used, model size, tuning procedures, recalibration frequency, and sensitivity analyses. Such reporting would make it easier to distinguish genuine methodological progress from gains caused by dataset choice, evaluation design, implementation details, window selection, quantile-grid choice, or computationally intensive tuning.

8. Conclusions

Probabilistic electricity price forecasting has become an essential component of modern electricity-market analysis because it represents not only expected prices but also uncertainty, tail risk, and temporal dependence. This is particularly important in markets affected by renewable generation, price spikes, negative prices, changing liquidity, and frequent structural shifts.

The reviewed literature shows a clear methodological transition from marginal quantiles and prediction intervals toward full predictive distributions and multivariate scenarios. This development increases the relevance of PEPF for bidding, trading, hedging, storage operation, portfolio allocation, and risk management. At the same time, it requires evaluation frameworks that combine proper scoring rules, calibration diagnostics, multivariate assessment, and measures of economic value.

Future research should give greater attention to robust calibration under regime changes, realistic scenario generation, cross-hour and cross-market dependence, and probabilistic forecasting in intraday order-book and market-microstructure settings. Further progress also depends on transparent benchmark datasets, open code, reproducible evaluation protocols, and more interpretable deep generative models. As electricity markets become increasingly uncertain, probabilistic forecasting is likely to become a necessary element of market operation and risk-aware decision-making.

Author Contributions: Conceptualization, G.D., D.B., M.K. and P.P.; methodology, G.D., D.B., M.K. and P.P.; validation, G.D., D.B., M.K. and P.P.; formal analysis, G.D., D.B., M.K. and P.P.; investigation, G.D., D.B., M.K. and P.P.; resources, G.D., D.B., M.K. and P.P.; data curation, G.D., M.K. and P.P.; writing—original draft preparation, G.D., D.B., M.K. and P.P.; writing—review and editing, G.D., D.B., M.K. and P.P.; supervision, G.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors acknowledge the use of Gemini 3.1 Pro (Google) and ChatGPT 5.5 (OpenAI) as language-editing tools to support the translation of the manuscript from Polish into English and to improve its clarity, readability, and grammatical accuracy. ChatGPT was also used as an auxiliary tool to support the identification, organization, and preliminary screening of relevant publications during the literature review process. All scientific conclusions, interpretations, publication selections, and findings remain the sole responsibility of the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Nowotarski, J.; Weron, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1548–1568. [\[CrossRef\]](#)
- Maciejowska, K.; Uniejewski, B.; Weron, R. Forecasting Electricity Prices. In *Oxford Research Encyclopedia of Economics and Finance*; Banerjee, A., Ed.; Oxford University Press: Oxford, UK, 2022. [\[CrossRef\]](#)
- Lago, J.; Marcjasz, G.; De Schutter, B.; Weron, R. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Appl. Energy* **2021**, *293*, 116983. [\[CrossRef\]](#)
- O'Connor, C.; Bahloul, M.; Prestwich, S.; Visentin, A. A Review of Electricity Price Forecasting Models in the Day-Ahead, Intra-Day, and Balancing Markets. *Energies* **2025**, *18*, 3097. [\[CrossRef\]](#)
- O'Connor, C.; Bahloul, M.; Prestwich, S.; Visentin, A. The Evolution of Probabilistic Price Forecasting Techniques: A Review of the Day-Ahead, Intra-Day, and Balancing Markets. *arXiv* **2025**, arXiv:2511.05523.
- Uniejewski, B.; Weron, R. Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Econ.* **2021**, *95*, 105121. [\[CrossRef\]](#)
- Kath, C.; Ziel, F. Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *Int. J. Forecast.* **2021**, *37*, 777–799. [\[CrossRef\]](#)
- Marcjasz, G.; Narajewski, M.; Weron, R.; Ziel, F. Distributional neural networks for electricity price forecasting. *Energy Econ.* **2023**, *125*, 106843. [\[CrossRef\]](#)
- Lipiecki, A.; Uniejewski, B.; Weron, R. Postprocessing of point predictions for probabilistic forecasting of day-ahead electricity prices: The benefits of using isotonic distributional regression. *Energy Econ.* **2024**, *139*, 107934. [\[CrossRef\]](#)
- Nickelsen, D.; Müller, G. Bayesian hierarchical probabilistic forecasting of intraday electricity prices. *Appl. Energy* **2025**, *380*, 124975. [\[CrossRef\]](#)
- Hilger, H.; Witthaut, D.; Dahmen, M.; Gorjao, L.R.; Trebbien, J.; Cramer, E. Multivariate scenario generation of day-ahead electricity prices using normalizing flows. *Appl. Energy* **2024**, *367*, 123241. [\[CrossRef\]](#)
- Agakishiev, I.; Härdle, W.K.; Kopa, M.; Kozmik, K.; Petukhina, A. Multivariate probabilistic forecasting of electricity prices with trading applications. *Energy Econ.* **2025**, *141*, 108008. [\[CrossRef\]](#)
- Grothe, O.; Kächele, F.; Krüger, F. From point forecasts to multivariate probabilistic forecasts: The Schaafe shuffle for day-ahead electricity price forecasting. *Energy Econ.* **2023**, *120*, 106602. [\[CrossRef\]](#)
- Beykirch, M.; Janke, T.; Steinke, F. Bidding and Scheduling in Energy Markets: Which Probabilistic Forecast Do We Need? In *Proceedings of the 2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*; IEEE: New York, NY, USA, 2022; pp. 1–6. [\[CrossRef\]](#)
- Janczura, J.; Wójcik, E. Dynamic short-term risk management strategies for the choice of electricity market based on probabilistic forecasts of profit and risk measures. The German and the Polish market case study. *Energy Econ.* **2022**, *110*, 106015. [\[CrossRef\]](#)
- Janczura, J.; Puć, A. ARX-GARCH Probabilistic Price Forecasts for Diversification of Trade in Electricity Markets—Variance Stabilizing Transformation and Financial Risk-Minimizing Portfolio Allocation. *Energies* **2023**, *16*, 807. [\[CrossRef\]](#)
- Chen, J.; Lerch, S.; Schienle, M.; Serafin, T.; Weron, R. Probabilistic intraday electricity price forecasting using generative machine learning. *arXiv* **2025**, arXiv:2506.00044. [\[CrossRef\]](#)
- Serafin, T.; Marcjasz, G.; Weron, R. Trading on short-term path forecasts of intraday electricity prices. *Energy Econ.* **2022**, *112*, 106125. [\[CrossRef\]](#)
- Cui, H.; Qin, H.; Zhou, J.; Wan, C.; Hu, Z.; Pan, Y. Nonparametric probabilistic forecasting-driven electricity market clearing via improved two-stage interval optimization. *CSEE J. Power Energy Syst.* **2025**, 1–13. [\[CrossRef\]](#)
- Gao, R.; Dowling, A.W. Making Money in Energy Markets: Probabilistic Forecasting and Stochastic Programming Paradigms. In *2020 American Control Conference (ACC)*; IEEE: New York, NY, USA, 2020; pp. 168–173. [\[CrossRef\]](#)
- Maciejowska, K.; Serafin, T.; Uniejewski, B. Probabilistic forecasting with a hybrid Factor-QRA approach: Application to electricity trading. *Electr. Power Syst. Res.* **2024**, *234*, 110541. [\[CrossRef\]](#)
- Beykirch, M.; Bott, A.; Janke, T.; Steinke, F. The Value of Probabilistic Forecasts for Electricity Market Bidding and Scheduling Under Uncertainty. *IEEE Trans. Power Syst.* **2024**, *39*, 6986–6997. [\[CrossRef\]](#)

23. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [\[CrossRef\]](#)
24. Nowotarski, J.; Weron, R. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Comput. Stat.* **2015**, *30*, 791–803. [\[CrossRef\]](#)
25. Bernardi, M.; Lisi, F. Point and Interval Forecasting of Zonal Electricity Prices and Demand Using Heteroscedastic Models: The IPEX Case. *Energies* **2020**, *13*, 6191. [\[CrossRef\]](#)
26. Muniain, P.; Ziel, F. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. *Int. J. Forecast.* **2020**, *36*, 1193–1210. [\[CrossRef\]](#)
27. Narajewski, M.; Ziel, F. Ensemble forecasting for intraday electricity prices: Simulating trajectories. *Appl. Energy* **2020**, *279*, 115801. [\[CrossRef\]](#)
28. Moraglio, F.; Ragusa, C.S. Day-Ahead Electricity Price Forecasting in the Contemporary Italian Market. *IEEE Access* **2024**, *12*, 74606–74620. [\[CrossRef\]](#)
29. Haben, S.; Caudron, J.; Verma, J. Probabilistic Day-Ahead Wholesale Price Forecast: A Case Study in Great Britain. *Forecasting* **2021**, *3*, 596–632. [\[CrossRef\]](#)
30. O'Connor, C.; Bahloul, M.; Rossi, R.; Prestwich, S.; Visentin, A. Conformal Prediction for electricity price forecasting in the day-ahead and real-time balancing market. *Energy AI* **2025**, *21*, 100571. [\[CrossRef\]](#)
31. Dutot, G.; Zaffran, M.; Féron, O.; Goude, Y. Adaptive probabilistic forecasting of French electricity spot prices. *arXiv* **2024**, arXiv:2405.15359.
32. Brusaferrri, A.; Ballarino, A.; Grossi, L.; Laurini, F. On-line conformalized neural networks ensembles for probabilistic forecasting of day-ahead electricity prices. *Appl. Energy* **2025**, *398*, 126412. [\[CrossRef\]](#)
33. Marcjasz, G.; Uniejewski, B.; Weron, R. Probabilistic electricity price forecasting with NARX networks: Combine point or probabilistic forecasts? *Int. J. Forecast.* **2020**, *36*, 466–479. [\[CrossRef\]](#)
34. Cornell, C.; Dinh, N.T.; Pourmousavi, S.A. A probabilistic forecast methodology for volatile electricity prices in the Australian National Electricity Market. *Int. J. Forecast.* **2024**, *40*, 1421–1437. [\[CrossRef\]](#)
35. Uniejewski, B. Smoothing quantile regression averaging: A new approach to probabilistic forecasting of electricity prices. *J. Commod. Mark.* **2025**, *39*, 100501. [\[CrossRef\]](#)
36. Jiang, H.; Dong, Y.; Wang, J. Electricity price forecasting using quantile regression averaging with nonconvex regularization. *J. Forecast.* **2024**, *43*, 1859–1879. [\[CrossRef\]](#)
37. Zakrzewski, M.; Skonieczka, K.; Małkiński, M.; Mańdziuk, J. ReModels: Quantile Regression Averaging models. *SoftwareX* **2024**, *28*, 101905. [\[CrossRef\]](#)
38. Janczura, J. Expectile regression averaging method for probabilistic electricity price forecasting. *Comput. Stat.* **2025**, *40*, 681–705. [\[CrossRef\]](#)
39. Osone, Y.; Kodaira, D. Quantile Regression for Probabilistic Electricity Price Forecasting in the U.K. Electricity Market. *IEEE Access* **2025**, *13*, 10083–10093. [\[CrossRef\]](#)
40. Chen, Z.; Zhang, B.; Du, C.; Yang, C.; Gui, W. Outlier-adaptive-based non-crossing quantiles method for day-ahead electricity price forecasting. *Appl. Energy* **2025**, *382*, 125328. [\[CrossRef\]](#)
41. Berrisch, J.; Ziel, F. Multivariate probabilistic CRPS learning with an application to day-ahead electricity prices. *Int. J. Forecast.* **2024**, *40*, 1568–1586. [\[CrossRef\]](#)
42. Nitka, W.; Weron, R. Combining predictive distributions of electricity prices: Does minimizing the CRPS lead to optimal decisions in day-ahead bidding? *Oper. Res. Decis.* **2023**, *33*, 105–118. [\[CrossRef\]](#)
43. Xu, Y.; Li, J.; Wang, H.; Du, P. A novel probabilistic forecasting system based on quantile combination in electricity price. *Comput. Ind. Eng.* **2024**, *186*, 109834. [\[CrossRef\]](#)
44. Narajewski, M. Probabilistic Forecasting of German Electricity Imbalance Prices. *Energies* **2022**, *15*, 4976. [\[CrossRef\]](#)
45. Chaweewat, P.; Singh, J.G. An electricity price interval forecasting by using residual neural network. *Int. Trans. Electr. Energy Syst.* **2020**, *30*, e12506. [\[CrossRef\]](#)
46. Yang, W.; Wang, J.; Niu, T.; Du, P. A novel system for multi-step electricity price forecasting for electricity market management. *Appl. Soft Comput.* **2020**, *88*, 106029. [\[CrossRef\]](#)
47. Zhang, Y.; Fu, Y. Probabilistic Electricity Price Forecast with Optimal Prediction Interval. *IEEE Trans. Power Syst.* **2024**, *39*, 442–452. [\[CrossRef\]](#)
48. Ghimire, S.; Nguyen-Huy, T.; Deo, R.C.; Casillas-Pérez, D.; Prasad, R.; Salcedo-Sanz, S. Point and confidence interval prediction of electricity price based on hybrid dual-decomposition and deep learning model. *Energy* **2026**, *350*, 140692. [\[CrossRef\]](#)
49. Jiang, H.; Dong, Y.; Dong, Y.; Wang, J. Probabilistic electricity price forecasting by integrating interpretable model. *Technol. Forecast. Soc. Change* **2025**, *210*, 123846. [\[CrossRef\]](#)
50. Wang, Z.; Xu, X.; Trajcevski, G.; Zhang, K.; Zhong, T.; Zhou, F. PrEF: Probabilistic Electricity Forecasting via Copula-Augmented State Space. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 12200–12207. [\[CrossRef\]](#)

51. Klein, N.; Smith, M.S.; Nott, D.J. Deep distributional time series models and the probabilistic forecasting of intraday electricity prices. *J. Appl. Econom.* **2023**, *38*, 493–511. [[CrossRef](#)]
52. He, H.; Lu, N.; Jiang, Y.; Chen, B.; Jiao, R. End-to-end probabilistic forecasting of electricity price via convolutional neural network and label distribution learning. *Energy Rep.* **2020**, *6*, 1176–1183. [[CrossRef](#)]
53. Brusaferrri, A.; Matteucci, M.; Ramin, D.; Spinelli, S.; Vitali, A. Probabilistic day-ahead energy price forecast by a Mixture Density Recurrent Neural Network. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*; IEEE: New York, NY, USA, 2020; pp. 523–528. [[CrossRef](#)]
54. Afrasiabi, M.; Aghaei, J.; Afrasiabi, S.; Mohammadi, M. Probability density function forecasting of electricity price: Deep gabor convolutional mixture network. *Electr. Power Syst. Res.* **2022**, *213*, 108325. [[CrossRef](#)]
55. Trebbien, J.; Pütz, S.; Schäfer, B.; Nygård, H.S.; Gorjão, L.R.; Witthaut, D. Probabilistic Forecasting of Day-Ahead Electricity Prices and their Volatility with LSTMs. In *Proceedings of the 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Grenoble, France, 23–26 October 2023. [[CrossRef](#)]
56. Toubeau, J.F.; Morstyn, T.; Bottieau, J.; Zheng, K.; Apostolopoulou, D.; De Greve, Z.; Wang, Y.; Vallee, F.C. Capturing Spatio-Temporal Dependencies in the Probabilistic Forecasting of Distribution Locational Marginal Prices. *IEEE Trans. Smart Grid* **2021**, *12*, 2663–2674. [[CrossRef](#)]
57. Liu, H.; Shen, X.; Tang, X.; Liu, J. Day-Ahead Electricity Price Probabilistic Forecasting Based on SHAP Feature Selection and LSTNet Quantile Regression. *Energies* **2023**, *16*, 5152. [[CrossRef](#)]
58. Cramer, E.; Witthaut, D.; Mitsos, A.; Dahmen, M. Multivariate probabilistic forecasting of intraday electricity prices using normalizing flows. *Appl. Energy* **2023**, *346*, 121370. [[CrossRef](#)]
59. Walter, D.; Wagner, A. Probabilistic simulation of electricity price scenarios using Conditional Generative Adversarial Networks. *Energy AI* **2024**, *18*, 100422. [[CrossRef](#)]
60. Lu, X.; Qiu, J.; Lei, G.; Zhu, J. Scenarios modelling for forecasting day-ahead electricity prices: Case studies in Australia. *Appl. Energy* **2022**, *308*, 118296. [[CrossRef](#)]
61. Mashlakov, A.; Kuronen, T.; Lensu, L.; Kaarna, A.; Honkapuro, S. Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. *Appl. Energy* **2021**, *285*, 116405. [[CrossRef](#)]
62. Yu, R.; Tao, Y.; Leimgruber, F.; Esterl, T.; Stiasny, J.; Bunn, D.W.; Wen, Q.; Guo, H.; Cremer, J.L. OrderFusion: Encoding Orderbook for End-to-End Probabilistic Intraday Electricity Price Forecasting. *arXiv* **2025**, arXiv:2502.06830. [[CrossRef](#)]
63. Zhang, R.; Li, G.; Ma, Z. A Deep Learning Based Hybrid Framework for Day-Ahead Electricity Price Forecasting. *IEEE Access* **2020**, *8*, 143423–143436. [[CrossRef](#)]
64. Cantillo-Luna, S.; Moreno-Chuquen, R.; Lopez-Sotelo, J.; Celeita, D. An Intra-Day Electricity Price Forecasting Based on a Probabilistic Transformer Neural Network Architecture. *Energies* **2023**, *16*, 6767. [[CrossRef](#)]
65. Jiang, H.; Pan, S.; Dong, Y.; Wang, J. Probabilistic electricity price forecasting based on penalized temporal fusion transformer. *J. Forecast.* **2024**, *43*, 1465–1491. [[CrossRef](#)]
66. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [[CrossRef](#)]
67. Giacomini, R.; White, H. Tests of Conditional Predictive Ability. *Econometrica* **2006**, *74*, 1545–1578. [[CrossRef](#)]
68. Hansen, P.R.; Lunde, A.; Nason, J.M. The Model Confidence Set. *Econometrica* **2011**, *79*, 453–497. [[CrossRef](#)]
69. Lebedev, A.; Das, A.; Pappert, S.; Schlüter, S. Analyzing Uncertainty Quantification in Statistical and Deep Learning Models for Probabilistic Electricity Price Forecasting. *IEEE Access* **2026**, *14*, 52162–52189. [[CrossRef](#)]
70. Hirsch, S.; Ziel, F. Probabilistic Forecasting for Day-ahead Electricity Prices, Battery Trading Strategies and the Economic Evaluation of Predictive Accuracy. *arXiv* **2026**, arXiv:2604.19580.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.