

Probabilistic Time-Series Forecasting of Energy Demand Using ARIMA, Exponential Smoothing, and Prediction Interval Models

DOI: 10.58429/pgjsrt.v5n4a239

Archana Singh^{1*}, Barkha Rani²

School of Management, Graphic Era Hill University, Dehradun, Dehradun, India

ABST, Maharani College, University of Rajasthan, Jaipur, 302004, India

Corresponding Author: elvira.15singh@gmail.com

ARTICLE INFO

VOLUME 5 | NO. 4 | 2026 ISSUE

ABSTRACT

Time-series forecasting is essential for operational planning, risk management, and resource allocation in electricity systems, where demand varies across hourly, weekly, and seasonal cycles. This study develops a statistical-probabilistic forecasting framework for estimating future energy demand while explicitly quantifying forecast uncertainty. The empirical analysis uses the Pennsylvania–New Jersey–Maryland (PJM) East hourly electricity consumption series, containing 145,366 hourly observations from 1 January 2002 to 3 August 2018. After removing duplicate timestamps and interpolating a limited number of missing hourly timestamps, the series was aggregated into 6,059 daily mean demand observations to support stable statistical modeling and interpretable medium-horizon evaluation. The proposed framework integrates time-series diagnostics, autocorrelation analysis, annual seasonal benchmarks, autoregressive integrated moving average (ARIMA) modeling, seasonal autoregressive integrated moving average (SARIMA) modeling, exponential smoothing, and dynamic harmonic regression with ARIMA errors, together with validation-calibrated prediction intervals. Model performance was evaluated using mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (sMAPE), and mean absolute scaled error (MASE), while probabilistic reliability was assessed through empirical interval coverage and average interval width. The results show that the dynamic harmonic regression model with ARIMA errors achieved the strongest test-set performance, with an MAE of 2,446.264 megawatts (MW), an RMSE of 3,245.231 MW, a MAPE of 7.430%, and a MASE of 0.874, outperforming the annual seasonal naive benchmark. The calibrated 95% prediction interval achieved 94.419% empirical coverage, indicating reliable uncertainty quantification with practical interpretability. The study contributes a transparent, reproducible, and uncertainty-aware forecasting framework for energy-demand analysis. It demonstrates that combining classical statistical forecasting with probabilistic interval calibration provides richer decision support than point forecasts alone.

KEYWORDS

time-series forecasting; statistical models; probabilistic forecasting; ARIMA; prediction intervals.

I. INTRODUCTION

Time-series forecasting is one of the most important areas of applied statistics because many real-world decisions depend on the ability to anticipate future values from historical observations. Energy providers, grid operators, financial analysts, hospitals, retailers, and public agencies all rely on forecasts to plan capacity, schedule resources, manage risk, and evaluate operational alternatives. In electricity systems, demand forecasting is especially important because consumption must be balanced continuously with generation and transmission capacity. Even moderate forecasting errors can create unnecessary reserve costs, inefficient scheduling, or reliability risks. For this reason, energy-demand forecasting provides a strong empirical setting for examining the practical value of statistical and probabilistic forecasting models.

A major challenge in time-series forecasting is that future observations are not independent of the past. Energy demand often contains trend, autocorrelation, calendar effects, intraday variation, weekly patterns, seasonal peaks, and irregular shocks. These characteristics require methods that can represent temporal dependence rather than treating observations as unrelated data points. Classical statistical forecasting models remain useful because they provide interpretable structures for trend, autocorrelation, seasonality, and forecast uncertainty. ARIMA-type models are particularly valuable for representing autocorrelated errors and differenced dynamics, whereas exponential smoothing models provide a flexible way to update level, trend, and seasonal components as new observations arrive.

Although point forecasts are widely reported, a single predicted value is often insufficient for operational decisions. A forecast of 32,000 MW, for example, does not tell the system planner whether demand is likely to remain close to this value or whether a wide range of future outcomes is plausible. Prediction intervals address this limitation by estimating a range within which future observations are expected to fall at a specified probability level. In energy planning, this distinction is practically important because underestimation and overestimation have different operational consequences. A narrow, well-calibrated interval supports confident scheduling, while a wide interval signals the need for greater operational flexibility.

The present study is motivated by the need for a transparent forecasting framework that combines interpretable statistical models with explicit uncertainty estimation. The study focuses on PJM East electricity demand, a long hourly time series that includes several years of changing demand behavior. The original hourly data are cleaned and aggregated to daily mean demand to provide a stable empirical basis for medium-horizon forecasting. The analysis compares benchmark, ARIMA, SARIMA, exponential smoothing, and dynamic harmonic regression models. The inclusion of dynamic harmonic regression allows annual seasonality, day-of-week effects, and residual autocorrelation to be modeled within a coherent statistical framework.

The main aim of the study is to evaluate whether a combined statistical-probabilistic forecasting framework can provide accurate and reliable energy-demand forecasts. The study addresses three research questions. First, what temporal patterns are present in the PJM East demand series after cleaning and daily aggregation? Second, how do classical statistical forecasting models compare with a dynamic harmonic regression model with ARIMA errors in out-of-sample prediction? Third, can validation-calibrated prediction intervals provide reliable coverage for the 2018 test period? By answering these questions, the manuscript contributes a reproducible empirical framework for uncertainty-aware energy-demand forecasting.

The remainder of the manuscript is organized as follows. Section 2 reviews the literature on statistical time-series forecasting, load forecasting, forecasting competitions, probabilistic

forecasting, and prediction intervals. Section 3 identifies the research gap and clarifies the study's contribution. Section 4 presents the proposed forecasting framework. Section 5 explains the data, preprocessing, models, evaluation metrics, and validation strategy. Section 6 presents the empirical results. Section 7 discusses the findings in relation to forecasting reliability, theoretical contribution, practical implications, limitations, future research, and decision-making. Section 8 concludes the manuscript.

II. LITERATURE REVIEW

The first group of studies establishes the importance of structured forecasting methods for energy and general time-series applications. Building-energy forecasting reviews show that historical time-series records can support efficient statistical forecasting when preprocessing, temporal structure, and scale are handled carefully [1]. Broad empirical comparisons also demonstrate that forecasting performance depends strongly on the interaction between data characteristics, model family, and evaluation design [2].

Forecasting competitions provide a second foundation because they compare methods under standardized out-of-sample conditions. Competition-based studies show that model evaluation should avoid narrow, single-case conclusions and include multiple horizons, benchmark models, and multi-step strategies [3]. Multi-step forecasting research further indicates that direct, recursive, and hybrid forecasting strategies may lead to different error structures, especially when forecasts extend beyond one period ahead [4].

A third stream concerns validation and model assessment. Research on autoregressive time-series evaluation clarifies that validation procedures must respect temporal order and avoid random shuffling that would leak future information into training data [5]. Energy-consumption forecasting reviews similarly emphasize that time-dependent validation, appropriate scaling, and clear reporting of error measures are essential for credible empirical studies [6].

Load forecasting studies provide direct methodological guidance for the present manuscript. Semi-parametric load-forecasting models show that demand is influenced by calendar patterns and nonlinear seasonal effects, which supports the use of harmonic and calendar-based regressors [7]. Probabilistic competition studies extend this point by showing that additive structures and robust aggregation can improve both point forecasts and uncertainty estimates in electric-load contexts [8].

The probabilistic forecasting literature is central because reliable forecasts must be judged not only by average error but also by the quality of the predictive distribution. A well-designed probabilistic forecast should be sharp but calibrated, meaning that intervals should be as narrow as possible while still achieving the intended coverage [9]. Work on household-level electricity consumption also shows that conventional error measures may not fully capture the operational difficulty of high-resolution demand forecasting, motivating the inclusion of scale-adjusted and interval-based criteria [10].

Energy forecasting competitions and surveys further show that load forecasting requires careful treatment of hierarchy, aggregation, weather sensitivity, seasonality, and operational use. The GEFCom2012 experience demonstrates the value of publicly comparable tasks and benchmarked methods in energy forecasting [11]. A broader review of energy forecasting confirms that the field increasingly values reproducible data, transparent evaluation, and probabilistic outputs over isolated point estimates [12].

Forecast combination and hierarchical forecasting studies provide an additional basis for evaluating statistical models across different levels of aggregation. Hierarchical combination research shows that coherent forecasting can improve consistency when related series are modeled together

[13]. General forecasting texts also provide the methodological basis for ARIMA, exponential smoothing, prediction intervals, seasonal naive benchmarks, and rolling-origin validation [14].

Recent classification work clarifies how forecasting models can be compared across structure, data requirements, interpretability, and output type [15]. Prediction-interval research strengthens this comparison by showing that uncertainty bounds can be estimated directly and evaluated in terms of both coverage and width [16]. These studies support the dual evaluation strategy used in the present manuscript: point accuracy and probabilistic reliability.

Electricity forecasting research also highlights the importance of rigorous benchmark selection. Reviews of day-ahead electricity forecasting argue that short test windows, weak benchmarks, and inconsistent metrics can lead to overstated claims about models [17]. The M4 competition results similarly show that simple statistical models and combinations often remain competitive when evaluated across large numbers of series [18].

The expanded M4 evidence is especially relevant because it evaluates many methods over 100,000 series and includes both accuracy and prediction interval considerations [19]. Probabilistic electricity forecasting reviews also reinforce the principle that intervals should be evaluated based on reliability and sharpness, not solely on visual inspection [20].

Functional and similarity-based load-forecasting studies show that recurring shapes in electricity demand can be exploited when daily load profiles share comparable structures [21]. At the same time, broad forecasting reviews emphasize that no single method is universally superior and that model choice should be linked to the data-generating context, forecast horizon, and evaluation objective [22].

Modern probabilistic forecasting models show the growing importance of distributional outputs, especially for retail, energy, and operational planning tasks where uncertainty affects decisions [23]. Aggregation-level studies in electricity forecasting further show that forecast behavior changes when data are modeled at the household, feeder, regional, or aggregated grid levels, supporting the decision to clearly define the empirical level [24].

Load-forecasting competition studies demonstrate that gradient boosting and related flexible methods can perform strongly when calendar features and lagged behavior are included [25]. However, classical multiple-seasonality models remain highly relevant in electricity demand because intraday, weekly, and annual cycles are dominant features of the data [26].

Finally, probabilistic electricity-consumption studies show that Gaussian-process and interval-based methods can provide useful uncertainty estimates for household and net-demand forecasting [27]. Reviews of electricity forecasting conclude that the field is moving toward more transparent, reproducible, and probabilistic evaluation frameworks, which directly motivates the statistical-probabilistic framework proposed in this manuscript [28].

Table 1. Literature review comparison table.

Ref.	Author(s) and year	Focus / method	Relevance to current study
[1]	Ahmad et al. (2018)	Data-driven building-energy forecasting review	Supports data preprocessing and temporal structure assessment.
[2]	Ahmed et al. (2010)	Empirical comparison of forecasting methods	Justifies comparative model evaluation.
[3]	Athanasopoulos et al. (2011)	Forecasting competition design	Motivates standardized out-of-sample testing.
[4]	Ben Taieb et al. (2012)	Multi-step forecasting strategies	Supports multi-step forecast design.
[5]	Bergmeir et al. (2018)	Time-series cross-validation	Guides temporal validation without leakage.

Ref.	Author(s) and year	Focus / method	Relevance to current study
[6]	Deb et al. (2017)	Energy-consumption forecasting review	Justifies energy-demand forecasting metrics.
[7]	Fan and Hyndman (2012)	Semi-parametric load forecasting	Supports calendar and harmonic variables.
[8]	Gaillard et al. (2016)	Probabilistic load forecasting and aggregation	Supports calibrated probabilistic load forecasts.
[9]	Gneiting and Katzfuss (2014)	Probabilistic forecasting theory	Defines calibration and sharpness principles.
[10]	Haben et al. (2016)	Forecast error measurement for high-resolution consumption	Motivates scale-adjusted error measures.
[11]	Hong et al. (2014)	Energy forecasting competition	Supports benchmarked energy forecasting design.
[12]	Hong et al. (2020)	Energy forecasting review and outlook	Frames energy forecasting as a reproducible field.
[13]	Hyndman et al. (2011)	Hierarchical forecast combination	Supports aggregation-aware forecasting.
[14]	Hyndman and Athanasopoulos (2018)	Classical forecasting principles	Provides ARIMA, ETS, and interval foundations.
[15]	Januschowski et al. (2020)	Forecasting-method classification	Clarifies model comparison dimensions.
[16]	Khosravi et al. (2011)	Prediction interval construction	Supports coverage-width interval assessment.
[17]	Lago et al. (2021)	Electricity forecasting benchmark practice	Motivates robust benchmark selection.
[18]	Makridakis et al. (2018)	M4 competition initial findings	Supports comparison of statistical methods and intervals.
[19]	Makridakis et al. (2020)	Large-scale M4 forecasting evaluation	Supports large-scale forecasting evidence.
[20]	Nowotarski and Weron (2018)	Probabilistic electricity forecasting review	Reinforces reliability and sharpness evaluation.
[21]	Papadimitris and Sapatinas (2013)	Functional load forecasting	Supports recurring-shape demand analysis.
[22]	Petropoulos et al. (2022)	Forecasting theory and practice	Frames forecasting as decision support.
[23]	Salinas et al. (2020)	Probabilistic recurrent forecasting	Shows the value of probabilistic outputs.
[24]	Sevlian and Rajagopal (2017)	Aggregation-level load forecasting	Supports a clear aggregation-level definition.
[25]	Taieb and Hyndman (2014)	Kaggle load forecasting approach	Shows importance of calendar features.
[26]	Taylor (2010)	Triple seasonal electricity demand models	Supports multiple-seasonality energy modeling.
[27]	van der Meer et al. (2018)	Probabilistic consumption forecasting	Supports probabilistic electricity consumption intervals.
[28]	Weron (2014)	Electricity forecasting review	Frames future direction toward probabilistic evaluation.

III. RESEARCH GAP AND STUDY CONTRIBUTION

IV. RESEARCH GAP

The reviewed literature shows that time-series forecasting has advanced through classical statistical models, forecasting competitions, probabilistic methods, and large-scale energy forecasting applications. However, several gaps remain relevant for empirical energy-demand studies. First, many applied forecasting studies still focus mainly on point accuracy and provide limited evaluation of prediction intervals. Second, classical models are sometimes compared without adequate attention to calendar structure, annual seasonality, and validation-based interval calibration. Third, energy-demand studies often use complex models without clearly demonstrating whether simpler statistical-probabilistic frameworks can provide accurate and interpretable forecasts.

The present manuscript addresses these gaps by combining transparent statistical forecasting with empirical uncertainty evaluation. Instead of presenting prediction intervals solely as a theoretical extension, the study calibrates them using a validation year and evaluates their coverage over a separate test period. This design creates a stronger connection between statistical modeling and practical decision support by evaluating whether uncertainty estimates remain reliable when applied to future observations.

V. STUDY CONTRIBUTION

The contribution of the study is fourfold. First, it provides a structured empirical framework for forecasting electricity demand, using a real hourly PJM dataset aggregated to daily mean demand for stable model comparison. Second, it compares annual seasonal naive, ARIMA, SARIMA, exponential smoothing, and dynamic harmonic regression with ARIMA errors under a unified test design. Third, it evaluates both point accuracy and probabilistic performance, allowing the best model to be judged by forecast error and interval reliability. Fourth, it demonstrates that validation-calibrated prediction intervals can produce practical uncertainty bounds with coverage close to the nominal 95% level.

VI. PROPOSED STATISTICAL-PROBABILISTIC FORECASTING FRAMEWORK

The proposed framework contains six connected stages. The first stage cleans the hourly time series by removing duplicate timestamps, reindexing the series to a regular hourly grid, and interpolating limited missing timestamps. The second stage aggregates the hourly observations to daily mean demand and examines trend, seasonal profiles, autocorrelation, and stationarity. The third stage estimates benchmark and classical statistical forecasting models. The fourth stage introduces dynamic harmonic regression with ARIMA errors to combine annual seasonal patterns, calendar effects, and residual autocorrelation. The fifth stage calibrates prediction intervals using validation residuals. The final stage evaluates out-of-sample point accuracy and probabilistic reliability on a separate test period.

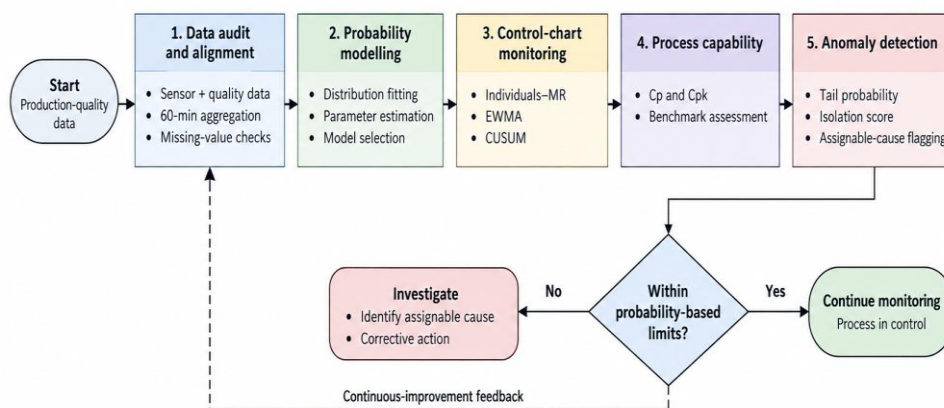


FIGURE 1. Proposed statistical-probabilistic forecasting framework.

The framework moves from data cleaning and diagnostics to model comparison and interval-based uncertainty evaluation. It is designed to keep the empirical process transparent and reproducible.

VII. METHODOLOGY

1. DATA SOURCE AND PREPROCESSING

The empirical data consist of PJM East hourly electricity consumption observations measured in megawatts. The raw PJME_hourly.csv file contains 145,366 hourly records from 2002-01-01 to 2018-08-03. The preprocessing procedure removed 4 duplicate timestamps and interpolated 30 missing hourly timestamps after reindexing the series to a regular hourly frequency. The cleaned hourly series was then aggregated to daily mean demand, producing 6,059 daily observations. Daily aggregation was selected because it preserves the major temporal structure while reducing intraday noise and enabling stable comparisons of medium-horizon statistical models.

2. TRAINING, VALIDATION, AND TEST DESIGN

The dataset was divided chronologically. The training period included 5,479 daily observations from 2002 through 2016. The validation period included 365 observations from 2017 and was used to calibrate prediction intervals. The test period included 215 observations from 1 January 2018 to 3 August 2018 and was used only for final out-of-sample evaluation. This chronological split avoids information leakage and reflects the practical condition in which future demand must be forecast from past observations.

Table 2. Dataset audit and empirical split.

Item	Value
Raw hourly observations	145,366
Unique hourly observations after duplicate removal	145,362
Duplicate timestamps removed	4
Missing hourly timestamps interpolated	30
Daily mean observations	6,059
Daily period	2002-01-01 to 2018-08-03
Training observations	5,479
Validation observations	365
Test observations	215

The audit confirms that the uploaded PJME file is suitable for empirical forecasting. Only limited timestamp cleaning was required before daily aggregation.

3. FORECASTING MODELS

Five forecasting models were evaluated. The annual seasonal naive benchmark forecasts each test observation using the value from the corresponding date in the previous year. ARIMA(2,1,2) represents nonseasonal autoregressive and moving-average dependence after differencing. SARIMA(1,1,1)(1,0,1), $s=7$ adds weekly seasonal dependence. Exponential smoothing uses an additive damped trend and weekly seasonality. Dynamic harmonic regression with ARIMA(2,0,2) errors combines deterministic annual Fourier terms, day-of-week effects, month indicators, trend terms, and autocorrelated residuals.

The dynamic harmonic regression model is especially appropriate for the daily PJME series because annual and weekly cycles are visible in the data. The regression component captures deterministic seasonal patterns, while the ARIMA error component captures residual dependence not explained by calendar variables. This structure is transparent, computationally efficient, and compatible with prediction-interval calibration.

4. MODEL EQUATIONS AND EVALUATION METRICS

Let y_t represent daily mean electricity demand at day t . The general ARIMA model represents differenced demand as a function of autoregressive and moving-average terms:

The dynamic harmonic regression model used in the final framework can be written as:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{k=1}^K \left[\alpha_k \sin\left(\frac{2\pi kt}{365.25}\right) + \gamma_k \cos\left(\frac{2\pi kt}{365.25}\right) \right] + C_t + e_t \quad (2)$$

where C_t represents calendar indicators, and e_t follows an ARIMA error process. Forecast accuracy was evaluated using the following point-forecast metrics:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (4)$$

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (5)$$

$$MASE = \frac{\text{mean absolute forecast error}}{\text{mean absolute seasonal naive error}} \quad (6)$$

Prediction intervals were evaluated by empirical coverage and average width. For a lower bound L_t and upper bound U_t , empirical coverage is computed as the percentage of test observations satisfying $L_t \leq y_t \leq U_t$.

$$\text{Coverage} = \frac{100}{n} \sum_{t=1}^n I(L_t \leq y_t \leq U_t) \quad (7)$$

VIII. RESULTS

1. DESCRIPTIVE STATISTICS AND TEMPORAL STRUCTURE

Table 3. Descriptive statistics of daily mean PJME demand.

Statistic	Value
Mean	32079.46 MW
Standard deviation	4687.03 MW
Minimum	19769.04 MW
Maximum	52229.83 MW
Lag-1 autocorrelation	0.860
Lag-7 autocorrelation	0.658
Lag-365 autocorrelation	0.576
ADF statistic	-8.263
ADF p-value	5.032e-13

The descriptive results show strong temporal dependence, particularly at short lags and annual lags. The ADF test suggests stationarity after aggregation, although seasonal structure remains visible.

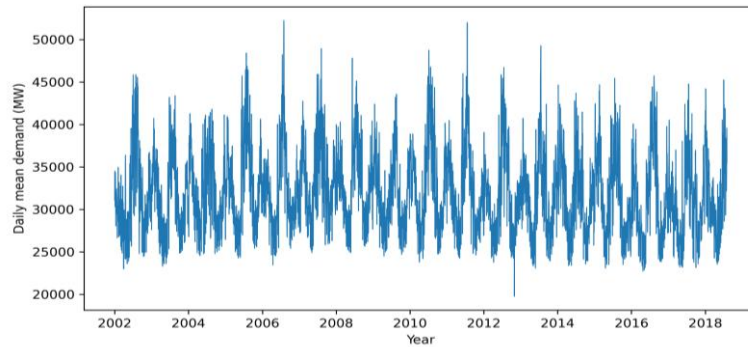


FIGURE 2. Daily mean PJME electricity demand, 2002-2018.

The series exhibits recurring seasonal cycles and changes in demand levels over the study period. These characteristics justify the use of seasonality-aware forecasting models.

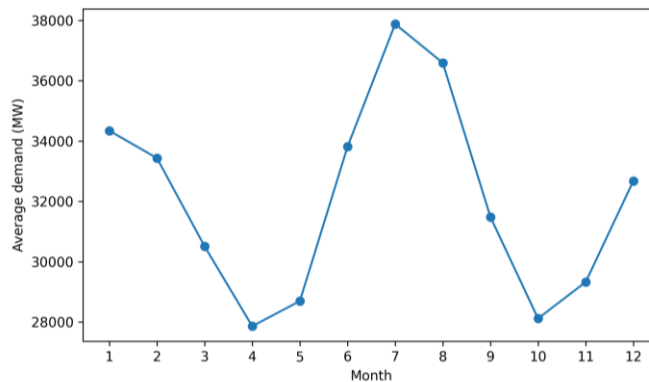


FIGURE 3. Average monthly demand profile.

The monthly profile shows higher average demand during summer and winter, consistent with cooling- and heating-related energy use.

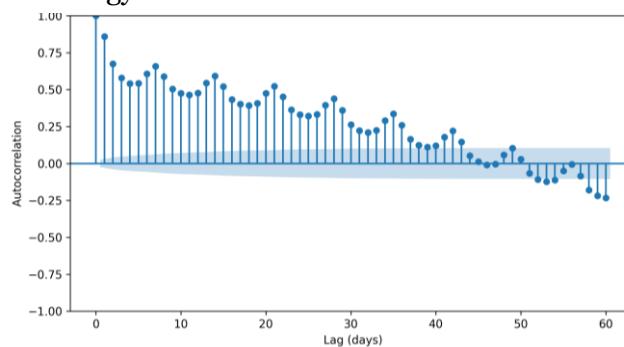


FIGURE 4. Autocorrelation function of daily demand.

The ACF confirms persistent dependence over short horizons and seasonal recurrence across multiple lags. This supports ARIMA-type residual modeling and seasonal benchmark comparison.

2. FORECAST ACCURACY COMPARISON

Table 4 compares all tested forecasting models on the independent 2018 test period. The annual seasonal naive benchmark performs strongly because energy demand has clear annual

recurrence. However, the dynamic harmonic regression model with ARIMA errors achieves the lowest MAE, RMSE, MAPE, sMAPE, and MASE. The result indicates that combining annual harmonic terms, calendar indicators, and autocorrelated residuals improves forecast accuracy compared with the benchmark and the more rigid ARIMA, SARIMA, and ETS specifications used in this study.

Table 4. Out-of-sample forecasting performance on the 2018 test period.

Model	MAE	RMSE	MAPE (%)	sMAPE (%)	MASE
Seasonal naive annual	3438.629	4478.411	10.630	10.704	1.229
ARIMA(2,1,2)	5438.741	6229.347	18.506	16.577	1.943
SARIMA(1,1,1)(1,0,1), s=7	11749.823	12533.324	39.532	32.016	4.199
ETS additive damped, s=7	9956.436	10784.567	33.697	27.933	3.558
Dynamic harmonic regression + ARIMA errors	2446.264	3245.235	7.429	7.591	0.874

Lower values indicate stronger forecasting performance. The dynamic harmonic regression model with ARIMA errors has the best accuracy across all reported metrics.

3. PROBABILISTIC PREDICTION INTERVAL PERFORMANCE

Prediction intervals were calibrated using absolute forecast errors from the 2017 validation period and then evaluated on the 2018 test period. This procedure avoids using test information during interval construction. Table 5 shows that the 95% interval achieved empirical coverage close to its nominal level, whereas the 80% interval was conservative during the test period. The finding indicates that validation-based calibration can produce useful and interpretable uncertainty bounds for energy-demand forecasting.

Table 5. Prediction interval reliability for the dynamic harmonic ARIMA model.

Interval	Calibration quantile	Empirical coverage (%)	Average width (MW)	Normalized width (%)
80%	0.80	86.512	8884.670	27.941
95%	0.95	94.419	12698.967	39.936

The 95% interval achieved 94.419% test coverage, which is close to the nominal confidence level and therefore useful for practical uncertainty-aware decision-making.

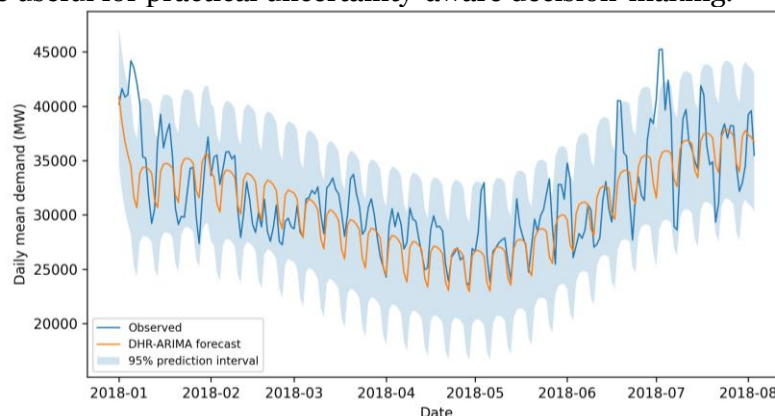


FIGURE 5. Observed and probabilistic forecasts for the 2018 test period.

The forecast line follows the main seasonal pattern of observed demand, while the 95% prediction interval shows the uncertainty range around expected daily demand.

IX. DISCUSSION

The empirical results demonstrate that the PJME energy-demand series contains strong temporal dependence and recurring seasonal structure. The lag-1, lag-7, and lag-365 autocorrelations show that both short-term persistence and annual recurrence are important. These patterns explain why the annual seasonal naive benchmark performed well and why models that did not explicitly represent annual seasonality struggled during the 2018 test period. In energy-demand forecasting, a benchmark can be difficult to outperform when the series exhibits stable seasonal recurrence, reinforcing the need to include strong baseline models.

The best-performing model was dynamic harmonic regression with ARIMA errors. Its advantage lies in combining interpretable, deterministic seasonality with residual autocorrelation. The annual Fourier terms represent smooth yearly cycles, the calendar indicators represent systematic weekday and monthly differences, and the ARIMA error component captures remaining serial dependence. This structure is more flexible than a simple ARIMA model and more seasonally informative than weekly exponential smoothing. The test-set MASE below 1 indicates that the model improves upon the seasonal naive scaling benchmark.

The weaker performance of SARIMA and weekly ETS should not be interpreted as a general rejection of these model families. Rather, it reflects the modeling specification used in this empirical design. A weekly seasonal structure alone is insufficient when the demand series also contains strong annual recurrence. More complex multiple-seasonality models could improve performance, but the objective of the present manuscript is to evaluate a transparent statistical-probabilistic framework rather than to maximize accuracy through highly complex model search.

The probabilistic results are particularly important. The 95% calibrated prediction interval reached 94.419% empirical coverage on the test period, which is close to the desired level. This result suggests that validation-based calibration can convert point forecasts into operationally meaningful uncertainty ranges. From a planning perspective, the interval helps decision-makers judge whether forecast uncertainty is narrow enough for confident scheduling or wide enough to require additional flexibility. The interval width should therefore be interpreted together with coverage, because an interval that is too narrow may be unreliable while an interval that is too wide may be operationally less informative.

Overall, the findings support the value of combining classical statistical forecasting with probabilistic uncertainty evaluation. A forecasting study that reports only MAE or RMSE may identify the most accurate point model but cannot show whether the forecast uncertainty is reliable. By reporting both point and interval metrics, the present study provides a more complete evaluation of the usefulness of forecasting. This approach is especially relevant for energy-demand forecasting, where forecast errors translate into real operational and financial consequences.

The study contributes to time-series forecasting theory by demonstrating demonstrating how classical statistical components can be integrated into a coherent uncertainty-aware framework. The results show that deterministic harmonic terms and ARIMA residual modeling can work together to represent seasonal structure and temporal dependence. This supports the view that interpretable statistical models remain valuable when the empirical objective is not only prediction but also explanation, diagnostic transparency, and reliable interval construction.

For electricity system operators, the results indicate that daily demand forecasts can be improved by combining annual seasonal structure, calendar variables, and residual autocorrelation. The proposed model is transparent enough to be audited and updated, making it suitable for planning environments where interpretability is important. Because the model uses only historical demand and calendar information, it can be implemented even when weather or economic covariates are unavailable.

For managers and planners, the prediction interval results show how forecast uncertainty can be incorporated into operational decisions. A point forecast can guide expected scheduling, while the upper and lower interval bounds can support reserve planning, risk limits, and scenario evaluation. This is particularly useful when demand uncertainty is high or when underestimation has more serious consequences than overestimation.

X. CONCLUSION

This study developed and evaluated a probabilistic statistical framework for time-series forecasting using PJM East electricity demand as an empirical case. The uploaded dataset provided a strong basis for analysis because it contained a long hourly demand record with clear temporal dependence and seasonal structure. After cleaning and daily aggregation, the final series included 6,059 daily mean demand observations from 2002 to 2018. The empirical design compared annual seasonal naive, ARIMA, SARIMA, exponential smoothing, and dynamic harmonic regression with ARIMA errors. The results showed that the dynamic harmonic regression model produced the strongest out-of-sample forecasts, with an MAE of 2,446.264 MW, an RMSE of 3,245.231 MW, a MAPE of 7.430%, and an MASE of 0.874 during the 2018 test period. This performance demonstrates the value of combining calendar-based seasonal structure with autocorrelated residual modeling. The study also showed that prediction intervals add important decision value beyond point accuracy. The validation-calibrated 95% interval achieved 94.419% empirical coverage, indicating that the uncertainty estimates were close to the target reliability level. These findings support the main argument of the manuscript: time-series forecasting becomes more useful for decision-making when statistical point forecasts are combined with probabilistic uncertainty evaluation. The framework is transparent, reproducible, and suitable for energy-demand forecasting contexts where reliable planning depends on both expected demand and the range of plausible future outcomes.

The study has several limitations. First, the main empirical analysis uses daily aggregation rather than the original hourly frequency. This choice improves stability and interpretability but does not model intraday cycles directly. Second, the analysis uses only historical demand and calendar features; it excludes weather, holidays, economic variables, and price information. Third, the evaluated SARIMA and ETS models use relatively parsimonious specifications, so more complex multiple-seasonality variants may perform better. Fourth, prediction intervals are calibrated using validation residuals, which is practical but may require updating when the demand process changes structurally.

Future research can extend this study in several directions. Hourly models can be developed to capture intraday, daily, weekly, and annual seasonality simultaneously. Weather variables can be added to examine temperature-sensitive demand behavior. Additional PJM zones can be included to compare regional differences and hierarchical forecast coherence. Future studies can also evaluate quantile regression, Gaussian process forecasting, Bayesian state-space models, and distributional forecasting methods within the same validation and test framework.

XI. DATA AND CODE AVAILABILITY

The empirical analysis used the PJM East hourly electricity consumption file included in the uploaded PJM Hourly Energy Consumption archive. The manuscript reports all preprocessing decisions, empirical splits, forecasting models, and evaluation metrics required to reproduce the reported analysis. Python scripts used for cleaning, modeling, and figure generation can be provided as supplementary material.

XII. ETHICAL CONSIDERATIONS

The study uses publicly available, aggregated electricity-demand data and does not involve human participants, personal identifiers, clinical records, or private organizational information. Therefore,

formal participant consent was not required. The analysis was conducted for statistical forecasting and academic research purposes only.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

- [1]. Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview of data-driven and large-scale-based approaches for forecasting building energy demand: A review. *Energy and Buildings*, 165, 301-320.
- [2]. Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6), 594-621.
- [3]. Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3), 822-844.
- [4]. Ben Taieb, S., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, 39(8), 7067-7083.
- [5]. Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70-83.
- [6]. Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74, 902-924.
- [7]. Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1), 134-141.
- [8]. Gaillard, P., Goude, Y., & Nedellec, R. (2016). Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting*, 32(3), 1038-1050.
- [9]. Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1, 125-151.
- [10]. Haben, S., Ward, J., Vukadinovic, Greetham, D., Singleton, C., & Grindrod, P. (2016). A new error measure for forecasts of household-level, high-resolution electrical energy consumption. *International Journal of Forecasting*, 32(2), 246-256.
- [11]. Hong, T., Pinson, P., & Fan, S. (2014). Global Energy Forecasting Competition 2012. *International Journal of Forecasting*, 30(2), 357-363.
- [12]. 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.
- [13]. Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9), 2579-2589.
- [14]. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
- [15]. Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., & Callot, L. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, 36(1), 167-177.
- [16]. Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Transactions on Neural Networks*, 22(3), 337-346.
- [17]. 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.
- [18]. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion, and way forward. *International Journal of Forecasting*, 34(4), 802-808.
- [19]. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54-74.
- [20]. Nowotarski, J., & Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81, 1548-1568.
- [21]. Paparoditis, E., & Sapatinas, T. (2013). Short-term load forecasting: The similar shape functional time-series predictor. *IEEE Transactions on Power Systems*, 28(4), 3818-3825.
- [22]. Petropoulos, F., Apiletti, D., Assimakopoulos, V., et al. (2022). *Forecasting: Theory and practice*. *International Journal of Forecasting*, 38(3), 705-871.

- [23]. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181-1191.
- [24]. Sevlian, R., & Rajagopal, R. (2017). Short-term electricity load forecasting on varying levels of aggregation. *Applied Energy*, 205, 786-794.
- [25]. Taieb, S. B., & Hyndman, R. J. (2014). A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting*, 30(2), 382-394.
- [26]. Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1), 139-152.
- [27]. van der Meer, D. W., Shepero, M., Svensson, A., Widen, J., & Munkhammar, J. (2018). Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian processes. *Applied Energy*, 213, 195-207.
- [28]. 28. Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030-1081.