WORKing papers in Management Science

WORMS/26/01

From biased point forecasts of electricity demand to accurate predictive distributions: Using LASSO and GAMLSS

Katarzyna Chęć[1]
Bartosz Uniejewski[1]
Rafał Weron[1,2]

[1] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Poland
[2] Center for Research in Energy (CoRE), Aarhus School of Business and Social Sciences, Denmark

WORMS is a joint initiative of the Management Science departments of the Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

From biased point forecasts of electricity demand to accurate predictive distributions: Using LASSO and GAMLSS

Katarzyna Chęć [a], Bartosz Uniejewski [a], Rafał Weron [a,b,*]

[a]Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland
[b]Center for Research in Energy (CoRE), Aarhus University, 8000 Aarhus C, Denmark

Abstract

Electricity demand forecasts are crucial for power system operations. Market participants frequently rely on day-ahead predictions provided by Transmission System Operators (TSOs), but these can be systematically biased and – as recent studies report – may be improved using parsimonious autoregressive models. Despite the fact that many operational and economic decisions require well-calibrated uncertainty estimates, previous work has focused on point forecasts. The key question is how to derive accurate quantile and density predictions. Here we show that processing TSO forecasts with the Least Absolute Shrinkage and Selection Operator (LASSO) brings further accuracy gains and provides strong inputs for probabilistic forecasts. Drawing on ten years of data (2016-2025) from three European and North American power markets, we find that Generalized Additive Models for Location, Scale, and Shape (GAMLSS) deliver consistently better probabilistic performance than commonly used econometric and machine learning approaches. Together, these findings highlight how regularization and flexible distributional modeling can improve uncertainty quantification of electricity demand.

Keywords: Electricity demand, Day-ahead market, LASSO, Probabilistic forecasting, GAMLSS

1. Introduction

Electricity demand (or load) forecasting remains a key part of power system management and has supported operational and strategic decisions for many decades. Market participants rely on demand predictions in their daily decision-making processes, including system planning, scheduling generation units, trading activities, or revenue prediction (Kirschen and Strbac, 2019). Historically, research on load forecasting has focused primarily on point forecasts, which summarize future demand with a single number – typically the expected value (Hong and Fan, 2016). However, probabilistic forecasts in the form of quantiles or predictive distributions are crucial for operational planning and risk management, as they alone provide a comprehensive view of potential future outcomes (Winkler et al., 2019; Petropoulos et al., 2022). Ultimately, they enable more

*Corresponding author; *email:* rafal.weron@pwr.edu.pl

informed decision-making across the energy sector (Hong et al., 2020; Uniejewski and Weron, 2021; Gneiting et al., 2023a; Nawaz et al., 2023; Maciejowska and Nitka, 2026).

In practice, market participants often rely on the freely available electricity demand forecasts provided by Transmission System Operators (TSOs). However, these official forecasts can be of limited accuracy and in many markets exhibit a systematic bias. Maciejowska et al. (2021) showed that even relatively simple autoregressive (AR) models could significantly improve TSO forecasts, lowering the MAE by as much as 37%. The analysis, however, was restricted to one market (Germany, 10.2015-09.2019) and focused solely on point forecasts, leaving the probabilistic perspective unexplored. This raises two questions. First, whether more advanced modeling techniques than simple linear regression can yield additional accuracy gains. Second, whether similar improvements can be achieved within a probabilistic forecasting framework.

In our study, we improve the approach of Maciejowska et al. (2021) by considering a much richer pool of potential regressors and use the Least Absolute Shrinkage and Selection Operator (LASSO) to identify the relevant ones, instead of using a parsimonious regression model based on expert knowledge. Secondly, we use both the TSO and the LASSO-based point forecasts as inputs to probabilistic models for quantile and density forecasting. We focus in particular on Generalized Additive Models for Location, Scale and Shape (GAMLSS; Rigby and Stasinopoulos, 2005), which have seen only limited application in electricity demand forecasting.

The GAMLSS framework allows each parameter of the chosen distribution to be modeled as a function of explanatory variables, with potentially different covariate sets. It also enables non-linearity to be included in the estimation process, making it a highly flexible tool for forecasting applications (Ziel, 2022). Despite these advantages, the literature on electricity demand forecasting with GAMLSS is sparse.

The study most closely related to our work is Browell and Fasiolo (2021), who improve probabilistic forecasts of regional net load in Great Britain by modeling the distribution's tails more accurately. The authors generate point forecasts using Generalized Additive Models (GAMs; Hastie and Tibshirani, 1990) and utilize quantile regression (QR; Koenker, 2005) to estimate the central quantiles. To better approximate extreme quantiles, they use the Generalized Pareto Distribution (GPD), either in a static variant with constant parameters or a conditional one with parameters estimated by GAMLSS. They conclude that the latter approach yields sharper prediction intervals (PIs) than static GPDs while retaining calibration, and enable more efficient reserve setting.

Other related electricity demand forecasting papers include Beykirch et al. (2021), who applied GAMLSS-based predictions for anomaly detection in building heat load time series, and Gilbert et al. (2023), who employed GAMLSS to predict the timing and level of daily peak demand at low voltage levels. In a wider energy forecasting context, GAMLSS has seen limited application in short-term electricity price forecasting (Serinaldi, 2011; Hirsch et al., 2024), solar irradiance forecasting (Brabec et al., 2015; Bakker et al., 2019; Yagli et al., 2020; Yang and van der Meer, 2021), and wind forecasting (Gilbert et al., 2021; Dong et al., 2024).

Although the above papers tackle important problems in energy forecasting, they do not explicitly consider GAMLSS-based models to predict day-ahead electricity demand. In our study, we consider GAMLSS with three parametric distributions: Normal, Student's T and Johnson's SU (JSU), and benchmark them against well and less known nonparametric (historical simulation – HS, conformal prediction – CP, isotonic distributional regression – IDR, and quantile regression

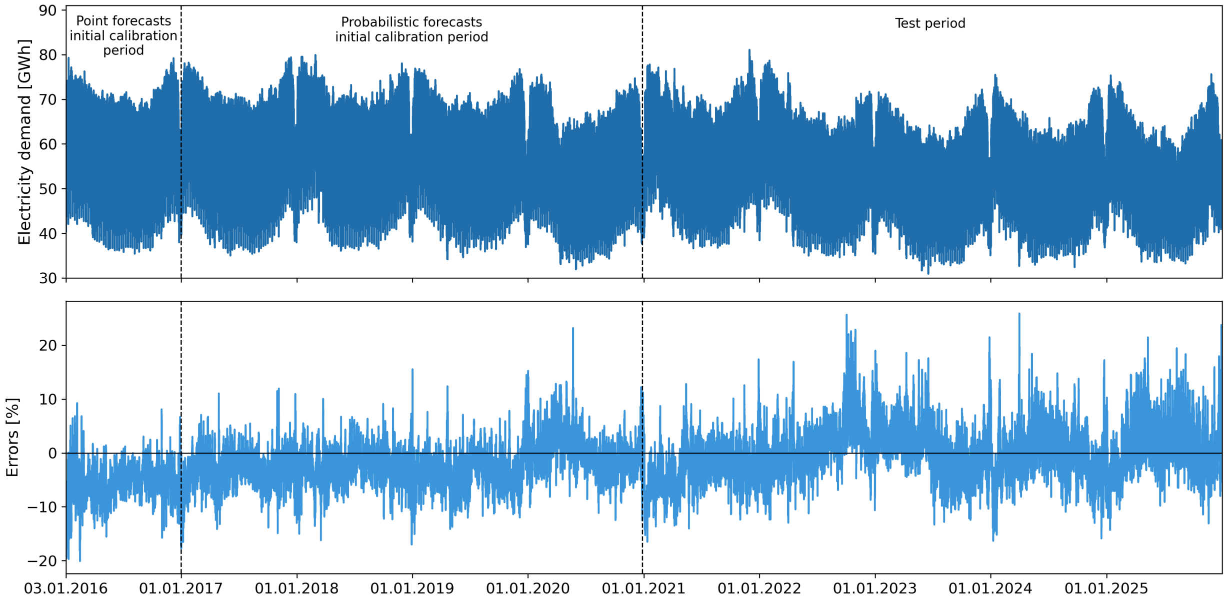

Figure 1: Electricity demand (*top*) and percentage errors $\varepsilon^{\mathrm{TSO}}_{d,h} = (\hat{L}^{\mathrm{TSO}}_{d,h} - L_{d,h})/L_{d,h}$ of the TSO's day-ahead forecasts relative to realized demand (*bottom*) in the German electricity market (EPEX-DE). Vertical dashed lines indicate the end dates of the initial calibration (training) windows for the point and probabilistic forecasts. Each day, forecasts for the 24 hours of the following day are computed, and the windows are then rolled forward by one day.

– QR), as well as parametric (GARCH and Natural Gradient Boosting – NGBoost) probabilistic forecasting methods. To evaluate the performance of these approaches, we analyze ten years of historical data (2016–2025) from two European (EPEX-DE, Germany and POLEX, Poland) and one North American electricity market (ISO-NE, New England). Predictive accuracy is assessed via the pinball score (Nowotarski and Weron, 2018; Gneiting et al., 2023b), and statistical significance is tested with the multivariate version of the Diebold-Mariano (DM) test (Ziel and Weron, 2018; Lago et al., 2021).

The remainder of the paper is structured as follows. In Section 2 we introduce the datasets. In Sections 3 and 4 we describe the point and probabilistic forecasting methods. Then, is Section 5 we discuss the results in terms of prediction accuracy and statistical significance. Finally, in Section 6 we wrap up the findings and conclude.

2. Markets and data

The German market (EPEX-DE; see Figure 1) is one of the most extensively studied in the literature (Berk et al., 2018; Maciejowska et al., 2021; Berrisch and Ziel, 2024; Chęć et al., 2025; Maciejowska and Nitka, 2026), largely due to its central position in Europe and high liquidity across various market segments. Over the last decade, the proportion of electricity generated from renewable energy sources (mainly solar and wind) has grown significantly in this market, reaching over 60% in 2024. In contrast, the Polish market (POLEX; see Figure 2) has received less attention (Dudek, 2016; Uniejewski and Weron, 2021; Janczura and Wójcik, 2022; Maciejowska

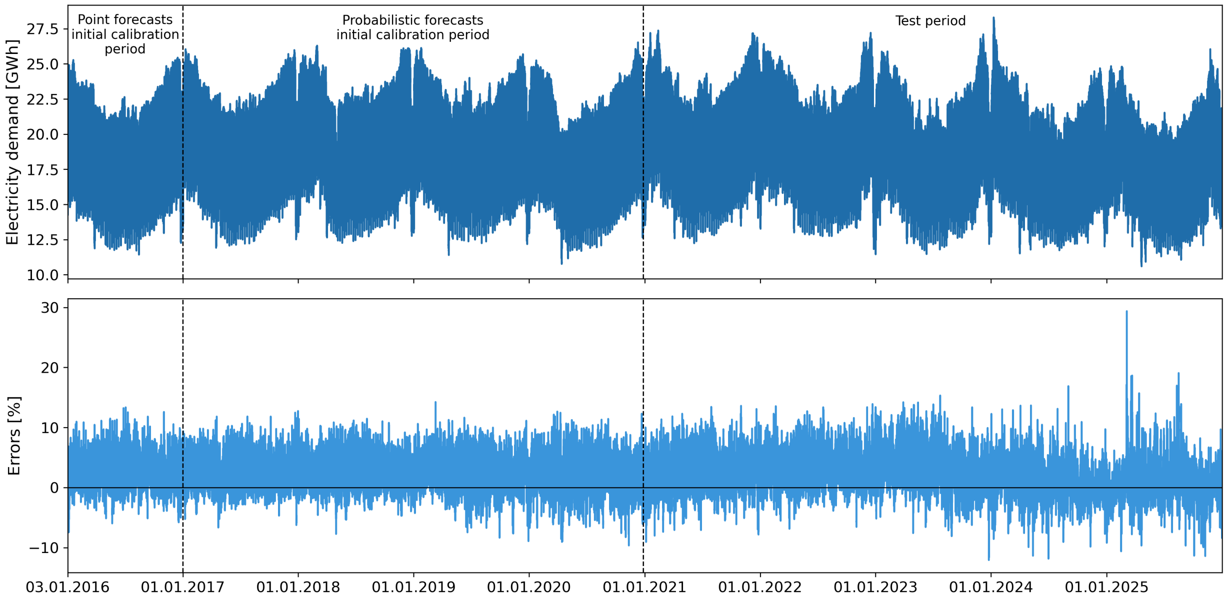

Figure 2: Electricity demand (*top*) and percentage errors of the TSO's day-ahead forecasts relative to realized demand (*bottom*) in the Polish electricity market (POLEX). Compare with Fig. 1.

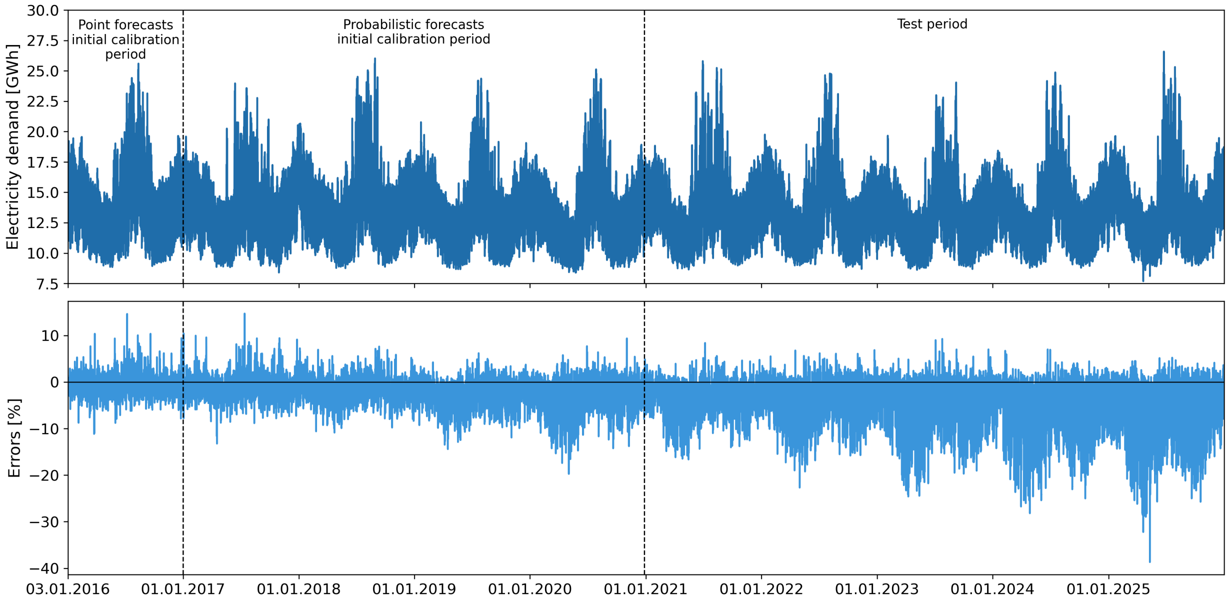

Figure 3: Electricity demand (*top*) and percentage errors of the TSO's day-ahead forecasts relative to realized demand (*bottom*) in the New England electricity market (ISO-NE). Compare with Figs. 1 and 2.

Table 1: Six representative cities for each market. The system-wide temperature forecast is the arithmetic mean of the forecasts for these six cities.

Market	Cities
Germany (EPEX-DE)	Berlin, Cologne, Frankfurt am Main, Hamburg, Leipzig, Munich
Poland (POLEX)	Gdańsk, Katowice, Lublin, Poznań, Warsaw, Wrocław
New England (ISO-NE)	Bangor, Boston, Burlington, Concord, Hartford, Providence

et al., 2024), despite also being centrally located within Europe. Its generation has traditionally relied on fossil fuels, but it has been gradually shifting towards a more diverse energy mix. In 2016, almost 90% of its generation came from fossil fuels, but by 2025, this figure had dropped to 71%. Poland shares similar economic and geographic contexts with Germany, which often results in comparable market responses to regional developments. Including the New England (ISO-NE; see Figure 3) market provides an important extension of the analysis, as it operates under different market structure, regulatory rules, climate and consumer behavior. Unlike the other two markets, it remains strongly dependent on natural gas-fired generation and also draws heavily on nuclear energy.

European electricity data are sourced from ENTSO-E (`transparency.entsoe.eu`) and data for the New England market from the official website of the Independent System Operator (ISO; `iso-ne.com`). The day-ahead temperature forecasts are obtained from Open-Meteo (`open-meteo.com`; specifically, we use the JMA Seamless model). All three data sources are open access. To approximate system-wide weather conditions for a given hour, we average temperature forecasts from six representative cities in each market, see Table 1.

All datasets are downloaded in hourly resolution and cover the period from 3.01.2016 to 31.12.2025. Both point and probabilistic forecasts are generated using a rolling-window approach. Each day, forecasts for the 24 hours of the following day are computed, and the windows are then rolled forward by one day. The initial training window for the point-forecasting models (see Section 3) spans the first 52 weeks of the dataset (3.01.2016–31.12.2016). The initial training window for the post-processing methods used to produce probabilistic forecasts (see Section 4) covers the subsequent 208 weeks (1.01.2017–26.12.2020); this long training period is essential to ensure stable and reliable estimation. The out-of-sample test period, spanning roughly four years, begins on 27.12.2020 and ends on 31.12.2025. The datasets were pre-processed to address missing and duplicate values arising from transitions to and from daylight saving time: missing observations were replaced by the mean of the two adjacent values, and duplicate observations by their mean.

3. Improving TSO point forecasts

The electricity demand point forecasts publicly disclosed by the Transmission System Operators (TSOs) are consistently inaccurate and systematically biased. This is evident in the lower panels of Figures 1–3 in which we plot the percentage errors $\varepsilon^{\mathrm{TSO}}_{d,h} = (\hat{L}^{\mathrm{TSO}}_{d,h} - L_{d,h})/L_{d,h}$ of the TSOs' day-ahead forecasts relative to realized demand. As shown by Maciejowska et al. (2021), these forecasts can be substantially improved using simple regression models. Building on this insight, we first generate refined point predictions and then use them as inputs to probabilistic forecasting models. As is common in energy forecasting (Ziel and Weron, 2018), both point and probabilistic forecasts are generated independently for each hour of the target day.

3.1. The LASSO-estimated AutoRegressive (LEAR) model

To obtain more accurate point forecasts of electricity demand, we use a parameter-rich LASSO-Estimated AutoRegressive (LEAR) model, introduced to the load forecasting literature by Dudek (2016) and Ziel and Liu (2016). Inspired by the variant proposed for price forecasting by Lago et al. (2021), who coined the acronym LEAR, we adopt the following model structure which is able to capture cross-hour interactions:

$$\begin{aligned} L_{d,h} = \phi_0 + & \underbrace{\sum\nolimits_{i=1}^{24} \left(\phi_i \hat{L}^{\text{TSO}}_{d,i} + \phi_{i+24} \hat{L}^{\text{TSO}}_{d-1,i} + \phi_{i+48} \hat{L}^{\text{TSO}}_{d-2,i} + \phi_{i+72} \hat{L}^{\text{TSO}}_{d-7,i}\right)}_{\text{electricity demand forecasts}} \\ & + \underbrace{\sum\nolimits_{i=1}^{24} \left(\phi_{i+96} L_{d-2,i} + \phi_{i+120} L_{d-7,i}\right)}_{\text{autoregressive effects}} + \underbrace{\sum\nolimits_{i=1}^{24} \phi_{144+j} \hat{T}_{d,i}}_{\text{temperature forecasts}} + \underbrace{\sum\nolimits_{j=1}^{7} \phi_{168+j} D_j}_{\text{daily dummies}} + \varepsilon_{d,h}, \end{aligned} \tag{1}$$

where the initial 96 regressors with coefficients $\phi_1, ..., \phi_{96}$ are the day-ahead TSO forecasts of electricity demand $\hat{L}^{\text{TSO}}$ for days d, $d-1$, $d-2$ and $d-7$, and hours $i = 1, ..., 24$, while the next 48 regressors with coefficients $\phi_{97}, ..., \phi_{144}$ represent the autoregressive effects and capture the impact of the realized demand from days $d-2$ and $d-7$. Note that at the time the predictions are computed, i.e., 9am on day $d-1$, yesterday's realized demand $L_{d-1,i}$ is not known for all hours i. The subsequent 24 inputs with coefficients $\phi_{145}, ..., \phi_{168}$ correspond to day-ahead temperature forecasts $\hat{T}_{d,i}$ for day d. The final 7 features with coefficients $\phi_{169}, ..., \phi_{175}$ are day-of-the-week dummies; alternatively we could use an intercept and 6 dummies. All regressors, except the dummies, are standardized before estimating the coefficients.

The LEAR model is estimated in Python using a combination of the `LassoLarsIC()` and `Lasso()` functions, where the former is used to select the regularization parameter based on the Bayesian Information Criterion (BIC), and the latter to fit the linear model with the Least Absolute Shrinkage and Selection Operator (LASSO) of Tibshirani (1996); the intercept is not regularized in this setting. See Lago et al. (2021) and the associated `epftoolbox` package for reference. Note that LASSO is only one of many regularization schemes. Still, despite its simplicity, it is an all-round performer. For a recent comparison in an energy forecasting context and possible alternatives, see Uniejewski (2024).

3.2. The benchmark ARX model

We benchmark point predictions from the LEAR model against a parsimonious autoregressive model with exogenous variables (ARX):

$$\begin{aligned} L_{d,h} = \phi_0 + & \underbrace{\phi_1 \hat{L}^{\text{TSO}}_{d,h} + \phi_2 \hat{L}^{\text{TSO}}_{d-1,h} + \phi_3 \hat{L}^{\text{TSO}}_{d-2,h} + \phi_4 \hat{L}^{\text{TSO}}_{d-7,h}}_{\text{electricity demand forecasts}} \\ & + \underbrace{\phi_5 L_{d-2,h} + \phi_6 L_{d-7,h}}_{\text{autoregressive effects}} + \underbrace{\phi_7 \hat{T}_{d,h}}_{\text{temperature forecast}} + \underbrace{\sum\nolimits_{i=1}^{6} \phi_{i+7} D_i}_{\text{daily dummies}} + \varepsilon_{d,h}. \end{aligned} \tag{2}$$

The ARX model follows a similar specification to Eq. (1), but instead of taking into account all 24 hourly values from day d, it uses only the value corresponding to hour h. The ARX model is estimated in Python via ordinary least squares (OLS). Note that compared to the LEAR model, we drop one dummy variable to avoid collinearity.

4. Probabilistic forecasting

Our objective is to estimate 99 predictive quantiles of electricity demand for each of the 24 hours of day d, using information available at 9am on day $d-1$. To this end, we consider both parametric (GAMLSS, GARCH, NGBoost) and nonparametric (HS, CP, IDR, QR) methods. Quantile forecasting (HS, CP, QR) allow us to obtain the 99 percentiles directly, whereas density forecasting (GAMLSS, GARCH, NGBoost) yields parameter estimates of a chosen distribution, which can then be used to compute the desired percentiles. The remaining approach (IDR) yields values of the cumulative distribution function (CDF) at a finite set of response points observed in the calibration window; quantiles at other probability levels are then obtained via interpolation.

4.1. Parametric methods

4.1.1. GAMLSS

The Generalized Additive Model for Location, Scale and Shape (GAMLSS) was introduced by Rigby and Stasinopoulos (2005) as a novel approach to distributional forecasting. The central idea of this framework is that each parameter of a chosen distribution – such as location, scale, and shape – can be modeled as a function of explanatory variables.

Building on the foundations of Generalized Additive Models (GAM), GAMLSS allows for a wide variety of distributions to be modeled. It can be applied in both linear and nonlinear settings, with the latter supporting smooth functions of continuous predictors. The key strength of GAMLSS lies in its adaptability: each distributional parameter can be modeled using different sets of covariates and functional forms, offering tailored specifications for different parameters of the distribution. Despite this flexibility, GAMLSS remains computationally efficient, making it well-suited for estimating predictive distributions (Ziel, 2022).

GAMLSS assumes that observations of the response variable, i.e., electricity demand $\boldsymbol{L} = \{L_{d,h}\}$, are independent and $\mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$-distributed for all days d and hours h, where $\boldsymbol{\mu} = \{\mu_{d,h}\}$ and $\boldsymbol{\sigma} = \{\sigma_{d,h}\}$ are respectively the location and scale parameters, and $\boldsymbol{\nu} = \{\nu_{d,h}\}$ and $\boldsymbol{\tau} = \{\tau_{d,h}\}$ are shape parameters related to the skewness and kurtosis. We work with three distributions $\mathcal{D}$ of varying flexibility: normal with the location and scale parameters (N), Student's t with the location, scale and shape (kurtosis) parameter (T), and Johnson's SU with the location, scale, skewness and shape parameters (JSU). We estimate the models using the `gamlss` package in R and adopt the default link functions $g_k(\cdot)$ for the distributions considered (see Table 6.1 in Stasinopoulos et al., 2017): the identity link for the location and the skewness parameter (JSU), and the logarithmic link for the scale and the kurtosis parameter (T, JSU).

For each distribution, we have two scenarios. In the first, only $\boldsymbol{\mu}$ is modeled as a function of explanatory variables:

$$g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta_1} + s_1(\hat{\boldsymbol{T}}), \tag{3}$$

where $g_1(\boldsymbol{\mu}) = \boldsymbol{\mu}$ is the identity link function, $s_1(\cdot)$ is a nonparametric smoothing function based on convex splines (or C-splines; Meyer, 2008)[1] and

$$\boldsymbol{X\beta_k} = \beta_{k,0} + \underbrace{\beta_{k,1}\tilde{L}_{d,h} + \beta_{k,2}\tilde{L}_{d-1,h} + \beta_{k,3}\tilde{L}_{d-2,h} + \beta_{k,4}\tilde{L}_{d-7,h}}_{\text{TSO or LEAR-corrected demand forecasts}} + \underbrace{\beta_{k,5}L_{d-2,h} + \beta_{k,6}L_{d-7,h}}_{\text{autoregressive effects}} + \underbrace{\textstyle\sum_{j=1}^{6}\beta_{k,j+6}D_j}_{\text{daily dummies}}. \tag{4}$$

The remaining parameters (when applicable) are modeled as constants: $g_2(\boldsymbol{\sigma}) = \beta_{2,0}$, $g_3(\boldsymbol{\nu}) = \beta_{3,0}$, and $g_4(\boldsymbol{\tau}) = \beta_{4,0}$. Regressors $\tilde{L}_{d,h} \in \{\hat{L}^{\text{TSO}}_{d,h}, \hat{L}^{\text{LEAR}}_{d,h}\}$ are either TSO or LEAR-corrected demand forecasts, see Eq. (1). Note the similarity of Eq. (4) to Eq. (1). In other words, we assume that the expected value of electricity demand can be explained using the same variables as the location parameter – or the location and scale parameters – of its distribution. The main difference is that we now let the data determine the relationship between the response and the explanatory variables, rather than imposing a linear form.

In the second scenario, both $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are expressed as functions of covariates:

$$\begin{cases} g_1(\boldsymbol{\mu}) = \boldsymbol{X\beta_1} + s_1(\hat{\boldsymbol{T}}), \\ g_2(\boldsymbol{\sigma}) = \boldsymbol{X\beta_2} + s_2(\hat{\boldsymbol{T}}), \end{cases} \tag{5}$$

where $g_1(\boldsymbol{\mu}) = \boldsymbol{\mu}$ and $g_2(\boldsymbol{\sigma}) = \log \boldsymbol{\sigma}$ are respectively the identity and the logarithmic link functions, and $s_1(\cdot)$ and $s_2(\cdot)$ are nonparametric smoothing functions based on convex splines. Any remaining parameters are modeled as constants. We also experimented with scenarios in which all three (T) or all four (JSU) parameters were modeled as functions of explanatory variables. However, this did not improve predictive performance. Nor did using P-splines, as in Zimmermann and Ziel (2025) for mid-term forecasting of hourly electricity load using GAMs.

In addition, for the normal distribution we consider a simplified scenario, called GAMLSS$_0$, which uses a single input variable – the TSO or LEAR-corrected demand forecasts – instead of Eq. (3): $g_1(\boldsymbol{\mu}) = \boldsymbol{\mu} = \beta_0 + \beta_1\tilde{L}_{d,h}$. The scale parameter is modeled with a constant, i.e., $g_2(\boldsymbol{\sigma}) = \log \boldsymbol{\sigma} = \beta_0$.

4.1.2. GARCH

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model has become a standard tool for describing time-varying volatility in financial econometrics (Andersen et al., 2006). Among its many specifications, GARCH(1,1) is widely regarded as the workhorse model because it strikes a balance between realism and parsimony. In out-of-sample forecasting, it often performs better than more complex variants (Hansen and Lunde, 2005).

In the GARCH(1,1) model, the conditional variance of the response variable on day d and hour h is defined as:

$$\sigma^2_{d,h} = \alpha_0 + \alpha_1\epsilon^2_{d-1,h} + \alpha_2\sigma^2_{d-1,h}, \tag{6}$$

where $\alpha_0 > 0$, $\alpha_1, \alpha_2 \geq 0$, and $\epsilon_{d-1,h}$ denotes the prediction error for day $d-1$ and hour h. The parameters admit an intuitive interpretation: α_1 controls how strongly volatility responds to new shocks, while α_2 governs the persistence of volatility over time. To obtain the full predictive distribution, we assume that electricity demand $L_{d,h}$ for day d and hour h is distributed as $\text{N}(\tilde{L}_{d,h}, \hat{\sigma}_{d,h})$, where $\tilde{L}_{d,h} \in \{\hat{L}^{\text{TSO}}_{d,h}, \hat{L}^{\text{LEAR}}_{d,h}\}$ and $\hat{\sigma}_{d,h}$ is obtained from Eq. (6).

[1] We use the function cSpline in R with default options.

4.1.3. NGBoost

Following a recent trend in the forecasting literature (Brusaferri et al., 2022; Marcjasz et al., 2023; Morier and Valls Pereira, 2025), we adopt a distributional machine learning approach that directly estimates the parameters of a chosen predictive distribution. Specifically, we use Natural Gradient Boosting (NGBoost) of Duan et al. (2020), a probabilistic regression framework that extends gradient boosting. The algorithm proceeds iteratively by sequentially adding decision trees, referred to as base learners, with each new learner trained to correct the errors of the current ensemble. In this way, predictive performance is gradually improved through boosting.

When training probabilistic models, different distribution parameters can affect the predictive distribution on different scales, which can make standard gradient-based optimization unstable. NGBoost addresses this issue by using the so-called natural gradient, which modifies the usual gradient so that updates are measured in terms of changes to the predicted probability distribution rather than changes to the raw model parameters.

In our setup, NGBoost outputs the location and scale parameters of a normal distribution. These parameters are learned jointly by maximizing the likelihood and are allowed to depend on the input features. We use the open-source `ngboost` package in Python with default hyperparameters, in order to focus on the model's probabilistic forecasting capabilities rather than the effects of hyperparameter optimization.

4.2. Nonparametric methods

4.2.1. Historical simulation and conformal prediction

Historical simulation (HS) and conformal prediction (CP) are very simple, model-independent approaches in which the predicted α-quantile is obtained by adding a point forecast to the empirical α-quantile of the forecast errors (HS) or of the absolute forecast errors (CP). Although similar in concept, the methods originated in two different communities. HS can be traced back at least to the early 1990s and the beginnings of Value-at-Risk (VaR) in risk management practice (Hendricks, 1996), while CP originated in the machine learning literature (Vovk et al., 2005) and consequently uses different terminology.

Specifically, the HS-derived conditional α-quantile is given by:

$$\hat{Q}_{\alpha|\hat{L}_{d,h}} = \hat{L}_{d,h} + Q_{\alpha}^{\text{emp}}(\varepsilon), \tag{7}$$

where $\hat{L}_{d,h}$ is the point forecast of electricity demand and $Q_{\alpha}^{\text{emp}}(\varepsilon)$ represents the sample α-quantile of the prediction errors $\varepsilon_{d,h} = L_{d,h} - \hat{L}_{d,h}$.

In contrast, conformal prediction works with so-called non-conformity scores $\lambda_{d,h}$'s. In its basic form, $\lambda_{d,h} \equiv |\varepsilon_{d,h}| = |L_{d,h} - \hat{L}_{d,h}|$; see Kato et al. (2023) for alternative choices of non-conformity measures. The CP-derived conditional α-quantile is defined by:

$$\hat{Q}_{\alpha|\hat{L}_{d,h}} = \hat{L}_{d,h} - \mathbb{1}_{\alpha\leq 0.5} Q_{2\alpha}^{\text{emp}}(\lambda) + \mathbb{1}_{\alpha\geq 0.5} Q_{2(1-\alpha)}^{\text{emp}}(\lambda), \tag{8}$$

where $Q_{\alpha}^{\text{emp}}(\lambda)$ is the sample α-quantile of the $\lambda_{d,h}$'s. Note that $[\hat{L}_{d,h} - Q_{\alpha}^{\text{emp}}(\lambda), \hat{L}_{d,h} + Q_{\alpha}^{\text{emp}}(\lambda)]$ is a valid $(1-\alpha)$-prediction interval (PI), i.e., the non-conformity score defines the width of the interval. Consequently, unlike in historical simulation, the PIs are centered on the point forecasts. To compute the HS and CP forecasts for the 99 percentiles, we use the `point2quant()` function in the PostForecasts.jl package in Julia (Lipiecki and Weron, 2025).

4.2.2. IDR

Isotonic Distributional Regression (IDR) was introduced by Henzi et al. (2021) as a method for estimating the conditional cumulative distribution function (CDF) under an isotonicity constraint. This constraint implies that the quantiles of the conditional distribution are non-decreasing with respect to the regressor; see Lipiecki et al. (2024) for a detailed description and a schematic representation of the algorithm in an energy forecasting context.

The IDR provides estimated CDF values at a finite set of response points observed in the calibration window. Quantiles at other probability levels, e.g., the 99 percentiles, are obtained via interpolation. As for historical simulation and conformal prediction, we compute the IDR forecasts for the 99 percentiles using the `point2quant()` function from the PostForecasts.jl package in Julia (Lipiecki and Weron, 2025).

4.2.3. Quantile regression

The quantile regression (QR; Koenker, 2005) approach models conditional quantiles of the dependent variable given a set of explanatory variables. For a given quantile level $\alpha \in (0, 1)$, the conditional quantile of electricity load is assumed to be linear in the covariates:

$$\hat{Q}_{\alpha|\mathbf{X}_{d,h}} = \mathbf{X}_{d,h}\boldsymbol{\beta}_{\alpha}. \tag{9}$$

The estimation procedure involves minimizing an asymmetric loss function tailored to a specific quantile level α, i.e., the so-called pinball score (also known as the quantile loss; Nowotarski and Weron, 2018; Gneiting et al., 2023b):

$$\min_{\boldsymbol{\beta}_{\alpha}} \Big[\sum_{d,h} \underbrace{\Big(\alpha - \mathbb{1}_{L_{d,h} < \mathbf{X}_{d,h}\boldsymbol{\beta}_{\alpha}}\Big)\Big(L_{d,h} - \mathbf{X}_{d,h}\boldsymbol{\beta}_{\alpha}\Big)}_{\text{pinball score}} \Big], \tag{10}$$

where $\mathbf{X}_{d,h} = [1, \tilde{L}_{d,h}, \tilde{L}_{d-1,h}, \tilde{L}_{d-2,h}, \tilde{L}_{d-7,h}, L_{d-2,h}, L_{d-7,h}, \hat{T}_{d,h}, D_1, ..., D_6]$ denotes the vector of explanatory variables for day d and hour h, $\boldsymbol{\beta}_{\alpha}$ is the vector of regression coefficients, and $L_{d,h}$ is the observed electricity demand for day d and hour h. Notably, to construct a full picture of the target distribution, this minimization must be carried out separately for each desired quantile level. We use the `QuantileRegressor()` function from the `scikit-learn` package in Python to compute the quantile forecasts.

Similarly to GAMLSS, see Section 4.1.1, we also consider a variant of QR that uses a single vector of point predictions as the only input variable, i.e., $\tilde{L}_{d,h}$. This special case goes in line with the Quantile Regression Averaging (QRA) method of Nowotarski and Weron (2015), who were the first to propose using point predictions as inputs to quantile regression. We refer to this special case as QR_0. Note that as for the other non-parametric methods, see Sections 4.2.1 and 4.2.2, the QR_0 forecasts can be computed using the `point2quant()` function from the PostForecasts.jl package in Julia (Lipiecki and Weron, 2025).

Table 2: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for TSO, ARX-based, and LEAR-based point forecasts over the test period 27.12.2020–31.12.2025, with cell coloring applied independently for each column and each market (red → high, green → low). Columns labeled "%chng." show the percentage difference (improvement) with respect to the corresponding error metric of the TSO forecasts.

	MAE	%chng.	RMSE	%chng.
		EPEX-DE		
TSO	2006.98		2568.72	
ARX	1489.86	-25.77%	1966.93	-23.43%
LEAR	1385.72	-30.96%	1836.64	-28.50%
		POLEX		
TSO	505.35		657.35	
ARX	352.77	-30.19%	469.60	-28.56%
LEAR	343.87	-31.95%	457.91	-30.34%
		ISO-NE		
TSO	477.48		714.55	
ARX	202.50	-57.59%	295.69	-58.62%
LEAR	151.22	-68.33%	218.99	-69.35%

5. Results

5.1. Point forecasts

The prediction accuracy of point forecasts is measured by the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE):

$$\text{MAE} = \frac{1}{24N_d} \sum_{d=1}^{N_d} \sum_{h=1}^{24} |L_{d,h} - \hat{L}_{d,h}|, \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{1}{24N_d} \sum_{d=1}^{N_d} \sum_{h=1}^{24} (L_{d,h} - \hat{L}_{d,h})^2}, \tag{12}$$

where $N_d = 1831$ stands for the number of days in the test period (i.e., 27.12.2020–31.12.2025; see Figures 1-3), while $L_{d,h}$ and $\hat{L}_{d,h}$ respectively denote the actual and predicted electricity demand for day d and hour h.

Table 2 reports the point forecast accuracy of the TSO, ARX, and LEAR models. Even the parsimonious ARX model substantially improves over the TSO benchmark. In terms of MAE, it reduces forecast errors by approximately 26% for Germany, 30% for Poland, and 58% for New England; reductions is RMSE are of a similar magnitude. Compared to ARX, LEAR achieves additional reductions in all markets – the largest in New England and the smallest in Poland. Compared to the TSO benchmark, LEAR reduces MAE by 31% for Germany, 32% for Poland, and 68% for New England; again reductions is RMSE are of a similar magnitude.

Overall, these results indicate that allowing the model to select relevant predictors from a large information set leads to substantial accuracy gains over both TSO and ARX forecasts. Because LEAR consistently outperforms ARX, we excluded ARX from the subsequent probabilistic forecasting analysis.

5.2. Probabilistic forecasts

Probabilistic forecast performance is usually assessed based on sharpness, reflecting the concentration or spread of the predictive distribution (Nowotarski and Weron, 2018; Gneiting et al.,

Table 3: The aggregate pinball score (APS), i.e., the pinball score defined by Eq. (10) averaged for the entire test period and across all 99 percentiles, for the German (EPEX-DE), Polish (POLEX) and New England (ISO-NE) markets. Cell coloring is applied independently for each market (red → high, green → low).

	Non-parametric forecasting							Parametric forecasting						
	HS	CP	IDR	QR_0	QR	GARCH	$GAMLSS_0$	$GAMLSS^N$		$GAMLSS^T$		$GAMLSS^{JSU}$		NGBoost
								μ	μ,σ	μ	μ,σ	μ	μ,σ	
	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\mathbf{X}_{d,h}$	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$
EPEX-DE														
TSO	721.21	722.09	728.43	721.67	537.26	705.92	720.03	527.68	525.76	528.65	526.54	529.61	528.00	571.07
LEAR	505.52	504.86	515.41	505.43	497.69	499.29	505.55	488.20	484.58	487.88	484.69	488.66	485.88	506.35
POLEX														
TSO	153.29	181.51	151.71	147.86	128.70	184.63	147.59	125.40	125.45	125.62	125.59	126.00	125.71	137.76
LEAR	124.47	124.44	129.39	124.57	122.09	123.35	124.43	118.70	119.08	118.93	119.17	119.34	119.39	126.60
ISO-NE														
TSO	93.93	162.51	83.00	86.17	74.10	178.10	87.45	62.29	60.96	61.62	60.60	61.66	60.71	66.59
LEAR	55.69	55.69	58.27	55.28	54.33	55.33	56.11	54.42	53.70	53.94	53.42	54.00	53.47	57.99

2023b). For a single quantile α, we can measure sharpness using the pinball score defined in Eq. (10). To evaluate the full predictive distribution, pinball scores can be averaged across different α's; for an equidistant dense grid of quantiles the latter leads to the continuous ranked probability score (CRPS; see Nitka and Weron, 2023; Berrisch and Ziel, 2024). Further, pinball scores can be averaged across a certain time period, e.g., all hours in the test period. In this study, we report the results of the so-called aggregate pinball score (APS; Uniejewski and Weron, 2021) for the entire test period and all 99 percentiles (→ APS), or the extreme 20 percentiles, i.e., 0.01,0.02,...,0.10 and 0.90,0.91,...,0.99 (→ APS20).

Tables 3 and 4 report aggregate pinball scores for the German (EPEX-DE), Polish (POLEX), and New England (ISO-NE) electricity markets, using either TSO or LEAR point forecasts as inputs to the probabilistic models. Table 3 summarizes results across all 99 quantiles, while Table 4 focuses on the 20 most extreme quantiles. Each table is divided into two sections. The left panel presents results for the non-parametric methods: historical simulation (HS), conformal prediction (CP), Isotonic Distributional Regression (IDR), and quantile regression (QR_0 and QR). The right panel reports results for the parametric approaches: GARCH(1,1), several GAMLSS specifications, and NGBoost. Finally, the symbols $\tilde{L}_{d,h}$ and $\mathbf{X}_{d,h}$ indicate the input sets used by each method. Models based on $\tilde{L}_{d,h}$ only use the point demand forecast for day d and hour h. Models based on $\mathbf{X}_{d,h}$ use a larger information set: $\mathbf{X}_{d,h} = [1, \tilde{L}_{d,h}, \tilde{L}_{d-1,h}, \tilde{L}_{d-2,h}, \tilde{L}_{d-7,h}, L_{d-2,h}, L_{d-7,h}, \hat{T}_{d,h}, D_1, ..., D_6]$.

Based on the results presented in Tables 3 and 4 we can observe that replacing TSO forecasts with LEAR-based predictions consistently reduces the APS, both for the full set of 99 quantiles and for the 20 extreme quantiles. This holds for both non-parametric and parametric approaches. The results indicate that the improvements in point load predictions directly translate into more accurate probabilistic predictions. The effect is particularly pronounced for methods that rely solely on $\tilde{L}_{d,h}$. This pattern suggests that bias in TSO forecasts propagates into probabilistic models when no additional variables are available to correct for it.

The results show that among all specifications, the GAMLSS models with normal ($GAMLSS^N$), Student's t ($GAMLSS^T$), and Johnson's SU ($GAMLSS^{JSU}$) distributions generally deliver the most

Table 4: The aggregate pinball score (APS20), i.e., the pinball score defined by Eq. (10) averaged for the entire test period and across extreme 20 percentiles (0.01,0.02,...,0.10 and 0.90,0.91,...,0.99), for the German (EPEX-DE), Polish (POLEX), and New England (ISO-NE) markets. Cell coloring is applied independently for each market (red → high, green → low).

	Non-parametric forecasting							Parametric forecasting						
	HS	CP	IDR	QR_0	QR	GARCH	$GAMLSS_0$	$GAMLSS^N$		$GAMLSS^T$		$GAMLSS^{JSU}$		NGBoost
								μ	μ, σ	μ	μ, σ	μ	μ, σ	
	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\mathbf{X}_{d,h}$	$\tilde{L}_{d,h}$	$\tilde{L}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$	$\mathbf{X}_{d,h}$
							EPEX-DE							
TSO	281.96	281.52	292.96	280.26	217.23	263.29	284.92	216.26	212.34	217.66	212.77	217.64	212.80	246.14
LEAR	205.81	205.99	218.79	205.63	201.80	194.78	205.58	199.97	195.23	200.92	195.36	201.09	195.46	220.65
							POLEX							
TSO	59.45	69.18	63.55	57.72	51.41	75.63	57.67	49.71	49.25	49.90	49.32	49.82	49.17	59.59
LEAR	49.26	49.21	55.05	49.33	48.93	47.40	49.11	47.06	46.86	47.22	46.92	47.21	46.87	55.05
							ISO-NE							
TSO	35.92	56.48	36.42	33.09	29.75	79.62	34.70	26.17	24.86	26.13	24.81	26.18	24.88	29.42
LEAR	23.51	23.47	26.11	22.79	22.50	22.69	23.67	23.16	22.00	23.06	21.94	23.10	21.98	25.31

accurate forecasts across all markets and input scenarios. The only exception occurs for the 20 extreme quantiles in the EPEX-DE market, where GARCH(1,1) slightly outperforms the GAMLSS variants. Although differences between distributional assumptions within GAMLSS are relatively small, the results indicate that modeling both the location and scale parameters with a full set of variables improves predictive accuracy compared to specifications in which the shape parameter depends on intercept only.

Finally, we can conclude that models based solely on the point forecast input ($\tilde{L}_{d,h}$) systematically underperform specifications that use the richer covariate set ($\mathbf{X}_{d,h}$). Although GARCH(1,1) is typically the strongest among single-input parametric models – especially with LEAR inputs – it remains clearly inferior to the best GAMLSS variants and is getting markedly worse when TSO forecasts are used. Comparisons of QR_0 versus QR and $GAMLSS_0$ versus full GAMLSS show that incorporating additional covariates consistently improves predictive accuracy across markets and quantile sets. Among the parametric approaches, NGBoost performs rather weakly. In most cases, it is dominated by the GAMLSS variants and often fails to outperform simpler specifications. While its performance improves when LEAR inputs are used, it does not close the gap relative to the best probabilistic models.

5.3. *Statistical significance*

To assess whether the observed improvement is statistically significant, we use the Diebold-Mariano test (DM; Diebold, 2015) – a well-established method in both point and probabilistic forecasting contexts. In line with Ziel and Weron (2018) and Lago et al. (2021), the test is carried out in a multivariate setting, where the forecasts for all 24 hours in a day are evaluated together. Let $\pi_d^A = (\pi_{d,1}^A, ..., \pi_{d,24}^A)$ and $\pi_d^B = (\pi_{d,1}^B, ..., \pi_{d,24}^B)$ represent the daily vectors of pinball scores for models A and B, respectively. The multivariate loss differential series is calculated in the $||.||_1$-norm, and thus given by:

$$\Delta_d^{A,B} = ||\pi_d^A||_1 - ||\pi_d^B||_1, \tag{13}$$

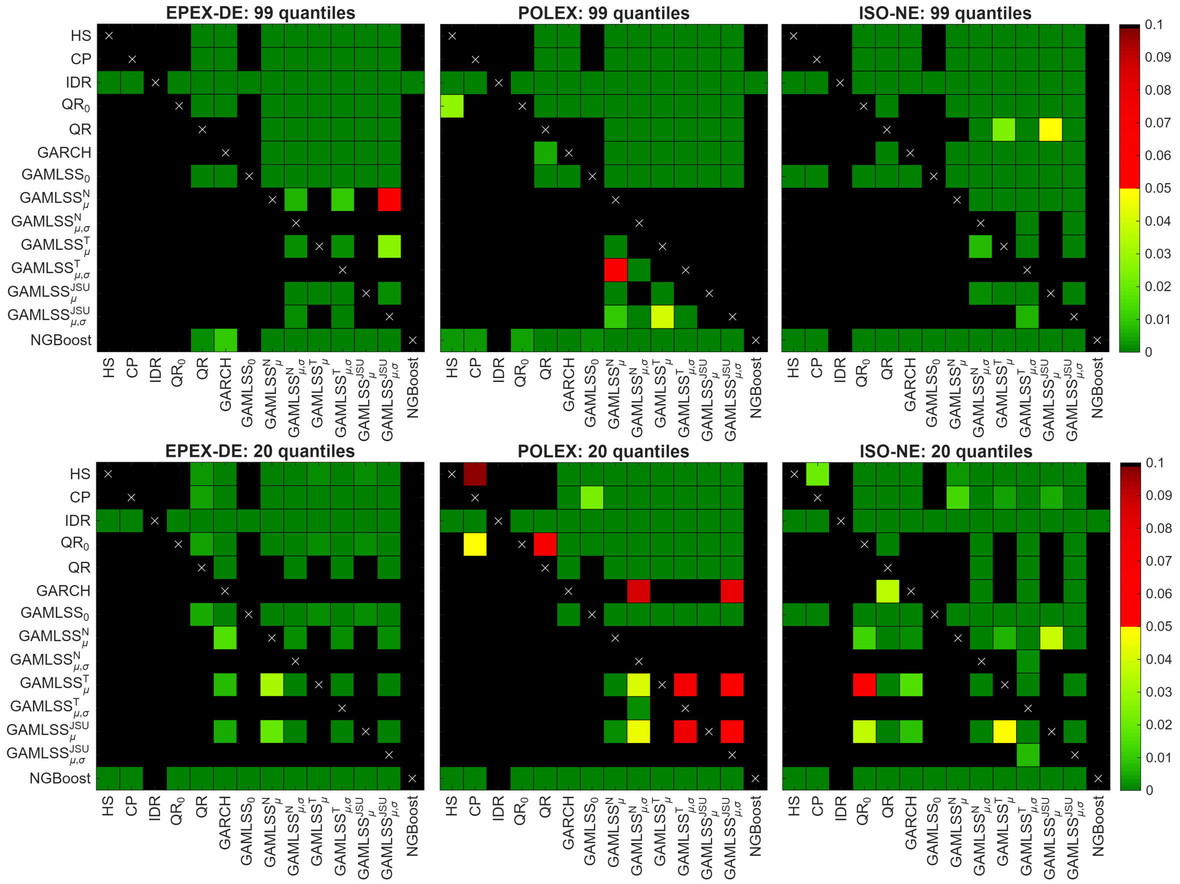

Figure 4: Results of the Diebold-Mariano (DM) test with the multivariate loss differential series given by Eq. (13), separately for all 99 percentiles (*top*) and for 20 extreme percentiles (*bottom*), for the three datasets: EPEX-DE, POLEX and ISO-NE. The results refer to the LEAR scenario, i.e., with the usage of improved point predictions as inputs. The outcomes of the test are presented using heat maps to indicate the range of the p-values. The closer they are to zero (→ dark green) the more significant is the difference between the forecasts of a model on the X-axis (better) and the forecasts of a model on the Y-axis (worse); p-values ≥ 0.10 are marked in black.

where $\|\pi_d^A\|_1 = \sum_{h=1}^{24} |\pi_{d,h}^A|$. The procedure assumes that the loss differential series is covariance stationary.

Figure 4 shows the results of the DM test for the LEAR-based models, separately for all 99 percentiles and 20 extreme percentiles. Following Ziel and Weron (2018). The DM tests largely confirm the ranking implied by Tables 3 and 4. In particular, the GAMLSS$_{\mu,\sigma}$ specifications are not significantly outperformed by any competing method, for either the full set of 99 quantiles or the 20 extreme percentiles. Consequently, the advantage of GARCH(1,1) in the EPEX-DE market for extreme quantiles is not statistically significant. Both GAMLSS$_\mu$ and GAMLSS$_{\mu,\sigma}$ significantly outperform the benchmark methods when all 99 quantiles are considered, reinforcing their strong performance in modeling the entire conditional distribution. For extreme quantiles, GAMLSS$_{\mu,\sigma}$ generally improves over GAMLSS$_\mu$.

Differences between the normal, Student's t, and JSU variants are generally small and market-dependent, with no single distribution uniformly dominating across datasets. For extreme quantiles, some distributional differences become statistically significant, but superiority remains context-specific. Consistent with earlier findings, NGBoost does not statistically dominate the GAMLSS approaches and is frequently outperformed by them.

6. Conclusions

This paper examined how improvements in point electricity demand forecasts translate into gains in probabilistic forecasting performance. We focused in particular on the Generalized Additive Models for Location, Scale and Shape (GAMLSS), a flexible framework that has received limited attention in day-ahead load forecasting. To generate enhanced point forecasts, we extended the approach of Maciejowska et al. (2021) by combining transmission system operator (TSO) forecasts, historical load realizations, and weather predictions within a high-dimensional regression setting. The Least Absolute Shrinkage and Selection Operator (LASSO) was used to select relevant predictors. Both TSO and LASSO-based forecasts were then used as inputs to a range of non-parametric and parametric probabilistic forecasting methods.

The empirical results lead to three main conclusions. First, improving point forecasts significantly increases probabilistic forecast accuracy. Using LEAR-based inputs reduces the aggregate pinball score more than TSO-based inputs do, especially for models that rely solely on point forecasts and cannot correct for bias in TSO predictions.

Second, directly modeling the conditional distribution – especially when allowing the location and scale to depend on explanatory variables – yields systematic improvements. GAMLSS specifications that incorporate covariates in both parameters rarely perform significantly worse than benchmark approaches, and they outperform the benchmarks when considering the full distribution (99 percentiles). Differences across specific distributional assumptions (normal, Student's t, and JSU) tend to be smaller and market-dependent. These results suggest that the key gains stem from flexible modeling of the conditional mean and variance rather than from a particular parametric choice.

Third, although GARCH(1,1) can perform competitively for extreme quantiles in certain markets, this advantage is not statistically significant. Likewise, a more flexible machine learning approach (NGBoost) did not systematically outperform GAMLSS models.

From an economic perspective, these findings underline the importance of accurate probabilistic load forecasts for electricity market participants. Improved modeling of conditional uncertainty supports better risk assessment and operational decision-making. The results also suggest that further applications of distributional regression frameworks – such as GAMLSS – to other TSO forecasts, including renewable generation, constitute a promising direction for future research.

Acknowledgments

This work was partially supported by the National Science Center (NCN, Poland) through grants No. 2018/30/A/HS4/00444 (to K.C. and R.W.) and No. 2023/49/N/HS4/02741 (to B.U.).

References

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting, in: Elliott, G., Granger, C., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Elsevier, pp. 777–878. doi:10.1016/S1574-0706(05)01015-3.

Bakker, K., Whan, K., Knap, W., Schmeits, M., 2019. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. Solar Energy 191, 138–150. doi:10.1016/j.solener.2019.08.044.

Berk, K., Hoffmann, A., Müller, A., 2018. Probabilistic forecasting of industrial electricity load with regime switching behavior. International Journal of Forecasting 34, 147–162. doi:10.1016/j.ijforecast.2017.09.006.

Berrisch, J., Ziel, F., 2024. Multivariate probabilistic CRPS learning with an application to day-ahead electricity prices. International Journal of Forecasting 40, 1568–1586. doi:10.1016/j.ijforecast.2024.01.005.

Beykirch, M., Janke, T., Tayeche, I., Steinke, F., 2021. Probabilistic forecast combination for anomaly detection in building heat load time series, in: 2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), pp. 1–6. doi:10.1109/ISGTEurope52324.2021.9640122.

Brabec, M., Paulescu, M., Badescu, V., 2015. Tailored vs black-box models for forecasting hourly average solar irradiance. Solar Energy 111, 320–331. doi:10.1016/j.solener.2014.11.003.

Browell, J., Fasiolo, M., 2021. Probabilistic forecasting of regional net-load with conditional extremes and gridded nwp. IEEE Transactions on Smart Grid 12, 5011–5019. doi:10.1109/TSG.2021.3107159.

Brusaferri, A., Matteucci, M., Spinelli, S., Vitali, A., 2022. Probabilistic electric load forecasting through Bayesian Mixture Density Networks. Applied Energy 309, 118341. doi:10.1016/j.apenergy.2021.118341.

Chęć, K., Uniejewski, B., Weron, R., 2025. Extrapolating the long-term seasonal component of electricity prices for forecasting in the day-ahead market. Journal of Commodity Markets 37, 100449. doi:10.1016/j.jcomm.2024.100449.

Diebold, F.X., 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. Journal of Business and Economic Statistics 33, 1–9.

Dong, S., Li, Y., Wang, M., Tao, S., 2024. Non-stationary extreme value models with periodic change for the extreme design wind speed. Ocean Engineering 311, 118823. doi:10.1016/j.oceaneng.2024.118823.

Duan, T., Avati, A., Ding, D.Y., Thai, K.K., Basu, S., Ng, A.Y., Schuler, A., 2020. NGBoost: Natural gradient boosting for probabilistic prediction. Proceedings of Machine Learning Research 119, 2690–2700. URL: https://proceedings.mlr.press/v119/duan20a.html.

Dudek, G., 2016. Pattern-based local linear regression models for short-term load forecasting. Electric Power Systems Research 130, 139–147. doi:10.1016/j.epsr.2015.09.001.

Gilbert, C., Browell, J., McMillan, D., 2021. Probabilistic access forecasting for improved offshore operations. International Journal of Forecasting 37, 134–150. doi:10.1016/j.ijforecast.2020.03.007.

Gilbert, C., Browell, J., Stephen, B., 2023. Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks. Sustainable Energy, Grids and Networks 34, 100998. doi:10.1016/j.segan.2023.100998.

Gneiting, T., Lerch, S., Schulz, B., 2023a. Probabilistic solar forecasting: Benchmarks, post-processing, verification. Solar Energy 252, 72–80. doi:10.1016/j.solener.2022.12.054.

Gneiting, T., Wolffram, D., Resin, J., Kraus, K., Bracher, J., Dimitriadis, T., Hagenmeyer, V., Jordan, A.I., Lerch, S., Phipps, K., Schienle, M., 2023b. Model diagnostics and forecast evaluation for quantiles. Annual Review of Statistics and Its Application 10, 597–621. doi:10.1146/annurev-statistics-032921-020240.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20, 873–889. doi:10.1002/jae.800.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized additive models. volume 43. CRC Press. doi:10.1201/9780203753781.

Hendricks, D., 1996. Evaluation of value-at-risk models using historical data. Economic Policy Review 2.

Henzi, A., Ziegel, J.F., Gneiting, T., 2021. Isotonic Distributional Regression. Journal of the Royal Statistical Society Series B: Statistical Methodology 83, 963–993. doi:10.1111/rssb.12450.

Hirsch, S., Berrisch, J., Ziel, F., 2024. Online Distributional Regression. URL: https://arxiv.org/abs/2407.08750, arXiv:2407.08750.

Hong, T., Fan, S., 2016. Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting 32, 914–938. doi:10.1016/j.ijforecast.2015.11.011.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H., 2020. Energy forecasting: A review and outlook. IEEE Open Access Journal of Power and Energy 7, 376–388. doi:10.1109/OAJPE.2020.3029979.

Janczura, J., Wójcik, E., 2022. 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 Economics 110, 106015. doi:10.1016/j.eneco.2022.106015.

Kato, Y., Tax, D.M., Loog, M., 2023. A review of nonconformity measures for conformal prediction in regression. Proceedings of Machine Learning Research 204, 369–383. URL: https://proceedings.mlr.press/v204/kato23a.html.

Kirschen, D., Strbac, G., 2019. Fundamentals of Power System Economics, 2nd ed. John Wiley & Sons, Chichester.

Koenker, R.W., 2005. Quantile Regression. Cambridge University Press.

Lago, J., Marcjasz, G., De Schutter, B., Weron, R., 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Applied Energy 293, 116983. doi:10.1016/j.apenergy.2021.116983.

Lipiecki, A., Uniejewski, B., Weron, R., 2024. Post-processing of point predictions for probabilistic forecasting of day-ahead electricity prices: The benefits of using isotonic distributional regression. Energy Economics 139, 107934. doi:10.1016/j.eneco.2024.107934.

Lipiecki, A., Weron, R., 2025. PostForecasts.jl: A Julia package for probabilistic forecasting by postprocessing point predictions. SoftwareX 31, 102200. doi:10.1016/j.softx.2025.102200.

Maciejowska, K., Nitka, W., 2026. Arbitrage in short-term electricity markets: Economic value of multidimensional ensemble forecasts. Operations Research and Decisions 36. doi:10.37190/ord/217347.

Maciejowska, K., Nitka, W., Weron, T., 2021. Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. Energy Economics 99, 105273. doi:10.1016/j.eneco.2021.105273.

Maciejowska, K., Serafin, T., Uniejewski, B., 2024. Probabilistic forecasting with a hybrid factor-qra approach: Application to electricity trading. Electric Power Systems Research 234, 110541. doi:10.1016/j.epsr.2024.110541.

Marcjasz, G., Narajewski, M., Weron, R., Ziel, F., 2023. Distributional neural networks for electricity price forecasting. Energy Economics 125, 106843.

Meyer, M.C., 2008. Inference using shape-restricted regression splines. Annals of Applied Statistics 2, 1013–1033. doi:10.1214/08-AOAS167.

Morier, B., Valls Pereira, P.L., 2025. Forecasting intraday volatility and densities using deep learning. Quarterly Review of Economics and Finance 104, 102076. doi:10.1016/j.qref.2025.102076.

Nawaz, K., Lahiani, A., Roubaud, D., 2023. Do natural resources determine energy consumption in Pakistan? the importance of quantile asymmetries. Quarterly Review of Economics and Finance 87, 200–211. doi:10.1016/j.qref.2020.10.003.

Nitka, W., Weron, R., 2023. Combining predictive distributions of electricity prices. Does minimizing the CRPS lead to optimal decisions in day-ahead bidding? Operations Research and Decisions 33, 103–116. doi:10.37190/ord230307.

Nowotarski, J., Weron, R., 2015. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. Computational Statistics 30, 791–803. doi:10.1007/s00180-014-0523-0.

Nowotarski, J., Weron, R., 2018. Recent advances in electricity price forecasting: A review of probabilistic forecasting. Renewable and Sustainable Energy Reviews 81, 1548–1568. doi:10.1016/j.rser.2017.05.234.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., *et al.*, 2022. Forecasting: Theory and practice. International Journal of Forecasting 38, 705–871. doi:10.1016/j.ijforecast.2021.11.001.

Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized Additive Models for Location Scale and Shape. Journal of the Royal Statistical Society Series C: Applied Statistics 54, 507–554. doi:10.1111/j.1467-9876.2005.00510.x.

Serinaldi, F., 2011. Distributional modeling and short-term forecasting of electricity prices by Generalized Additive Models for Location, Scale and Shape. Energy Economics 33, 1216–1226. doi:10.1016/j.eneco.2011.05.001.

Stasinopoulos, M.D., Rigby, R.A., Heller, G.Z., Voudouris, V., Bastiani, F.D., 2017. Flexible Regression and Smoothing: Using GAMLSS in R. 1st ed., Chapman and Hall/CRC. doi:10.1201/b21973.

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society B 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

Uniejewski, B., 2024. Regularization for electricity price forecasting. Operations Research and Decisions 34, 267–286. doi:10.37190/ord240314.

Uniejewski, B., Weron, R., 2021. Regularized quantile regression averaging for probabilistic electricity price forecasting. Energy Economics 95, 105121. doi:10.1016/j.eneco.2021.105121.

Vovk, V., Gammerman, A., Shafer, G., 2005. Algorithmic learning in a random world. Springer Science & Business Media.

Winkler, R.L., Grushka-Cockayne, Y., Lichtendahl, K.C., Jose, V.R.R., 2019. Probability forecasts and their combination: A research perspective. Decision Analysis 16, 239–260. doi:10.1287/deca.2019.0391.

Yagli, G.M., Yang, D., Srinivasan, D., 2020. Ensemble solar forecasting using data-driven models with probabilistic post-processing through gamlss. Solar Energy 208, 612–622. doi:10.1016/j.solener.2020.07.040.

Yang, D., van der Meer, D., 2021. Post-processing in solar forecasting: Ten overarching thinking tools. Renewable and Sustainable Energy Reviews 140, 110735. doi:10.1016/j.rser.2021.110735.

Ziel, F., 2022. M5 competition uncertainty: Overdispersion, distributional forecasting, GAMLSS, and beyond. International Journal of Forecasting 38, 1546–1554. doi:10.1016/j.ijforecast.2021.09.008.

Ziel, F., Liu, B., 2016. Lasso estimation for GEFCom2014 probabilistic electric load forecasting. International Journal of Forecasting 32, 1029–1037. doi:10.1016/j.ijforecast.2016.01.001.

Ziel, F., Weron, R., 2018. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. Energy Economics 70, 396–420. doi:10.1016/j.eneco.2017.12.016.

Zimmermann, M., Ziel, F., 2025. Efficient mid-term forecasting of hourly electricity load using generalized additive models. Applied Energy 388, 125444. doi:10.1016/j.apenergy.2025.125444.