Probabilistic Energy Balance Forecasting in Rural Grids using Conformal Prediction

César Felipe Lozano-Sánchez de la Morena
Dept. of Signal Processing and Communications
University of Alcalá
Alcalá de Henares, Spain
cesar.lozano@uah.es

Carlos Santos-Pérez
Dept. of Signal Processing and Communications
University of Alcalá
Alcalá de Henares, Spain
carlos.santos@uah.es

Jorge Pérez-Aracil
Dept. of Signal Processing and Communications
University of Alcalá
Alcalá de Henares, Spain
jorge.perezaracil@uah.es

Francisco Javier Rodríguez-Sánchez
Dept. of Electronics
University of Alcalá
Alcalá de Henares, Spain
franciscoj.rodriguez@uah.es

Silvia Jiménez-Fernández
Dept. of Signal Processing and Communications
University of Alcalá
Alcalá de Henares, Spain
silvia.jimenez@uah.es

Alejandro Martín-Zamora
Dept. of Signal Processing and Communications
University of Alcalá
Alcalá de Henares, Spain
a.martinz@uah.es

Reinier Herrera-Casanova
Faculty of Mechanical and Electrical Engineering
Autonomous University of Nuevo León
San Nicolás de los Garza, México
herreracasanovareinier@gmail.com

Arturo Conde
Faculty of Mechanical and Electrical Engineering
Autonomous University of Nuevo León
San Nicolás de los Garza, México
con_de@yahoo.com

***Abstract*— The growing integration of distributed energy resources demands improved forecasting methods for effective grid management. This paper applies conformal prediction (CP) techniques, enhanced with dynamic sliding windows, combined with deep learning models, to produce probabilistic forecasts of solar energy generation, energy demand, and net energy balance over a 36-hour horizon. The methodology is evaluated on a representative rural case study using publicly available datasets (ERA5 and DATADIS). A comparative analysis shows that the proposed CP framework statistically outperforms quantile regression (QR) across all forecasting targets, delivering well-calibrated prediction intervals with lower normalized interval scores. Computational profiling further confirms that the entire pipeline operates with sub-40 ms latency per prediction sequence, validating its suitability for real-time Virtual Power Plant control. These results indicate that CP is a viable and efficient framework for uncertainty quantification in energy planning and renewable integration.**

***Index Terms*—Conformal Prediction, Solar Generation Forecasting, Demand Forecasting, Smart Grids, Machine Learning**

I. Introduction

The global energy landscape is undergoing a significant transformation driven by the critical need to integrate distributed energy resources, such as solar and wind power. While these renewable sources are essential for decarbonization, their inherent variability poses substantial challenges to grid stability and efficient energy management [1]–[4]. The proliferation of distributed generation can lead to operational instabilities, energy curtailment, and volatile market prices, affecting electrical network reliability across all scales, from large transmission systems to local smart grids and Virtual Power Plants (VPPs) [5]. Consequently, effective energy management in this evolving paradigm demands advanced forecasting tools. Traditional point forecasts, which provide only single expected values, are often insufficient for robust decision-making under such uncertainty. Probabilistic forecasts, which quantify uncertainty through prediction intervals or full predictive distributions, provide a more reliable foundation for critical operations, including unit commitment [6], reserve management, storage optimization, and energy market participation [7].

This paper addresses the challenge of enhancing energy management by proposing and evaluating conformal prediction (CP) techniques to generate probabilistic forecasts for solar energy generation, energy demand, and net energy balance over a 36-hour horizon. The central hypothesis is that augmenting machine learning (ML) models with CP and dynamic sliding calibration windows, and leveraging publicly available meteorological and energy consumption data, yields well-calibrated prediction intervals capable of handling the non-stationarity and intermittency inherent to renewable-rich systems. This approach targets improved energy management efficiency in decentralized systems, exemplified by the rural municipality of Uceda, Spain, which has high solar penetration potential and is considered here within a future VPP scenario with a 3 MWp photovoltaic capacity. The application of CP is motivated by its ability to provide statistically rigorous prediction intervals with finite samples and without strong distributional assumptions, making it suitable for operational energy forecasting [8]–[11].

The main contributions of this work are: (i) the development and application of CP models with dynamic sliding windows for hourly probabilistic forecasting of solar generation and energy demand; (ii) the proposal and evaluation of a direct probabilistic forecasting model for the net energy balance, also leveraging CP, after an initial investigation revealed limitations in an indirect, derived approach due to error propagation; (iii) a comparative analysis demonstrating the statistical advantage of the CP framework over state-of-the-art QR methods, validated through the Diebold–Mariano test [29]; and (iv) an assessment of the computational complexity of the proposed pipeline, confirming its suitability for real-time applications within VPPs.

II. Methodology

A. Data Acquisition, Preprocessing, and Feature Engineering

The study utilizes historical hourly data (March 2022–March 2025) from the municipality of Uceda, Guadalajara, comprising energy demand records from the Spanish DATADIS platform and meteorological variables from the widely validated ERA5 reanalysis dataset [14]–[16].

A systematic data processing and feature engineering pipeline, depicted in Figure 1, was implemented to prepare the datasets for subsequent modeling.

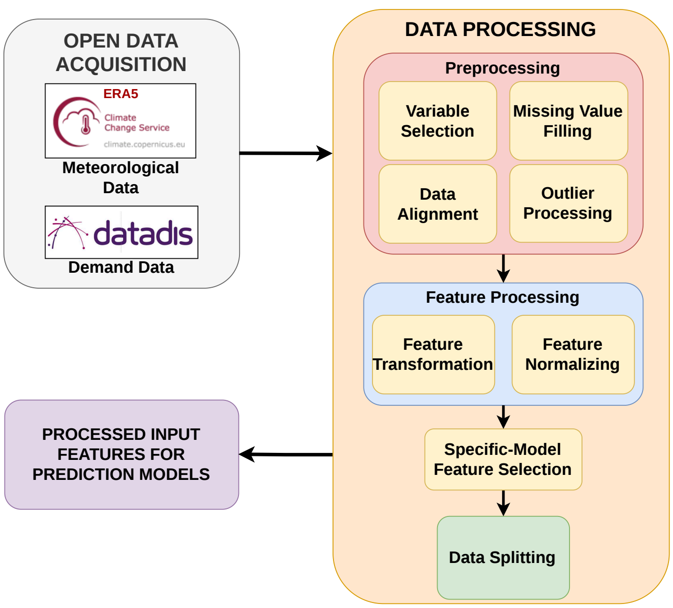

Fig. 1: Overall Data Processing and Feature Engineering Pipeline. This diagram outlines the key stages from raw data acquisition to the generation of processed features ready for input into prediction models.

Initial data cleaning involved linear interpolation of anomalous zero values in the energy demand series using the means of adjacent non-zero values. Feature engineering included creating lagged variables for energy demand and Surface Solar Radiation Downwards (SSRD), with lag selection guided by Autocorrelation (ACF) and Partial Autocorrelation (PACF) analysis of the target series to capture temporal dependencies. Calendar features, including binary indicators for public holidays and weekends, as well as cyclical time variables (Hour, Day, Month, Season), were also included. Subsequent exploratory data analysis guided the assessment of feature relevance. Pearson correlation analysis identified variables with low correlation to the targets (demand and SSRD) for potential exclusion. Principal Component Analysis was used not for transformation, but to assess contributions of the original features to dataset variance, thereby identifying less influential variables. This analytical process facilitated a targeted dimensionality reduction, retaining only the most salient predictors for each model. Three distinct, optimized datasets were prepared for generation forecasting, demand, and net energy balance. The specific input features are detailed in their respective model subsections.

B. Conformal Prediction for Probabilistic Forecasting

CP was selected for its ability to generate prediction intervals with finite-sample coverage guarantees without requiring specific distributional assumptions, rendering it suitable for complex energy forecasting tasks where error distributions are often unknown [12], [13]. CP can be overlaid onto any base model to yield statistically rigorous intervals for new data, assuming only data exchangeability [8].

In this study, we set the significance level to $\varepsilon = 0.05$, yielding a 95% coverage probability for the prediction intervals. We employ the standard (or split) CP regression framework, in which the non-conformity measure (NCM) is defined as the absolute error of the point forecasts from a base model.

The core concept involves calculating this NCM, $\alpha_i = |y_i - \hat{y}_i|$, for each data point in a calibration set. Subsequently, the empirical quantile from these non-conformity scores, $(1-\varepsilon)(1+1/N_{cal})$, where N_{cal} is the size of the calibration set, is determined. This quantile serves as a threshold, q_ε, which then defines the prediction interval for a new instance as $[\hat{y}_{new} - q_\varepsilon, \hat{y}_{new} + q_\varepsilon]$. To handle non-stationarity in the time series, we adopt a sliding calibration window that is updated as new observations become available. The window size was optimized per target via grid search over candidates ranging from one week to one month, selecting the value that minimized the MIS on a held-out validation set. The resulting optimal sizes were 504 samples (21 days) for generation, 744 samples (31 days) for demand, and 336 samples (14 days) for balance. This approach yielded consistently higher PICP values than a fixed global calibration set, supporting the benefit of adapting the calibration distribution to the local data regime.

$$\alpha_i(h) = |y_i(h) - \hat{y}_i(h)|, \quad i \in \mathcal{I}_{\text{cal}},\ h = 1, \ldots, H.$$

$$k_h = \lceil (1-\varepsilon)\,(N_{\text{cal}}+1) \rceil, \qquad \hat{q}_\varepsilon(h) = \alpha_{(k_h)}(h),$$

$$\Gamma_\varepsilon(x_j, h) = \big[\, \hat{y}_j(h) - \hat{q}_\varepsilon(h)\,,\ \hat{y}_j(h) + \hat{q}_\varepsilon(h) \,\big].$$

C. Probabilistic Generation Forecasting Model

For the solar power generation forecasting model, the selected input features include: Hour, Month, Season, 2-meter dewpoint temperature, 2-meter ambient temperature, the 10-meter U- and V-components of wind, total cloud cover, Top-of-Atmosphere incident solar radiation (TISR), lagged SSRD values at 1, 2, and 24 hours prior, and the target variable, SSRD.

To develop the generation forecasting model, we performed grid search on several ML models and applied Bayesian optimization to deep learning (DL) candidates to identify the best predictor of solar radiation [17], [18]. The ML models evaluated were: linear regression, ridge, lasso, elastic net, K-nearest neighbors, decision trees, random forests, XGBoost, and LightGBM. The DL architectures included: multilayer perceptron, convolutional neural network, gated recurrent unit (GRU), and long short-term memory (LSTM). We selected these models based on their reported performance in prior work [19]–[23]. Following this systematic evaluation, the LSTM network outperformed all other candidates and was selected as the base forecaster for solar radiation.

Subsequently, the standard CP method detailed in Section II-B was applied as a wrapper to the LSTM point forecasts using the absolute error as the NCM. The resulting prediction intervals depend on both the forecast generation time and the target horizon, widening as the forecast lead time increases, as illustrated in Fig. 2. We chose to predict SSRD directly rather than photovoltaic (PV) power, thereby decoupling the irradiance model from installation specifics and allowing flexible adjustment of peak installed capacity in future analyses. The predicted SSRD and its CP intervals were then converted to photovoltaic energy (kWh) via the Erbs model, a physical decomposition that uses TISR and SSRD together with installation-specific parameters [24].

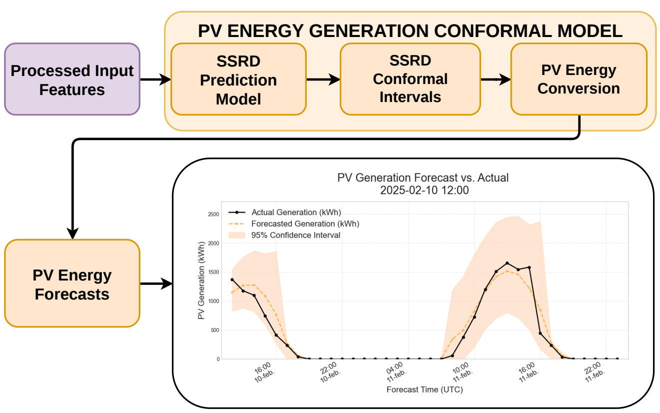

Fig. 2: Probabilistic PV Generation Forecasting Model Architecture, utilizing an LSTM network, CP, and Erbs model conversion, illustrating the widening of prediction intervals with the forecast horizon.

D. Probabilistic Demand Forecasting Model

For the energy demand forecasting model, the final curated set of input features comprises: Hour, Month, binary indicators for public holidays and weekends, Season, 2-meter ambient temperature, total cloud cover, SSRD, and lagged demand values at 1, 24, 48, 72, and 168 hours prior, in addition to the target variable of total energy demand.

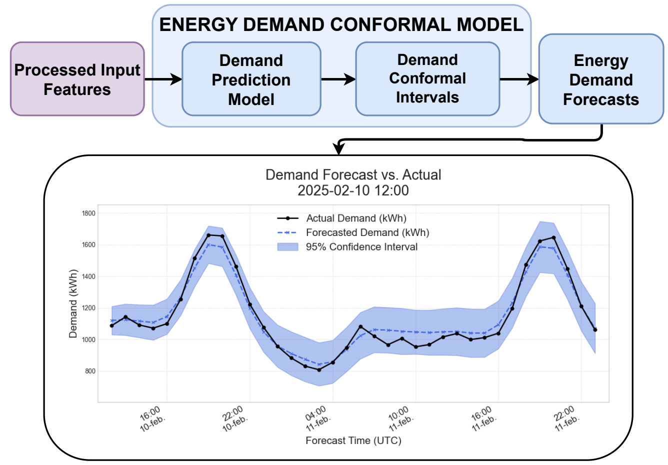

Fig. 3: Probabilistic Demand Forecasting Model Architecture, employing an LSTM network and CP, illustrating the widening of prediction intervals with the forecast horizon.

We employed the same hyperparameter optimization procedure with identical candidate models to develop the demand forecasting model. Following metric-based evaluation, the LSTM network was again identified as the best-performing architecture. We subsequently established prediction intervals for energy demand over the next 36 hours using the CP framework described in Section II-B, with the absolute error serving as the NCM (see Fig. 3).

E. Probabilistic Energy Balance Model

The direct probabilistic energy balance forecasting model employs a feature set derived from the inputs considered for the generation and demand models. The final selection includes: Hour, Month, binary indicators for public holidays and weekends, Season, 2-meter ambient temperature, and total cloud cover, supplemented by lagged energy balance values from 1, 2, and 24 hours prior. This direct modeling strategy was chosen because an initial investigation into an indirect method, which involved calculating the balance from $G_t - D_t$ forecasts, revealed that error propagation led to CP intervals of excessive width.

We mirrored the hyperparameter optimization strategy used previously for the direct energy balance model, applying grid search to the same set of ML models and Bayesian optimization to the DL candidates. Following this systematic evaluation, the GRU network emerged as the most accurate forecaster for the energy balance over the 36-hour horizon. We subsequently constructed prediction intervals using the CP technique with the absolute error as the NCM, as detailed in Section II-B; the resulting lower and upper bounds are illustrated in Fig. 4.

F. Evaluation Metrics

We assessed the performance of the probabilistic forecasting models and their prediction intervals using established metrics [25]. Reliability was evaluated via the Prediction Interval Coverage Probability (PICP), which measures the proportion of

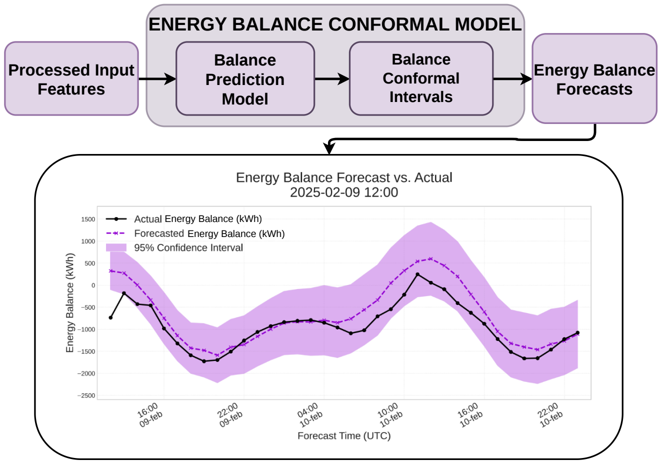

Fig. 4: Direct Probabilistic Energy Balance Model Architecture, based on a GRU network and CP, illustrating the widening of prediction intervals with the forecast horizon.

true values within the predicted bounds, ideally matching the nominal confidence level. Interval sharpness was quantified via the Prediction Interval Normalized Average Width (PINAW), where lower values indicate more informative intervals, assuming adequate PICP. The overall quality of the predictive distributions was assessed using the Continuous Ranked Probability Score (CRPS), with lower scores signifying better probabilistic accuracy [26]. We also report the logarithmic score (LogScore), which evaluates the sharpness of the predictive distribution. For point forecast comparison, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were calculated. A Daily Seasonal Naive Persistence (DSP) model, forecasting based on the value observed 24 hours prior, served as a benchmark justified by the strong daily seasonality identified in ACF and PACF analyses. Additionally, for the comparative analysis against QR [30], the Mean Interval Score (MIS), defined as the Winkler Score [26] normalized by the target range, was adopted as the primary evaluation metric, as it penalizes both excessive interval width and lack of coverage while enabling cross-target comparison. Statistical significance of performance differences was assessed using the Diebold–Mariano test [29] across all 36 forecasting horizons.

III. Results and Discussion

The evaluation employs the metrics defined in Section II-F, with a nominal confidence level of 95% for all prediction intervals. All models, including the DSP baseline, were augmented with CP. Percentage improvements are calculated with respect to this DSP baseline, chosen because of the pronounced daily seasonality and autocorrelation observed in the target variables.

A. Generation Forecasting Model Performance

For SSRD forecasting, which serves as a proxy for PV generation, the CP-augmented LSTM network outperformed the DSP baseline. The proposed model achieved an average RMSE of 56.84 W/m^2 and an MAE of 26.90 W/m^2, corresponding to improvements of 22.82% and 13.27% over the DSP model's 73.64 W/m^2 RMSE and 31.01 W/m^2 MAE. The CRPS was 30.03 W/m^2, a 24.04% improvement over the DSP's 39.54 W/m^2. Under the sliding window evaluation, the model attained an average PICP of 0.940 with a PINAW of 0.319. The LogScore improved by 5.40%, from 5.8258 to 5.5112. Following standard practice in solar forecasting, these CP metrics were computed exclusively on diurnal hours.

B. Demand Forecasting Model Performance

The proposed CP-augmented LSTM network for energy demand forecasting demonstrated improvements over the CP-augmented DSP baseline. For point forecast accuracy, the proposed model achieved an average RMSE of 129.39 kWh and an MAE of 89.54 kWh, representing improvements of 20.47% and 15.07%, respectively, over the DSP model's 162.69 kWh RMSE and 105.43 kWh MAE. The CRPS improved by 16.63%, from 83.20 kWh to 69.36 kWh. Under the sliding window evaluation, the model attained an average PICP of 0.923 with a PINAW of 0.298. The LogScore was 6.8806, a marginal increase of 2.27% compared to the DSP's 6.7279.

C. Energy Balance Model Performance

For energy balance forecasting, a CP-augmented GRU network was developed and compared against the CP-augmented DSP baseline. The proposed model achieved an average RMSE of 379.58 kWh, an 8.27% improvement over the DSP's 413.80 kWh. However, the MAE of 246.41 kWh was 12.81% higher than the DSP's 218.43 kWh. This expected behavior occurs because DL architectures tend to predict a smoothed conditional mean, effectively minimizing larger squared errors (RMSE), whereas naive DSP directly replicates the previous day's structural noise, inherently favoring purely absolute error metrics (MAE) in highly seasonal series. Despite this higher point MAE, the GRU model yields a superior predictive distribution: the CRPS improved by 7.38%, from 212.24 kWh to 196.58 kWh. Under the sliding window evaluation, the model attained an average PICP of 0.945 with a PINAW of 0.346. The LogScore showed a modest improvement of 1.39%, decreasing from 7.4766 to 7.3725.

D. Summary of Model Performance at Key Horizons

Table I summarizes the normalized performance of the demand, generation, and energy balance forecasting models, all employing the CP-based approach. Key metrics, normalized by the global target range observed in the evaluation set (nRMSE, nPINAW, nCRPS), are presented as percentages for 1-hour, 12-hour, 24-hour, and 36-hour forecast horizons.

E. Comparative Analysis with Quantile Regression

To address the need for comparison with state-of-the-art probabilistic methods, a gradient-boosting QR model [30] was implemented and evaluated under a unified framework. Both CP and QR were assessed on the same held-out test set, with PINAW normalized by the same global target range to ensure a fair comparison. Table II presents the results.

TABLE I: Normalized Performance Metrics [%] for Generation, Demand, and Energy Balance Models at Selected Horizons.

Model	Horizon	nRMSE [%]	nPINAW [%]	nCRPS [%]
Demand	1h	4.15	9.75	2.22
	12h	7.37	17.83	3.93
	24h	8.59	20.75	4.56
	36h	9.52	22.88	4.99
Generation	1h	5.67	13.69	3.07
	12h	9.58	19.68	5.14
	24h	10.19	21.77	5.40
	36h	11.10	24.17	5.93
Balance	1h	7.22	21.44	3.76
	12h	9.60	33.95	4.96
	24h	9.69	40.59	5.09
	36h	10.09	38.73	5.27

TABLE II: Performance Comparison: CP vs. QR at the 95% nominal level. MIS denotes the Winkler Score normalized by target range. Best values per target in bold.

Target	Method	PICP	PINAW	MIS
Generation	CP	0.940	**0.319**	**0.430**
	QR	**0.979**	0.428	0.450
Demand	CP	**0.923**	0.298	**0.474**
	QR	0.792	**0.266**	0.660
Balance	CP	**0.945**	0.346	**0.448**
	QR	0.784	**0.233**	0.529

The CP framework consistently achieves lower MIS across all three targets, indicating better overall interval quality when width and calibration are considered jointly. For the generation target, both methods achieve adequate coverage, yet CP produces 25% narrower intervals (PINAW of 0.319 vs. 0.428). For demand and balance, the QR model exhibits substantial under-coverage (PICP of 0.792 and 0.784, respectively), falling below the nominal 0.95 target, whereas the CP approach with optimized sliding windows achieves PICP values of 0.923 and 0.945. Although QR yields nominally narrower intervals (lower PINAW), this comes at the cost of inadequate coverage, as reflected in its higher MIS values. Using the MIS as the underlying loss function, the Diebold–Mariano test confirmed that CP statistically outperforms QR ($p < 0.05$) in 36/36 horizons for generation, 34/36 for demand, and 33/36 for balance. QR achieves lower MIS at short lead times (horizon 1), where small prediction errors favor its tighter quantile-specific bounds. Regarding computational requirements, the complete CP pipeline, comprising DL inference and CP interval construction, exhibits an end-to-end latency below 40 ms per 36-hour prediction sequence on an Intel Core i7-14650HX CPU. While both approaches require base model retraining under significant distribution shift, CP's calibration step alone adapts in under 3 ms by updating the sliding window, whereas QR additionally requires retraining of all quantile models (approximately 1–2 minutes). This makes CP particularly well suited for real-time VPP control and intra-hour energy dispatching.

F. Discussion of Applicability and Insights

The probabilistic forecasts generated by the proposed models offer potential benefits for stakeholders in rural energy systems, including municipalities, smart grid operators, and VPP managers [9], [27]. Their utility includes improved risk assessment for energy imbalances, more informed scheduling of energy storage systems by anticipating surplus or deficit periods, and enhanced participation in demand response programs through a clearer understanding of net demand ranges [28]. Furthermore, the sliding-window calibration mechanism mitigates violations of the exchangeability assumption required by CP, as the locally adapted score distributions consistently yielded higher PICP values than a fixed global calibration set, confirming that temporal locality is essential for maintaining valid coverage under seasonal and non-stationary energy dynamics.

The results highlight the critical trade-offs between point accuracy and probabilistic reliability in VPP forecasting. Although the strong daily seasonality of energy balance enables the naive DSP to achieve a lower point MAE, the CP-augmented DL architectures provide sharper prediction intervals (lower PINAW) and superior overall probabilistic skill (lower CRPS), as detailed in Table I. In operational contexts, these robust worst-case uncertainty bounds take precedence over exact central tendency. Furthermore, when compared to the QR baseline, which produces narrower intervals but suffers from severe under-coverage for demand and balance (PICP < 0.80), the dynamic sliding-window CP framework maintains coverage closer to the nominal level. It achieves PICP values closely aligned with the 0.95 nominal target alongside superior interval quality (lower MIS) across all tasks (Table II). The generation model attains the lowest absolute MIS (0.430), whereas demand and balance show the most significant relative improvements. Finally, computational analysis confirms the CP pipeline's viability for real-time, intra-hour deployment.

IV. Conclusion and Future Work

This study applied CP techniques with dynamic sliding calibration windows to generate probabilistic forecasts for solar power generation, energy demand, and net energy balance over a 36-hour horizon. A comparative analysis showed that the CP framework statistically outperforms QR across all three forecasting targets, as confirmed by Diebold–Mariano tests. Under a unified evaluation framework, the CP models achieved well-calibrated PICP values of 0.940, 0.923, and 0.945 for generation, demand, and balance, respectively, while consistently delivering superior MIS scores. Notably, direct forecasting of the net balance proved essential, as indirect derivation ($G_t - D_t$) propagated errors that yielded excessively wide intervals. Furthermore, computational profiling revealed that the complete pipeline operates with sub-40 ms latency, supporting its suitability for real-time deployment in VPPs and intra-hour energy management. These results suggest that CP provides an adaptive and computationally efficient framework for probabilistic energy forecasting in decentralized systems with high renewable penetration. Future work will focus on exploring adaptive non-conformity measures and alternative DL architectures to further improve calibration for the demand and balance targets, as well as integrating these probabilistic forecasts into energy management optimization algorithms for applications such as battery scheduling, local energy market participation, and deployment on edge computing devices for decentralized VPPs.

ACKNOWLEDGMENT

This research was supported by the Comunidad de Madrid through the PREDFLEX-CM project (TEC-2024/ECO-287) under the R&D activities programme “Tecnologías2024”; the “RuralVPP” project (ref. SBPLY/21/180501/000223), funded by the Junta de Comunidades de Castilla-La Mancha (JCCM) and the European Union through the European Regional Development Fund; the “IA-CONV” project (ref. CM/DEMG/2024-039); and the “UrbanFlex” project (ref. CM/DEMG/2024-030), funded by the Community of Madrid through the grant agreement for the promotion and advancement of research and technology transfer at the University of Alcalá. Additionally, this work was supported by a contract (ref. 16 UAH-INV-JCCM) subsidized under the “Programa Investigo” of Castilla-La Mancha, financed by the Fondo Social Europeo Plus (FSE+) of Castilla-La Mancha 2021–2027, with a co-financing rate of 85%.

REFERENCES

[1] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, ”Solar forecasting methods for renewable energy integration,” *Prog. Energy Combust. Sci.*, vol. 39, no. 6, pp. 535–576, Dec. 2013, doi: 10.1016/j.pecs.2013.06.002.

[2] F. Shahnazian, K. Das, R. Yan, and P. Sørensen, ”Challenges of renewable energy integration to weak grids,” in *Proc. 22nd Wind Solar Integr. Workshop (WIW 2023)*, 2023, vol. 2023, pp. 529–536, doi: 10.1049/icp.2023.2783.

[3] O. Smith, O. Cattell, E. Farcot, R. D. O’Dea, and K. I. Hopcraft, ”The effect of renewable energy incorporation on power grid stability and resilience,” *Sci. Adv.*, vol. 8, no. 9, Art. no. eabj6734, Mar. 2022, doi: 10.1126/sciadv.abj6734.

[4] M. Shafiullah, S. D. Ahmed, and F. A. Al-Sulaiman, ”Grid integration challenges and solution strategies for solar PV systems: A review,” *IEEE Access*, vol. 10, pp. 52233–52257, 2022, doi: 10.1109/ACCESS.2022.3174555.

[5] C. Li, H. Shi, Y. Cao, J. Wang, Y. Kuang, Y. Tan, and J. Wei, ”Comprehensive review of renewable energy curtailment and avoidance: A specific example in China,” *Renewable Sustainable Energy Rev.*, vol. 41, pp. 1067–1079, Jan. 2015, doi: 10.1016/j.rser.2014.09.009.

[6] M. Tahanan, W. van Ackooij, A. Frangioni, and F. Lacalandra, ”Large-scale unit commitment under uncertainty,” *4OR*, vol. 13, no. 2, pp. 115–171, Jun. 2015, doi: 10.1007/s10288-014-0279-y.

[7] Q. Wang *et al.*, ”Quantifying the value of probabilistic forecasting for power system operation planning,” *Appl. Energy*, vol. 343, Art. no. 121254, Aug. 2023, doi: 10.1016/j.apenergy.2023.121254.

[8] C. Kath and F. Ziel, ”Conformal prediction interval estimation and applications to day-ahead and intraday power markets,” *Int. J. Forecasting*, vol. 37, no. 2, pp. 777–799, Apr.-Jun. 2021, doi: 10.1016/j.ijforecast.2020.09.006.

[9] Y. Renkema, N. Brinkel, and T. Alskaif, ”Conformal prediction for stochastic decision-making of PV power in electricity markets,” *Electr. Power Syst. Res.*, vol. 234, Art. no. 110750, May 2024, doi: 10.1016/j.epsr.2024.110750.

[10] A. Faustine and L. Pereira, ”Conformal multilayer perceptron-based probabilistic net-load forecasting for low-voltage distribution systems with photovoltaic generation,” in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2024, pp. 59–64, doi: 10.1109/SmartGridComm60555.2024.10738106.

[11] M. Borrotti, ”Quantifying uncertainty with conformal prediction for heating and cooling load forecasting in building performance simulation,” *Energies*, vol. 17, no. 17, Art. no. 4348, Aug. 2024, doi: 10.3390/en17174348.

[12] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*. Springer, 2005, doi: 10.1007/b106715.

[13] G. Shafer and V. Vovk, ”A Tutorial on Conformal Prediction,” *Journal of Machine Learning Research*, vol. 9, pp. 371–421, Mar. 2008.

[14] H. Hersbach *et al.*, ”The ERA5 global reanalysis,” *Quart. J. Roy. Meteor. Soc.*, vol. 146, no. 730, pp. 1999–2049, Jun. 2020, doi: 10.1002/qj.3803.

[15] T. H. Soukissian, F. E. Karathanasi, and D. K. Zaragkas, ”Exploiting offshore wind and solar resources in the Mediterranean using ERA5 reanalysis data,” *Energy Convers. Manage.*, vol. 237, art. no. 114092, Jun. 2021, doi: 10.1016/j.enconman.2021.114092.

[16] K. Chreng, H. S. Lee, and S. Tuy, ”Electricity demand prediction for sustainable development in Cambodia using recurrent neural networks with ERA5 reanalysis climate variables,” *Energy Reports*, vol. 8, pp. 76–81, Jun. 2022, doi: 10.1016/j.egyr.2022.01.025.

[17] K. Ferkous, M. Guermoui, S. Menakh, A. Bellaour, and T. Boulmaiz, ”A novel learning approach for short-term photovoltaic power forecasting, a review and case studies,” *Eng. Appl. Artif. Intell.*, vol. 133, art. no. 108502, Jan. 2024, doi: 10.1016/j.engappai.2024.108502.

[18] A. Khosravi, R. N. N. Koury, L. Machado, and J. J. G. Pabon, ”Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms,” *J. Clean. Prod.*, vol. 176, pp. 63–75, Mar. 2018, doi: 10.1016/j.jclepro.2017.12.065.

[19] X. Li *et al.*, ”Probabilistic solar irradiance forecasting based on XGBoost,” *Energy Reports*, vol. 8, suppl. 3, pp. 360–367, Mar. 2022, doi: 10.1016/j.egyr.2022.02.251.

[20] B. Schulz, M. El Ayari, S. Lerch, and S. Baran, ”Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting,” *Solar Energy*, vol. 220, pp. 1016–1031, May 2021, doi: 10.1016/j.solener.2021.03.023.

[21] G. Moreno, P. Martín, C. Santos, F. J. Rodríguez, and E. Santiso, ”A day-ahead irradiance forecasting strategy for the integration of photovoltaic systems in virtual power plants,” *IEEE Access*, vol. 8, pp. 201101–201114, Nov. 2020, doi: 10.1109/ACCESS.2020.3036140.

[22] G. Moreno, C. Santos, P. Martín, F. J. Rodríguez, R. Peña, and B. Vuksanovic, ”Intra-day solar power forecasting strategy for managing virtual power plants,” *Sensors*, vol. 21, no. 16, art. no. 5648, Aug. 2021, doi: 10.3390/s21165648.

[23] A. K. Mandal, R. Sen, S. Goswami, and B. Chakraborty, ”Comparative study of univariate and multivariate long short-term memory for very short-term forecasting of global horizontal irradiance,” *Symmetry*, vol. 13, no. 8, art. no. 1544, Aug. 2021, doi: 10.3390/sym13081544.

[24] D. G. Erbs, S. A. Klein, and J. A. Duffie, ”Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation,” *Solar Energy*, vol. 28, no. 4, pp. 293–302, 1982, doi: 10.1016/0038-092X(82)90302-4.

[25] C. F. M. Coimbra, J. Kleissl, and R. Marquez, ”Overview of solar-forecasting methods and a metric for accuracy evaluation,” in *Solar Energy Forecasting and Resource Assessment*, J. Kleissl, Ed. Cambridge, MA: Academic Press, 2013, pp. 171–194, doi: 10.1016/B978-0-12-397177-7.00008-5.

[26] T. Gneiting *et al.*, ”Probabilistic forecasts, calibration and sharpness,” *J. R. Stat. Soc. Ser. B*, vol. 69, no. 2, pp. 243–268, Apr. 2007, doi: 10.1111/j.1467-9868.2007.00587.x.

[27] M. Jin, W. Feng, P. Liu, C. Marnay, and C. Spanos, ”MOD-DR: Microgrid optimal dispatch with demand response,” *Appl. Energy*, vol. 187, pp. 758–766, Feb. 2017, doi: 10.1016/j.apenergy.2016.11.093.

[28] A. Molderink, V. Bakker, M. G. C. Bosman, J. L. Hurink, and G. J. M. Smit, ”Management and control of domestic smart grid technology,” *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 109–119, Sep. 2010, doi: 10.1109/TSG.2010.2055904.

[29] F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 253–263, Jul. 1995, doi: 10.1080/07350015.1995.10524599.

[30] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.