Enhancing Biogas Energy Production Forecasting with Box-Cox Transformation and Prediction Interval Methods

Nazmus Sakib, Mohammad Anwar Hosen, Burhan Khan, Bruce Gunn, Michael Johnstone
Intelligent Systems Research and Innovation (IISRI), Deakin University,
Waurn Ponds, Victoria, Australia
{s.sakib, anwar.hosen, burhan.khan, b.gunn, michael.johnstone}@deakin.edu.au

2025 IEEE International systems Conference (SysCon) | 979-8-3315-0818-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/SYSCON64521.2025.11014866

***Abstract*—Biogas, a renewable energy source, plays a crucial role in reducing greenhouse gas emissions and promoting sustainable energy production. Its dual benefits of waste management and energy generation make biogas a key component of renewable energy strategies. However, forecasting biogas energy production remains challenging due to inherent variability and uncertainties in feedstock composition and co-digestion processes. This study addresses these challenges by comparing traditional point predictions with uncertainty-aware approaches. Initially, a simple neural network was applied to the original data, achieving good accuracy for point predictions but failing to account for production uncertainty, limiting its practical applicability. Recognizing the importance of capturing variability, this research integrates a hybrid approach utilizing the Box-Cox transformation and the Lower Upper Bound Estimation (*LUBE*) method. The Box-Cox transformation normalized the skewed data, improving the model's ability to capture complex variations and refine Prediction Interval Normalized Average Width (*PINAW*). The LUBE method, applied after transformation, provided prediction intervals that were both narrower and more reliable than traditional methods. These enhanced intervals significantly improve decision-making by providing better insights into production variability. Policymakers and investors can leverage these tools for strategic planning and resource allocation, aligning with sustainable development goals (SDG-7) by fostering confidence in renewable energy production from biogas.**

***Index Terms*—Biogas, Energy Production Forecast, Prediction Interval (*PI*), Lower Upper Bound Estimation (*LUBE*), Uncertainty Quantification, Box-Cox Transformation, Sustainable Energy Strategies.**

I. INTRODUCTION

Biogas is a versatile form of renewable energy, is produced through the anaerobic digestion of organic waste materials, including municipal solid waste, agricultural residues, and animal by-products. With the increasing global focus on reducing carbon emissions and promoting sustainable energy sources, biogas has garnered significant attention for its dual role in energy generation and greenhouse gas reduction [1] [2] [3]. Biogas plants, which convert waste into energy, not only mitigate methane emissions—a potent greenhouse gas—but also contribute to energy self-sufficiency, particularly for farms and rural areas [4]. Despite these benefits, the production of biogas is inherently variable, dependent on a range of factors such as feedstock composition, environmental conditions, and digestion technologies. As such, accurate forecasting of biogas production is essential for optimising energy output and ensuring operational efficiency [5] [6] [7].

One of the key challenges in biogas production lies in its variability. Local environmental conditions, including weather, feedstock type, digestion methods, and co-digestion processes, contribute to fluctuations in biogas output. Feedstock variability, for example, can result in significant differences in methane yield, making it difficult to predict biogas production consistently [1]. This variability poses operational challenges for biogas plants, particularly those that require steady energy output to meet electricity demands. Accurate forecasting models must, therefore, account for these nonlinear and complex relationships to improve the reliability of predictions [4] [7] [8].

Traditional statistical model and machine learning based prediction models, which often rely on point estimates, have been shown to be inadequate in capturing the uncertainties inherent in biogas production [2]. These models typically fail to address the variability of input parameters such as substrate retention time (SRT), volatile dry solids (VDS), and methane production rates, all of which interact in complex ways during the anaerobic digestion process [9]. Moreover, point predictions are susceptible to biases introduced by regional farm conditions, weather variability, and differences in digestion technology. As a result, there is a growing need for advanced models that can provide more reliable, uncertainty-aware forecasts [5] [7] [10].

In response to these challenges, this study introduces a hybrid approach that combines data preprocessing using Box-Cox transformation with prediction interval (*PI*) method like Lower Upper Bound Estimation (*LUBE*) to improve the accuracy of biogas production forecasting [8]. The dataset utilised in this research is derived from biogas plants that generate electricity using organic waste, particularly cattle manure. The data includes the number of cattle, biogas yield, and corresponding electricity output, offering a comprehensive view of energy production from biogas. However, due to the high skewness of the data and the variability in digestion methods, a preprocessing step was necessary [11] [12]. Outlier detection, missing data imputation, and the Box-Cox transformation were employed to normalise the skewed data distributions and ensure suitability for further modeling [13] [14].

Following data normalisation, this study leverages the Lower Upper Bound Estimation (*LUBE*) method, a prediction interval technique that provides a range of possible outcomes rather than a single-point estimate. By generating upper and lower bounds for the predicted biogas output, the *LUBE* method accounts for the inherent uncertainties in the biogas production process [15] [16]. Additionally, simulated annealing is used to optimise the prediction intervals, refining the bounds to ensure a robust and reliable forecast. This hybrid methodology offers several key advantages over traditional point prediction models, including improved accuracy, better uncertainty quantification, and the ability to capture complex nonlinear interactions in the data. By providing a more reliable forecast, this method allows biogas plants to optimise their operations, improve energy output, and better integrate biogas into the broader energy mix. In summary, the main contributions of this work are as follows:

- Firstly, We improve *PI* quality using a hybrid predictive framework, data processed utilising Box-Cox transformation and *PI* estimation through the *LUBE* method. This approach addresses data skewness and ensures more comprehensive uncertainty quantification for biogas production forecasting.
- Lastly, We showcase the practicality of the proposed method in providing narrower, yet reliable prediction intervals, thereby supporting decision-makers and investors in the renewable energy sector with actionable and confident forecasts that align with sustainable development goals (SDG), such as SDG-7 [17].

The remainder of this paper is structured as follows: Section II reviews the relevant literature on the biogas industry, highlighting trends, challenges, and opportunities. Section III outlines the methodology, followed by the analysis results and discussion in Section IV. Finally, Section V presents concluding remarks and recommendations for future research.

II. Literature Review

Biogas production has emerged as a crucial element in the global push towards renewable energy. With mounting concerns over climate change and the depletion of fossil fuel resources, the conversion of organic waste into energy offers a sustainable solution that addresses both environmental and energy-related challenges [2] [3]. The significance of biogas lies in its dual role—converting waste into electricity while reducing harmful methane emissions. This aligns with the principles of the circular economy, as it enables the reuse of waste products, particularly in sectors such as agriculture, municipal waste management, and energy production [5] [18]. Salvador et al. emphasised that biogas plants contribute significantly to environmental sustainability by reducing methane emissions from decomposing waste, which would otherwise contribute to the greenhouse effect. Despite its advantages, biogas production is not without challenges, particularly in terms of operational variability and forecasting difficulties [1] [8] [10].

Biogas production is influenced by several complex factors, including feedstock composition, local environmental conditions such as temperature and humidity, and the type of digestion technology employed [6]. This variability makes it difficult to predict biogas output consistently. Ming-Chuan Et el. [19] discusses how changes in temperature and retention time significantly impact biogas yield, and how the anaerobic digestion process is sensitive to fluctuations in microbial activity, which may be affected by variations in feedstock or operational conditions. Given these complexities, predicting biogas production accurately is a significant challenge that must be addressed to optimise biogas plants and meet energy production goals [11].

Machine learning and deep learning techniques have become increasingly popular for addressing the challenges in biogas production forecasting [3] [18]. Unlike traditional statistical methods, machine learning models can capture nonlinear relationships and learn from historical data, making them well-suited for the complex and variable nature of biogas production. Various machine learning models, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forests (RF), have been applied to biogas forecasting [2] [18] [20] [21]. These models excel in learning patterns from data, providing reasonable predictions for biogas output. However, they typically produce point estimates—single predictions of biogas yield—which do not capture the uncertainty or variability inherent in the data [12]. For instance, Yi et al. [11] demonstrated that while machine learning models are effective in many scenarios, their inability to account for uncertainty in predictions limits their practical application, especially when biogas production is subject to sudden fluctuations due to external factors such as weather changes or feedstock variations [4] [6].

To address the inherent uncertainties in biogas production, prediction interval methods have been developed to provide a range of possible outcomes, rather than a single-point estimate. This approach acknowledges that biogas production is influenced by numerous unpredictable factors and that forecasting models must account for this variability to provide robust and reliable predictions [10] [11]. The *LUBE* method is a widely used approach for generating prediction intervals. *LUBE* does not merely provide an average output prediction but instead calculates upper and lower bounds that encompass a range of likely outcomes [15] [22]. By generating these intervals, the *LUBE* method allows decision-makers to assess the confidence level of predictions, offering a clearer understanding of potential variations in biogas production.

In assessing prediction intervals, several key metrics are employed, including Prediction Interval Coverage Probability (*PICP*), Prediction Interval Normalised Average Width (*PINAW*), and Coverage Width-based Criteria (*CWC*) [16] [23]. These metrics provide a comprehensive evaluation of the accuracy and reliability of prediction intervals. *PICP*, for instance, measures the proportion of actual data points that fall within the predicted interval, which is critical for understanding how well the model captures the variability of biogas production. The equation for *PICP* is given as:

$$PICP = \frac{1}{n} \sum_{i=1}^{n} I(y_i \in [L_i, U_i]) \tag{1}$$

Where n is the number of observations, y_i represents the actual values, L_i and U_i are the lower and upper bounds of the prediction interval, respectively, and $I(\cdot)$ is the indicator function that equals 1 if the actual value falls within the bounds. Meanwhile, *PINAW* measures the width of the prediction interval relative to the actual values, calculated as:

$$PINAW = \frac{1}{n} \sum_{i=1}^{n} \frac{U_i - L_i}{\bar{y}_i} \tag{2}$$

This metric is essential for balancing the trade-off between interval width and coverage—narrow intervals may be desirable but must still capture the true values. Finally, the *CWC* metric incorporates both *PICP* and *PINAW* into a single criterion to optimise prediction intervals, expressed as:

$$CWC = PINAW + PICPe^{-\eta(PICP-\phi)} \tag{3}$$

Where η is a control parameter that penalises intervals with low coverage probabilities. This equation is critical in ensuring that the intervals are not only narrow but also provide sufficient coverage to be useful in real-world applications [16] [22].

However, before applying prediction models and generating intervals, it is crucial to address the quality of the input data. In the case of biogas production, the variability of feedstock and environmental conditions often results in skewed data distributions, which can negatively affect the performance of predictive models [1] [14]. Skewed data makes it difficult for models, especially those based on statistical or machine learning methods, to accurately capture the underlying patterns in the data. To mitigate this issue, data transformation techniques like the Box-Cox transformation are applied [24].

The Box-Cox transformation is a statistical method designed to stabilise variance and transform non-normal dependent variables into a normal distribution [25]. This is particularly useful in biogas production data, where skewness due to factors like sudden spikes in production or variations in feedstock composition can lead to inaccurate predictions. The transformation is defined as:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y) & \text{if } \lambda = 0 \end{cases} \tag{4}$$

Where y is the original data and λ is the transformation parameter. By applying this transformation, the data becomes more normally distributed, improving the performance of predictive models that assume normality, such as regression-based methods or machine learning algorithms [25]. In the context of renewable energy data, where fluctuations are common, the Box-Cox transformation helps stabilise these variations, making the data more suitable for predictive modeling.

To interpret the results in their original scale, the reverse Box-Cox transformation is applied [26]. The reverse transformation reverts the normalised data back to its original scale, ensuring that predictions are interpretable and actionable. The reverse transformation is defined as:

$$y = \begin{cases} (\lambda y(\lambda) + 1)^{\frac{1}{\lambda}} & \text{if } \lambda \neq 0 \\ \exp(y(\lambda)) & \text{if } \lambda = 0 \end{cases} \tag{5}$$

Here, $y(\lambda)$ is the transformed data, y is the original data, and λ is the transformation parameter. When $\lambda = 0$, the reverse transformation simplifies to the exponential function, as the original transformation corresponds to a natural logarithmic operation [26]. The reverse Box-Cox transformation is crucial for ensuring the practical applicability of predictive models, allowing results to be evaluated and interpreted in their original context.

Despite the advances in machine learning and the application of prediction intervals in biogas forecasting, there remain gaps in the existing body of research. While many studies have focused on general biogas production, there is limited research on predicting electricity generation from biogas, particularly in agricultural settings [1]. Farms that rely on biogas to meet their energy demands require accurate predictions of electricity output to ensure self-sufficiency and efficient resource utilisation [7] [10]. Furthermore, while data transformation techniques like the Box-Cox transformation are employed in some studies related to electricity load consumption prediction, the consistency of their use and the impact on model performance require further exploration [13] [25].

III. Methodology

In this section, we describe the dataset used in this study, the proposed model, and the step-by-step approach for forecasting biogas production. The flowchart of the proposed methodology is presented in Figure 1, followed by an explanation of each step, including data preprocessing, transformation, and the prediction interval estimation using the *LUBE* method.

A. Dataset Description

The dataset we have collected is from US-based farms over a period of one year. It contains data from 741 farms that manage various animals, including cattle, dairy, poultry, and swine. However, to focus on the most relevant subset, we have narrowed the dataset to 644 farms that primarily raise cattle, as this group represents the largest cattle population. These farms are located in states such as California, Arizona, Indiana, and North Carolina.

We further categorised the dataset based on whether the farms employ a digestion process. Specifically, we selected 295 farms that do not use a co-digestion process and 299 farms that employ the co-digestion method. The data from these farms include electricity generation from cattle manure, which presents non-linear characteristics. The multimodal nature of the data, compounded by skewed distributions and data gaps, makes it unsuitable for standard normalisation techniques like Min-Max scaling [25]. This necessitates more advanced preprocessing techniques to handle the uncertain behavior observed in the dataset.

A sample of the dataset is provided in Table I below:

TABLE I
SAMPLE DATA FROM US CATTLE FARMS DATASET

Location	Cattle Size	Co-Digestion	Electricity (KWh/Yr)
CA	7200	No	12,000,000
NC	5600	Yes	9,500,000
TX	8900	No	12,600,000

B. Proposed Model and Flowchart

The proposed model incorporates advanced preprocessing, data transformation, and prediction interval estimation methods. The process begins with data collection, followed by preprocessing to clean the data. A Box-Cox transformation is specifically applied to the electricity production feature to normalise its distribution. The *LUBE* method is then used to estimate prediction intervals, optimised using simulated annealing. The flowchart representing the proposed methodology is shown in Figure 1:

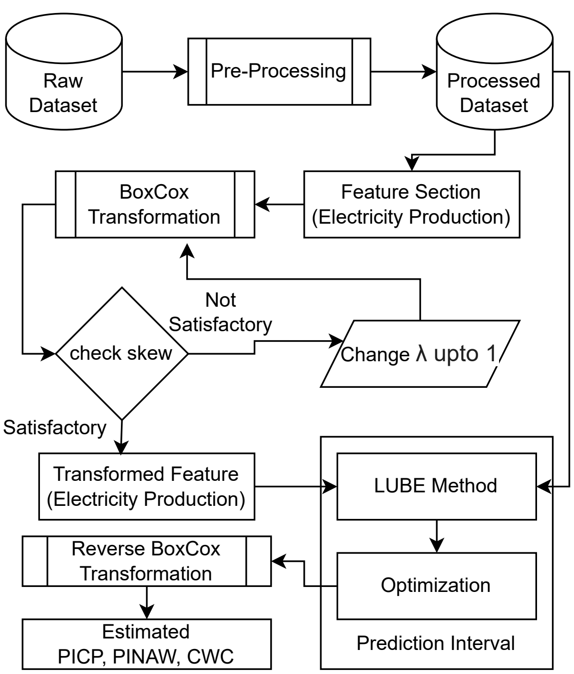

Fig. 1. Flowchart of the Proposed Methodology

C. Flowchart Description

The detailed steps of the proposed methodology are as follows:

1) Data Preprocessing: The initial step involves preparing the raw dataset to ensure data quality and suitability for modeling:

- **Outlier Detection and Removal**: Outliers were identified and removed using the interquartile range (IQR) method [27] to reduce their impact on model training.
- **Missing Data Imputation**: Missing data points were imputed using multiple imputation methods to maintain dataset integrity.
- **Feature Selection**: Key features were selected with a focus on the electricity production variable, as this is the primary target for prediction.

These preprocessing steps ensure that the dataset is ready for statistical analysis and model training, without nomalous data points or gaps.

2) Box-Cox Transformation: Given the multimodal nature and skewed distribution of the data, simple normalisation techniques are inadequate for preparing the data for modeling [24]. The Box-Cox transformation was applied exclusively to the electricity production feature due to its skewed distribution. This transformation stabilises variance and transforms skewed data into a distribution that approximates normality. The transformation parameter, λ, was iteratively optimised to achieve the best normalisation for this feature [13] [25]. This step is crucial for ensuring that the neural network model can learn effectively from the input data without being affected by distortions in data patterns.

We initiated the Box-Cox transformation with $\lambda = 0.1$, gradually increasing it to 0.5, where we observed the best fit with significantly reduced skewness in the data. At this point, the data began to approximate a normal distribution, as confirmed by visual inspection of histograms and statistical parameters. Figure 2 illustrates the distribution of the data before and after applying the Box-Cox transformation, highlighting the improvement in normality and reduction in skewness.

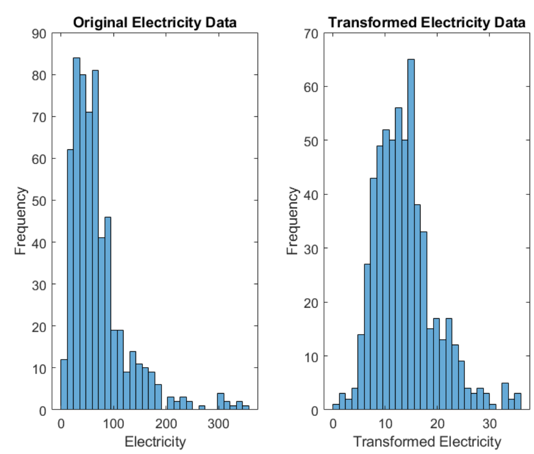

Fig. 2. Data Distribution Before and After Box-Cox Transformation

This targeted transformation ensured that the electricity production feature was suitable for subsequent modeling, minimising the impact of skewness while leaving other features in their original state to preserve their inherent characteristics.

3) LUBE Method and Optimisation: Once the data preprocessing and transformation were complete, the Lower Upper Bound Estimation (*LUBE*) method was employed to generate prediction intervals for electricity production. The *LUBE* method uses a feed-forward neural network to output two values for each input: an upper bound and a lower bound, forming the prediction interval [28] [16].

The dataset was split into 80% training and 20% testing to ensure robust model training and evaluation. The *LUBE* method, designed for interval prediction, accommodates the inherent uncertainties in biogas production, offering a more informative alternative to traditional point estimates.

Simulated annealing was employed as the optimisation technique for refining the prediction intervals generated by the *LUBE* method. This optimisation process balances the trade-off between interval width and coverage probability, ensuring that the prediction intervals are both narrow and comprehensive.

4) Reverse Transformation and Evaluation: After optimisation, a reverse transformation of the Box-Cox transformation was performed on the prediction intervals to revert them back to their original scale. This step ensures that the evaluation is conducted using real-world data, making the results more interpretable and actionable for realistic decision-making.

The prediction model's performance was evaluated using *PICP*, *PINAW*, and *CWC* to ensure the prediction intervals accurately capture the uncertainties in biogas forecasting. The evaluation process was repeated until these metrics met satisfactory levels; if not, further optimisation refined the model. Once performance was satisfactory, the final prediction intervals were produced, providing reliable forecasts for electricity production. This hybrid approach, using feature-specific transformation and advanced prediction interval techniques, improves decision-making by addressing data variability and uncertainty in biogas forecasting. The next section presents the model results and discusses their implications and accuracy.

IV. Results and Discussion

In this section, we present the results of applying different models to the dataset, including a traditional feed-forward neural network and the *LUBE* method, with and without data transformation using the Box-Cox model. The models were evaluated at various confidence levels (90% and 95%) using key metrics. The findings provide visual representations and a comprehensive comparison of the models' performance, highlighting the advantages of incorporating data transformation.

A. Performance of the Traditional Neural Network

A traditional feed-forward neural network was implemented as a baseline to evaluate point prediction performance. The results for the training and testing phases are summarised in Table II. While the network achieved reasonable performance, indicated by an R-squared (R^2) value of 0.8008 on the testing set, the testing Mean Squared Error (MSE) of 422.4145 and Mean Absolute Error (MAE) of 10.5115 reveal limitations in handling the volatile characteristics of biogas production data.

TABLE II
Performance Metrics of the Traditional Neural Network

Evaluation Metrics	Result Values
Training MSE	550.2291
Testing MSE	422.4145
Testing R^2	0.8008
Testing MAE	10.5115

The relatively lower R^2 value and high error metrics indicate that while the neural network can provide point predictions, it struggles to generalise well under the volatile nature of biogas production data, which includes factors like feedstock composition, co-digestion methods, and environmental changes. This variability limits the reliability of point predictions for decision-making, where understanding a range of possible outcomes is crucial for effective risk management and planning.

Figure 3 illustrates the actual versus predicted values using the neural network. While there is some alignment between the predicted and actual outputs, noticeable deviations occur, highlighting the challenges posed by data variability in capturing the full range of production scenarios.

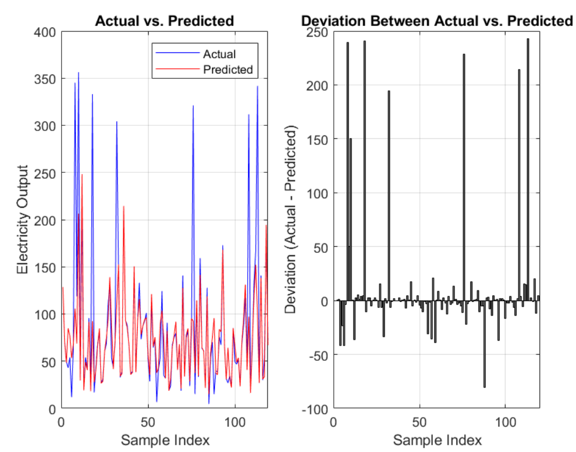

Fig. 3. Actual vs Predicted Values in Traditional Neural Network

These results underscore the need for models that not only achieve accurate point predictions but also provide prediction intervals to quantify uncertainty associated with volatile data. This capability is essential for renewable energy production scenarios, where variability significantly impacts operational decisions and energy planning.

B. LUBE Method on Original Data

The *LUBE* method applied to the original dataset to generate prediction intervals. The method was evaluated at confidence levels of 90% and 95%, and the results are shown in Table III.

At a 90% confidence level, the *LUBE* method achieved a *PINAW* of 35.2989 and a *CWC* of 93.1476, indicating

TABLE III
LUBE METHOD PERFORMANCE ON ORIGINAL DATA

Metric	90% Confidence	95% Confidence
CWC	93.1476	84.4544
PICP	86.3345	93.4512
PINAW	35.2989	53.0133

reasonably narrow prediction intervals but leaving room for improvement in practical applicability. At the 95% confidence level, ensuring better coverage, but this came at the cost of a wider *PINAW* of 53.0133 and a higher *CWC* of 84.4544. This demonstrates that while the *LUBE* method effectively captures most of the actual values, achieving higher coverage results in increased interval width, which may limit its utility in some practical scenarios. Figure 4 shows the prediction intervals generated by the *LUBE* method on the original data at 95% confidence level.

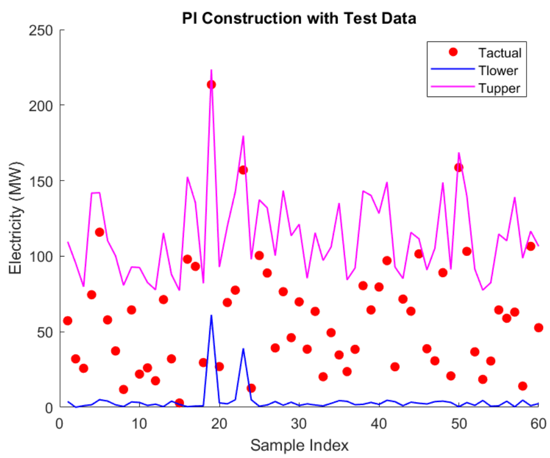

Fig. 4. Prediction Intervals using LUBE on Original Data

C. LUBE Method with Box-Cox Transformed Data

The dataset was transformed using the Box-Cox method to improve data distribution. The *LUBE* method was then re-applied to the transformed data, yielding the results shown in Table IV.

TABLE IV
LUBE METHOD PERFORMANCE ON BOX-COX TRANSFORMED DATA

Metric	90% Confidence	95% Confidence
CWC	67.0578	53.0905
PICP	88.4546	95.1232
PINAW	35.5238	52.5464

The results show significant improvement with the Box-Cox transformation. At a 90% confidence level, the model achieved a reduced *PINAW* of 35.5238 and a *CWC* of 67.0578, indicating prediction intervals of higher quality. Similarly, at a 95% confidence level, the *PINAW* was 52.5464 and the *CWC* was 53.0905. This demonstrates that data transformation enhances the performance of prediction interval methods, providing a better balance between coverage and interval width.

Figures 5 and 6 illustrate the prediction intervals and transformed vs. original PI comparison respectively.

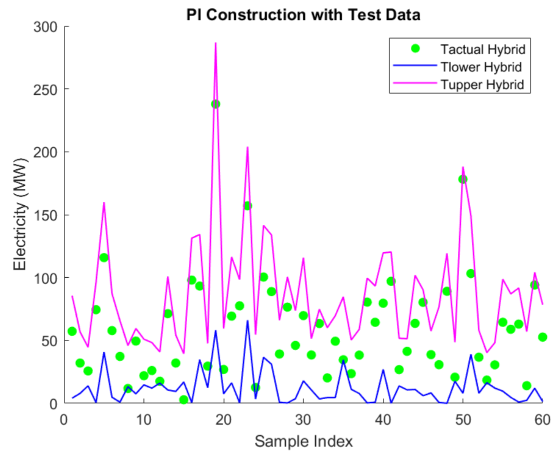

Fig. 5. Prediction Intervals using Box-Cox with LUBE

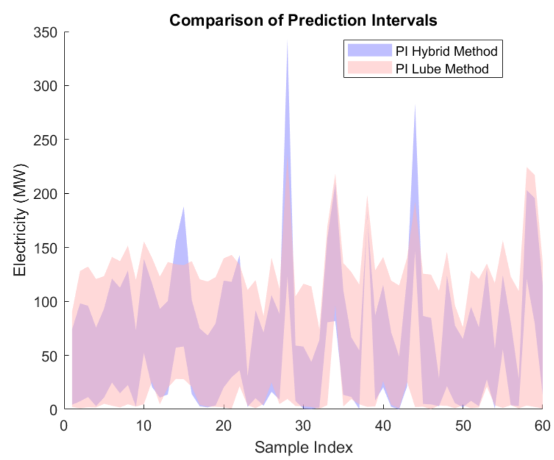

Fig. 6. Transformed vs. Original PI Comparison

D. Comparative Analysis and Discussion

This study conducted a comparative analysis of three models: a traditional neural network, the *LUBE* method applied to the original data, and the *LUBE* method with Box-Cox transformed data. The results, presented in Tables III and IV, highlight the importance of preprocessing and data transformation.

While the neural network demonstrated high accuracy in point predictions, it lacked the ability to capture the uncertainty inherent in the volatile characteristics of biogas production data. The *LUBE* method on original data provided more informative results but produced relatively wide prediction intervals, as shown by higher *PINAW* and *CWC* values. This indicates that although *LUBE* captures uncertainty, it is less precise without data transformation.

The Box-Cox transformation significantly improved model performance by reducing both *PINAW* and *CWC*, resulting in narrower and more reliable prediction intervals. This transformation allowed the *LUBE* method to maintain predefined coverage probability while enhancing the reliability and precision of the intervals. The improved performance at both 90% and 95% confidence levels demonstrates that advanced preprocessing, when combined with robust prediction interval estimation, provides a balanced approach that addresses both reliability and accuracy in forecasting.

These findings reinforce the importance of adopting hybrid methodologies in forecasting models. By effectively capturing and managing the inherent variability in renewable energy production, such methods can support more reliable decision-making. This approach not only aids energy planners and policymakers in resource allocation and strategic planning but also fosters sustainable energy practices, contributing to more resilient and informed energy strategies.

E. Practical Implications for Biogas Production

The findings of this study offer valuable insights for policymakers, energy planners, and biogas plant operators. The improved prediction intervals, especially those achieved with Box-Cox transformed data, provide a more precise and reliable tool for forecasting electricity generation from biogas. This enhanced accuracy allows decision-makers to plan with greater confidence, minimising the risk of over- or underestimating energy production.

For example, at a 95% confidence level, a farm with 10,000 cattle can expect to produce between 14.5438 MW(Mega Watt per Hour) and 86.5487 MW of electricity annually from cattle manure alone. This prediction interval offers flexibility for farm operators to manage resources effectively across varying production scenarios. When co-digestion processes are used, combining cattle manure with other organic materials like agricultural residues, the output may vary more due to complex biochemical interactions. For instance, a farm with 11,300 cattle utilising co-digestion might produce between 9.376 MW and 71.357 MW annually. This variability arises from factors such as substrate composition and degradability, which influence methane yields. Challenges like the carbon-to-nitrogen ratio and substrate compatibility can result in less consistent biogas yields, underscoring the need for prediction intervals that account for these additional uncertainties [1] [10] [14].

reliable prediction intervals also support effective resource allocation, investment decisions, and contingency planning, aligning with sustainable energy goals like SDG-7 (affordable and clean energy) [17]. The ability to forecast energy production with narrower, reliable intervals aids in long-term strategic planning, contributing to a stable and efficient energy supply.

V. Conclusion and Future Direction

This study effectively addressed the inherent variability and uncertainty in biogas electricity production by leveraging a hybrid methodology that integrates the Box-Cox transformation with the *LUBE* method. By transforming the skewed data and applying the *LUBE* method, optimised using simulated annealing, the model achieved significant improvements in prediction interval accuracy. The results demonstrate the enhanced capability of the model to provide reliable and precise prediction intervals, essential for informed decision-making in renewable energy forecasting.

The comparative analysis confirmed that while traditional neural networks offer high point prediction accuracy, they lack the ability to quantify uncertainty, which is crucial for practical applications in energy planning. The application of the *LUBE* method on Box-Cox transformed data produced narrower and more reliable prediction intervals compared to the original data, showcasing the importance of data preprocessing in optimising model performance. This approach offers a robust tool for policymakers and biogas plant operators, aiding in effective resource management, risk assessment, and strategic planning aligned with sustainable energy goals.

Future research can enhance this study by integrating real-time data from IoT-enabled biogas plants for continuous model updates and adaptive forecasting, improving accuracy and efficiency. Expanding this hybrid approach to other renewable energy sources, like wind and solar, can address similar forecasting challenges. Further exploration of complex model structures, incorporating factors such as microbial activity and feedstock variability, and advanced techniques like ensemble learning for uncertainty quantification, could improve robustness. Collaboration with industry partners for real-world validation would strengthen the model's applicability, supporting resilient and sustainable energy systems.

References

[1] R. Salvador, M. V. Barros, J. G. D. P. D. Rosário, C. M. Piekarski, L. M. da Luz, and A. C. de Francisco, "Life cycle assessment of electricity from biogas: A systematic literature review," *Environmental Progress & Sustainable Energy*, vol. 38, no. 4, p. 13133, 2019.

[2] Y. Sun, H.-l. Dai, H. Moayedi, B. N. Le, and R. M. Adnan, "Predicting steady-state biogas production from waste using advanced machine learning-metaheuristic approaches," *Fuel*, vol. 355, p. 129493, 2024.

[3] K. Jeong, A. Abbas, J. Shin, M. Son, Y. M. Kim, and K. H. Cho, "Prediction of biogas production in anaerobic co-digestion of organic wastes using deep learning models," *Water research*, vol. 205, p. 117697, 2021.

[4] I. Govender, G. A. Thopil, and R. Inglesi-Lotz, "Financial and economic appraisal of a biogas to electricity project," *Journal of Cleaner Production*, vol. 214, pp. 154–165, 2019.

[5] J. Sappl, M. Harders, and W. Rauch, "Machine learning for quantile regression of biogas production rates in anaerobic digesters," *Science of the Total Environment*, vol. 872, p. 161923, 2023.

[6] D. I. Massé, G. Talbot, and Y. Gilbert, "On farm biogas production: A method to reduce ghg emissions and develop more sustainable livestock operations," *Animal feed science and technology*, vol. 166, pp. 436–445, 2011.

[7] V. Aryai and M. Goldsworthy, "Day ahead carbon emission forecasting of the regional national electricity market using machine learning methods," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106314, 2023.

[8] L. Chen, R.-G. Cong, B. Shu, and Z.-F. Mi, "A sustainable biogas model in china: The case study of beijing deqingyuan biogas project," *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 773–779, 2017.

[9] P. Kianmehr, W. Mansoor, and F. A. Kfoury, "Prediction of biogas generation profiles in wastewater treatment plants using neural networks," *Journal of Clean Energy Technologies*, vol. 2, no. 3, p. 201, 2014.

[10] S. Abanades, H. Abbaspour, A. Ahmadi, B. Das, M. A. Ehyaei, F. Esmaeilion, M. El Haj Assad, T. Hajilounezhad, A. Hmida, M. A. Rosen *et al.*, "A conceptual review of sustainable electrical power generation from biogas," *Energy Science & Engineering*, vol. 10, no. 2, pp. 630–655, 2022.

[11] Y. Zhang, L. Li, Z. Ren, Y. Yu, Y. Li, J. Pan, Y. Lu, L. Feng, W. Zhang, and Y. Han, "Plant-scale biogas production prediction based on multiple hybrid machine learning technique," *Bioresource Technology*, vol. 363, p. 127899, 2022.

[12] F. Tufaner and Y. Demirci, "Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models," *Clean Technologies and Environmental Policy*, vol. 22, pp. 713–724, 2020.

[13] T. Proietti and H. Lütkepohl, "Does the box–cox transformation help in forecasting macroeconomic time series?" *International Journal of Forecasting*, vol. 29, no. 1, pp. 88–99, 2013.

[14] T. Zhang and B. Yang, "Box–cox transformation in big data," *Technometrics*, vol. 59, no. 2, pp. 189–201, 2017.

[15] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE transactions on neural networks*, vol. 22, no. 3, pp. 337–346, 2010.

[16] M. A. Hosen, A. Khosravi, H. D. Kabir, M. Johnstone, D. Creighton, S. Nahavandi, and P. Shi, "Nn-based prediction interval for nonlinear processes controller," *International Journal of Control, Automation and Systems*, vol. 19, no. 9, pp. 3239–3252, 2021.

[17] N. T. Runyowa and W. Fourie, "Responsible leadership and the implementation of sdg 7: the case of the undp botswana biogas project," in *Sustainable Development Goals for Society Vol. 2: Food security, energy, climate action and biodiversity*. Springer, 2021, pp. 91–105.

[18] D. De Clercq, D. Jalota, R. Shang, K. Ni, Z. Zhang, A. Khan, Z. Wen, L. Caicedo, and K. Yuan, "Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale chinese production data," *Journal of cleaner production*, vol. 218, pp. 390–399, 2019.

[19] M.-C. Chiu, C.-Y. Wen, H.-W. Hsu, and W.-C. Wang, "Key wastes selection and prediction improvement for biogas production through hybrid machine learning methods," *Sustainable Energy Technologies and Assessments*, vol. 52, p. 102223, 2022.

[20] T. Singh and R. V. Uppaluri, "Feed-forward ann and traditional machine learning-based prediction of biogas generation rate from meteorological and organic waste parameters," *The Journal of Supercomputing*, vol. 80, no. 2, pp. 2538–2571, 2024.

[21] N. Amran, H. Mohamed, Z. Rafaai, N. Yacob, H. Junoh, and A. Shamsuddin, "A review of machine learning models in predicting biogas production," in *AIP Conference Proceedings*, vol. 2934, no. 1. AIP Publishing, 2024.

[22] A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "A new fuzzy-based combined prediction interval for wind power forecasting," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 18–26, 2015.

[23] Z. Zhang, H. Qin, L. Yao, J. Lu, and L. Cheng, "Interval prediction method based on long-short term memory networks for system integrated of hydro, wind and solar power," *Energy Procedia*, vol. 158, pp. 6176–6182, 2019.

[24] C. Voyant, G. Notton, J.-L. Duchaud, J. Almorox, and Z. M. Yaseen, "Solar irradiation prediction intervals based on box–cox transformation and univariate representation of periodic autoregressive model," *Renewable Energy Focus*, vol. 33, pp. 43–53, 2020.

[25] R. M. Sakia, "The box-cox transformation technique: a review," *Journal of the Royal Statistical Society Series D: The Statistician*, vol. 41, no. 2, pp. 169–178, 1992.

[26] A. F. da Costa and A. F. Crepaldi, "The bias in reversing the box–cox transformation in time series forecasting: An empirical study based on neural networks," *Neurocomputing*, vol. 136, pp. 281–288, 2014.

[27] H. Vinutha, B. Poornima, and B. Sagar, "Detection of outliers using interquartile range technique from intrusion dataset," in *Information and decision sciences: Proceedings of the 6th international conference on ficta*. Springer, 2018, pp. 511–518.

[28] A. Khosravi and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 849–856, 2013.