

Article

Priority-Weighted Clustering and Prediction Intervals for AI-Driven Biogas Energy Forecasting

Mohammad Anwar Hosen ^{*,†}, Nazmus Sakib [†], Michael Johnstone, Burhan Khan and Douglas Creighton

Institute for Intelligent Systems, Deakin University, Geelong, VIC 3216, Australia;
s.sakib@deakin.edu.au (N.S.); michael.johnstone@deakin.edu.au (M.J.); burhan.khan@deakin.edu.au (B.K.);
douglas.creighton@deakin.edu.au (D.C.)

* Correspondence: anwar.hosen@deakin.edu.au

[†] These authors contributed equally to this work.

Abstract

Biogas plays an increasingly important role in advancing decarbonisation goals, particularly in rural regions where livestock and agricultural waste can be converted into renewable energy. However, predicting annual farm-level biogas electricity output remains challenging due to operational variability, uncertain co-digestion practices, and external constraints such as grid integration and environmental disruptions. Traditional point prediction models often provide limited support for practical decision-making because they do not explicitly represent uncertainty around the estimated output. To address this limitation, this paper proposes a data-driven prediction interval framework for annual farm-level biogas electricity-output estimation. The study does not model future temporal horizons; rather, it estimates annual electricity generation at the farm level and constructs prediction intervals around these estimates. The proposed framework combines Self-Organising Map (SOM)-based clustering with Lower Upper Bound Estimation (LUBE) to account for operational heterogeneity among biogas-producing farms. SOM clustering is performed using pre-prediction operational covariates, specifically cattle population and co-digestion status, while electricity output is used only as the prediction target and for post-hoc interpretation. For each operational cluster, a neural prediction interval model is trained using the LUBE approach. A priority-weighted extension is then incorporated to reflect cluster-level operational or strategic importance in the evaluation of interval performance. Experiments on a real-world dataset from U.S. biogas systems show that the proposed framework can support a trade-off between interval width and coverage performance across standard confidence levels. By combining operational clustering with uncertainty-aware interval estimation, the method improves the interpretability and practical relevance of annual farm-level biogas electricity-output prediction for infrastructure planning.



Academic Editors: Akshay Kumar Saha, Francesco Riganti Fulginei and Mamadou Bailo Camara

Received: 1 June 2026

Revised: 14 July 2026

Accepted: 24 July 2026

Published: 3 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

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

Keywords: prediction intervals; biogas energy output prediction; renewable energy uncertainty; self-organising map clustering; uncertainty quantification

1. Introduction

In rural regions, a biogas plant manager often faces a continual balancing act between resource availability and environmental compliance. One month, the digester performs well thanks to steady cattle feedstock and mild weather. The next, unpredictable co-digestion inputs or rainfall disruptions can swing production by 30%, just as government subsidy requirements shift without notice [1]. Meanwhile, national energy planners tasked with

hitting carbon-neutrality targets face decisions on where to invest limited infrastructure funds. Should they expand high-output plants, or stabilise smaller, volatile ones? In this landscape, estimating biogas energy is not merely a technical task, it's a high stakes decision making problem with environmental, economic, and regulatory consequences.

Biogas is increasingly recognised as a critical pillar of the global decarbonisation effort, converting organic waste into renewable electricity while reducing methane emissions [2,3]. Produced through the anaerobic digestion of livestock manure [4], crop residues, and organic food waste, biogas systems reduce methane emissions while contributing clean energy to rural and peri-urban grids [5,6]. Their scalable nature makes them particularly valuable in agricultural economies seeking low-carbon, circular energy solutions [7].

In agricultural nations like Australia, which hosts over 70,000 cattle farms [8,9], biogas offers a unique opportunity for both waste management and distributed energy production. Yet fewer than 10% of these farms are integrated into the energy grid, and predicting their electricity output remains highly uncertain [10]. Co-digestion variability, weather sensitivity, and operational inconsistencies make it difficult to plan reliably [11]. Even small forecasting errors can cascade into policy missteps, underestimations lead to missed investment opportunities [12], while overestimations can result in unmet grid commitments and wasted subsidies [13].

Australia is discussed here as a motivating example of the broader potential and challenges of biogas energy within agricultural economies. However, the empirical analysis presented in this study, including the clustering, prediction interval modelling, and performance evaluation, is conducted using publicly available livestock-based biogas data collected from farms in the United States.

Traditional forecasting methods, such as linear regression, ARIMA, and basic neural networks focus [14] on single-point predictions [15,16]. These models are efficient but blind to uncertainty, offering no indication of how much actual output may deviate from expected values. As a result, they provide little support for decisions that depend on risk awareness, such as grid stabilisation planning, subsidy allocation [17], or environmental risk management [18].

As illustrated in Figure 1, annual farm-level biogas electricity-output prediction is influenced by multiple sources of uncertainty, including operational variability, environmental factors, and policy-related disruptions, all of which significantly impact forecasting reliability and decision-making.

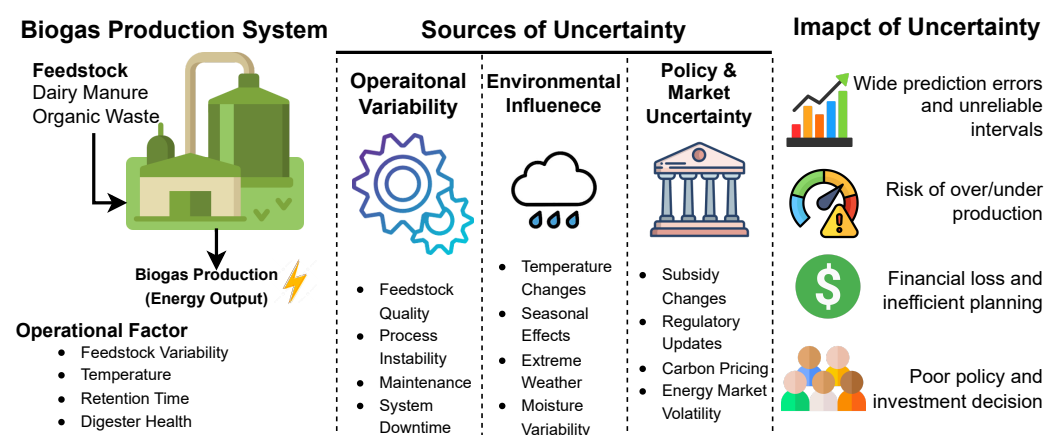


Figure 1. Overview of annual farm-level biogas electricity-output prediction, showing operational inputs, uncertainty sources, and prediction interval-based decision support.

To address these limitations, the research community has turned to Prediction Intervals (PIs), which define upper and lower bounds around forecasts with a target confidence level [19]. One of the most promising frameworks for PI generation is the Lower Upper Bound Estimation (LUBE) method, which uses neural networks to optimise interval bounds based on metrics such as Prediction Interval Coverage Probability (PICP) and Prediction Interval Normalised Average Width (PINAW) [20,21].

While PICP ensures that the predicted intervals capture the true values with a desired probability and PINAW evaluates the sharpness of these intervals, both metrics fail to quantify the severity of prediction failures when observations fall outside the estimated bounds [22]. In practical energy-output prediction settings, particularly in biogas systems where operational decisions depend on risk magnitude, it is crucial to distinguish between minor deviations and extreme forecasting errors [23].

To overcome this limitation, the Prediction Interval Normalised Average Failure Distance (PINAFD) has been introduced as an additional evaluation metric, measuring the magnitude of deviation for samples that fall outside the prediction interval [22]. By incorporating failure distance into the evaluation framework, PINAFD provides a more comprehensive understanding of uncertainty, enabling models to account not only for coverage and width but also for the severity of prediction errors.

However, most LUBE-based models [2] assume uniform behavior across all biogas systems, ignoring the rich operational diversity between farms. Additionally, they often neglect non-statistical uncertainties, such as policy volatility and environmental disruptions, which significantly influence real-world performance [24].

Forecasting biogas energy production is further complicated by its highly variable nature [6]. Yield is influenced by numerous interdependent factors, such as feedstock type and volume, digester temperature, microbial health, and the presence of co-digestion practices [2]. Two farms with identical cattle populations may produce vastly different energy outputs due to differences in digestion management, weather conditions, or local grid dynamics [4]. These uncertainties challenge the assumption of uniformity across farms and make traditional point forecasting methods unreliable for infrastructure planning [24].

To address these challenges, we propose an uncertainty-aware farm-level prediction interval framework that estimates biogas electricity output using available operational data and accounts for system-level uncertainty factors [25]. The approach combines clustering, neural modelling, and cost-sensitive learning to deliver forecasts that reflect both statistical uncertainty and broader infrastructural considerations.

The key contributions of this study are:

- A baseline LUBE model is developed to establish reference performance without behavioural segmentation.
- Self-Organising Maps (SOM) are applied to cluster farms based on cattle population and co-digestion practices, enabling behaviour-specific prediction interval estimation.
- Cluster-specific LUBE models are trained and evaluated to enhance interval reliability by capturing intra-group variability.
- A priority-weighted extension of the CWC criterion is introduced, representing the primary theoretical contribution of this work by integrating cluster-level operational significance into the evaluation of prediction interval performance.

To address the limitations of traditional point forecasting methods and the need for context-aware uncertainty quantification in heterogeneous biogas systems, this study seeks to answer the following research question:

How can prediction interval estimation for annual farm-level biogas electricity output be enhanced through behavioral clustering and context-aware optimisation to provide operationally meaningful uncertainty quantification for biogas energy systems?

Answering this question requires not only improving traditional statistical measures of prediction interval quality but also incorporating operational significance into the evaluation of interval performance, thereby enabling prediction models that are better aligned with real-world planning and decision-making requirements.

The remainder of the paper is structured as follows. Section 2 reviews relevant work in forecasting, clustering, and uncertainty modelling. Section 3 presents the proposed framework, including data preprocessing, SOM based clustering, prediction interval construction, and the priority weighted cost function. Section 4 discusses experimental results and comparative analyses. Section 5 concludes the study and outlines future research directions.

2. Related Work

Accurate forecasting of biogas production remains an open challenge due to the nonlinear, multi-modal, and highly uncertain nature [26] of the digestion process [15]. Biological complexity, environmental variability [4], and operational inconsistency [5] combine to make reliable energy estimation difficult, especially in systems that incorporate co-digestion practices or are exposed to fluctuating climatic and policy environments [2,11,13]. These challenges are particularly acute in planning contexts, where overestimation can strain grid capacity and underestimation may lead to underinvestment or infrastructure gaps [3,5].

2.1. Prediction Intervals and the LUBE Framework

Conventional forecasting approaches in biogas systems often provide only point estimates, offering no indication of how much actual output may vary around the prediction [18]. This becomes a notable limitation in decision-making contexts, particularly in energy policy, infrastructure investment, and grid integration, where stakeholders need to plan under uncertainty [24].

To address this gap, prediction interval (PI) methods have gained traction in renewable energy forecasting [27]. A prediction interval provides a bounded region $[L_i, U_i]$ such that the true target value y_i falls within this range with a given confidence level ϕ , typically 90% or 95% [21]. Formally, this is expressed as:

$$P(y_i \in [L_i, U_i]) \approx \phi \quad (1)$$

The goal is to construct intervals that are both *reliable*, meaning they contain the true value with high probability, and *sharp*, meaning they are not excessively wide. A widely adopted method for generating PIs is the LUBE framework, which employs neural networks to learn the bounds of the prediction interval directly [28]. In assessing the quality of these intervals, three core metrics are commonly used: Prediction Interval Coverage Probability (PICP), Prediction Interval Normalised Average Width (PINAW), and Coverage Width-based Criterion (CWC) [19,20].

2.2. Clustering for Behavioral Heterogeneity

In biogas systems, farms vary widely in terms of herd size, co-digestion strategies, and operational practices factors [29] that can influence energy output dramatically [6]. Applying a single model across all data fails to capture this heterogeneity, often leading to poorly generalised forecasts [3]. Clustering offers a pathway to address this by grouping similar operational patterns, allowing tailored models for each group [7].

Among clustering techniques, SOM [30] have gained traction in energy analytics for their ability to project high-dimensional data onto a lower-dimensional grid while

preserving topological relationships [31,32]. Given a dataset $D = \{(x_i, y_i)\}_{i=1}^n$, SOMs can partition the data into k clusters:

$$D = \bigcup_{j=1}^k D_j$$

Each cluster D_j can then be modeled separately using its own LUBE architecture, producing local prediction intervals $[L_i^j, U_i^j]$. This structure-aware modelling enhances both interpretability and accuracy in heterogeneous environments.

2.3. Gaps and Research Motivation

Although LUBE-based techniques have improved over time, their practical applicability remains limited due to unresolved challenges in modelling adaptation and external uncertainty integration. First, most LUBE implementations adopt a uniform forecasting approach, overlooking cluster-specific variations in system behaviour. Second, even advanced prediction interval models often fail to incorporate external uncertainties such as grid integration constraints, environmental disruptions, or economic variability. These omissions reduce the operational relevance of forecasting models, particularly in contexts where infrastructure decisions must be made under limited resources and uncertain conditions [27].

Our research addresses these limitations by combining unsupervised clustering with localised prediction interval modelling, and by introducing a cluster-level priority adjustment within the LUBE-based CWC criterion. This integrated approach supports prediction interval estimation that is statistically meaningful and contextually aligned with farm-level operational needs.

3. Methodology

This section presents the complete pipeline of the proposed annual farm-level electricity-output prediction framework, designed to generate uncertainty-aware prediction intervals that are both statistically reliable and operationally meaningful. The framework integrates baseline prediction interval modelling, behavior-driven clustering, and priority-weighted CWC evaluation to address the inherent variability and heterogeneity of biogas systems. By combining data-driven learning with contextual prioritisation, the approach aims to improve both predictive robustness and real-world applicability.

3.1. Proposed Methodological Overview

As illustrated in Figure 2, the proposed framework represents the core contribution of this study and follows a structured three-phase pipeline. In the first phase, a baseline LUBE model is trained on the full dataset to establish a reference for prediction interval quality. In the second phase, Self-Organising Map (SOM)-based clustering is employed to capture operational heterogeneity across biogas systems, enabling the development of cluster-specific models. In the final phase, a priority-weighted CWC criterion is introduced, where contextual importance is incorporated as a cluster-level adjustment to support priority-aware evaluation of prediction interval performance.

This hierarchical design allows the framework to progressively refine prediction interval quality, transitioning from a global baseline model to localised cluster-specific learning, and ultimately to a context-aware optimisation scheme that aligns forecasting performance with real-world operational priorities.

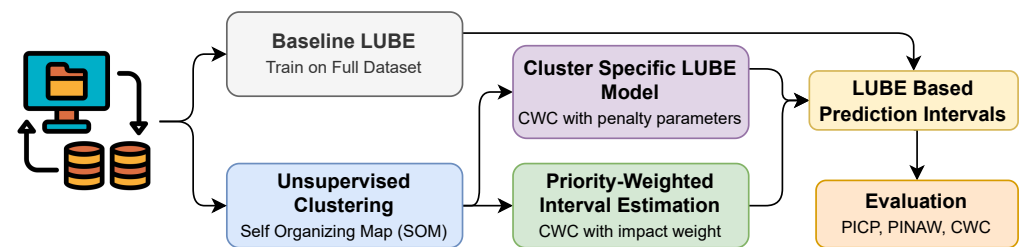


Figure 2. Conceptual overview of the proposed three-phase framework for uncertainty-aware annual farm-level biogas electricity-output prediction.

Figure 3 provides a detailed workflow representation of the implementation process. It outlines the sequential steps involved in data preprocessing, baseline model training, SOM-based clustering, cluster-specific learning, and evaluation using prediction interval metrics. While Figure 2 presents the conceptual design and methodological contribution, Figure 3 highlights the operational execution of the proposed system, ensuring clarity in both high-level understanding and practical implementation.

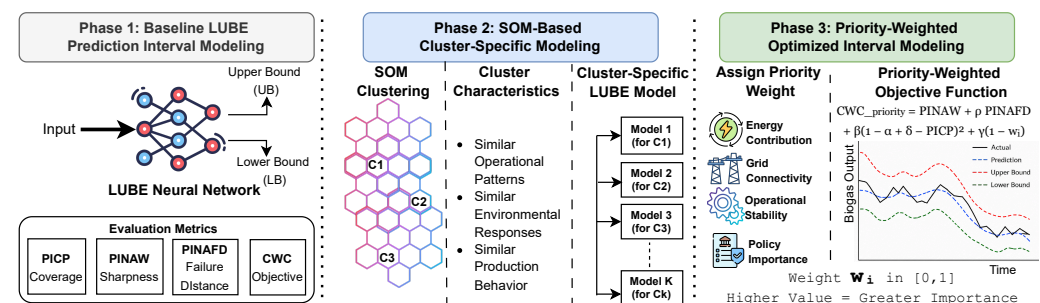


Figure 3. Detailed implementation workflow illustrating data processing, clustering, model training, and evaluation procedures.

3.2. Dataset Description

The dataset used in this study is sourced from livestock-based biogas production systems in the United States, covering a full year of operations. It includes data from 741 farms managing various animal types, including cattle, dairy, poultry, and swine. For this study, we focused on 644 farms that primarily raise cattle, representing the largest group in terms of both population and energy potential. These farms are geographically spread across states such as Texas, New Mexico, Idaho, and Pennsylvania [2]. After removing incomplete records for the selected modelling variables, 594 cattle-farm records were retained for model development, including 295 without co-digestion and 299 with co-digestion [33].

The present study does not perform temporal forecasting of future time-series observations. Instead, it focuses on predicting annual electricity generation output at the farm level using operational characteristics such as cattle population and co-digestion practices. Accordingly, the term *forecasting* in this study refers to the predictive estimation of farm-level energy production and its associated prediction intervals, rather than the forecasting of sequential future values over time. The input variables are cattle population and co-digestion status, while the target variable is annual electricity generation. The prediction unit is an individual farm, and no future temporal horizon is modeled.

To better capture operational distinctions, we categorised the farms based on their use of digestion technology. Specifically, 295 farms do not use co-digestion (i.e., they rely solely on cattle manure), while 299 farms actively incorporate co-digestion techniques by mixing additional organic inputs. The data include electricity generation values (in kWh/year),

cattle size, and co-digestion usage. These features display non-linear patterns and skewed distributions, necessitating advanced preprocessing.

Table 1 presents representative sample farms selected to illustrate the diversity of operational conditions within the dataset, including variations in cattle size, co-digestion practices, and annual electricity generation. These examples are provided for illustrative purposes only and do not correspond to any specific experimental subset.

Table 1. Sample data from US cattle farms dataset.

Location	Cattle Size	Co-Digestion	Electricity (kWh/year)
TX	8700	No	11,800,000
ID	6100	Yes	10,500,000
NM	7800	No	12,300,000

3.3. Phase 1: Baseline Prediction Interval Modelling

This initial phase establishes a reference using the LUBE framework [21]. A single feedforward neural network is trained on the full dataset to directly estimate lower and upper prediction bounds [27].

The reliability of the intervals is assessed using Prediction Interval Coverage Probability (PICP), which quantifies the proportion of actual values falling within predicted bounds:

$$\text{PICP} = \frac{1}{n} \sum_{i=1}^n I(y_i \in [L_i, U_i]) \quad (2)$$

where $I(\cdot)$ is an indicator function, and n is the number of observations. Prediction Interval Normalised Average Width (PINAW) evaluates the sharpness of intervals:

$$\text{PINAW} = \frac{1}{nR} \sum_{i=1}^n (U_i - L_i) \quad (3)$$

where R denotes the range of the target variable. Smaller PINAW values are preferred if adequate coverage is maintained.

In addition to coverage and width, it is important to quantify the magnitude of prediction failures when observations fall outside the estimated bounds. For this purpose, we incorporate the Prediction Interval Normalised Average Failure Distance (PINAFD), which captures the average deviation of missed samples from the predicted interval [22].

$$\text{PINAFD} = \frac{1}{nR} \sum_{i=1}^n F_i \quad (4)$$

where the failure distance F_i is defined as:

$$F_i = \begin{cases} y_i - U_i, & \text{if } y_i > U_i \\ L_i - y_i, & \text{if } y_i < L_i \\ 0, & \text{otherwise} \end{cases}$$

Unlike PICP and PINAW, PINAFD explicitly captures the magnitude of deviations outside the prediction bounds, providing a more comprehensive evaluation of interval reliability under uncertainty.

To jointly optimise coverage, interval width, and failure magnitude, we define a modified Coverage Width-based Criterion (CWC) that integrates PINAFD:

$$\text{CWC} = \text{PINAW} + \rho \cdot \text{PINAFD} + \beta \cdot \left((1 - \alpha + \delta - \text{PICP})^2 \right) \quad (5)$$

where ρ controls the influence of failure distance and was set to $\rho = 1.25$ in the experiments, β is a penalty coefficient enforcing coverage reliability, α is the target non-coverage rate (e.g., $\alpha = 0.05$ for 95% confidence), and δ is a small tolerance term to account for minor deviations.

The LUBE model was implemented using a feedforward neural network with a single hidden layer consisting of 10 neurons. The data were randomly divided into training (80%), validation (10%), and testing (10%) subsets. Prior to model training, the input and output variables were normalised using the `mapminmax` transformation. To avoid data leakage, all data-dependent preprocessing, clustering, and priority-weighting steps were fitted or computed using the training set only. Specifically, the `mapminmax` normalisation parameters were estimated from the training set and then applied unchanged to the validation and test sets. The SOM was trained only on the training-set operational covariates, namely cattle population and co-digestion status, and validation/test samples were assigned to the trained SOM using the same covariates. Cluster-level priority weights were computed using only the training partition and were then kept fixed during validation and testing.

The same train, validation, and test indices were used consistently across all model configurations to ensure fair comparison of the evaluated approaches. The network was initially trained using the Levenberg–Marquardt algorithm and subsequently refined using Simulated Annealing to optimise the prediction interval objective function. Experiments were conducted at 90% and 95% confidence levels, corresponding to $\alpha = 0.10$ and $\alpha = 0.05$, respectively.

In the modified CWC objective, the penalty coefficient was set to $\beta = 1000$, the tolerance parameter was set to $\delta = 0.001$, and the failure-distance coefficient was set to a positive penalty value of $\rho = 1.25$, ensuring that the PINAFD term remained active in the modified CWC objective. For the priority-weighted formulation, the weighting penalty coefficient was set to $\gamma = 10$.

The use of a positive ρ value ensures that the PINAFD term remains active in the modified CWC objective and penalises the magnitude of deviations for observations falling outside the prediction intervals. This formulation extends the traditional LUBE objective by explicitly penalising large prediction errors outside the interval bounds, thereby improving robustness in high-uncertainty environments such as biogas energy forecasting. This baseline model provides a statistically robust benchmark but does not account for external contextual uncertainties such as operational variability or regulatory constraints addressed in the subsequent phases.

3.4. Phase 2: Cluster-Based Interval Modelling

Biogas production systems exhibit operational diversity that a single model may fail to capture. To address this, unsupervised clustering was performed using Self-Organising Maps (SOM) [30]. The SOM was fitted using only pre-prediction operational covariates, namely, cattle population and co-digestion status, to group farms into behaviourally distinct operational clusters [31]. Annual electricity output was not used as an input to the SOM because it is the prediction target. Instead, electricity output was used only for prediction, evaluation, and post-hoc interpretation of the resulting cluster characteristics. Each cluster therefore reflects differences in available farm-level operational conditions rather than direct grouping by the target variable [34].

Rationale for Cluster Selection. The selection of clusters was guided by both visual and practical considerations. As illustrated in Figure 4, the SOM-based clustering with two groups provides a clear and meaningful separation between small-to-medium scale farms and larger facilities. Cluster 1 (red) captures farms with lower cattle counts and correspondingly low electricity output, while Cluster 2 (green) includes higher-yield systems, often employing co-digestion strategies.

In contrast, the three-cluster configuration shown in Figure 4 introduces a third group (blue) that overlaps substantially with Cluster 2 when the resulting clusters are visualised against cattle count and annual electricity output for post-hoc interpretation. This overlap reduces the distinctiveness of clusters, making it difficult to justify separate model training without introducing overfitting or sparsity issues, especially given the limited dataset size. Additionally, cluster 3 contains relatively few points, diluting its statistical significance for robust training.

Therefore, the two-cluster approach was chosen to ensure model reliability, interpretability, and stability, especially in low-data regimes. This segmentation into low and high output groups also facilitates more targeted interval modelling and strategic planning, such as prioritised resource allocation and uncertainty-sensitive forecasting.



Figure 4. Comparison of SOM Clustering with 2 Clusters (left) and 3 Clusters (right). Overlap in the 3-cluster version introduces ambiguity and sparsity.

Each cluster is then modelled independently using the LUBE framework, allowing the neural network to tailor interval estimates to specific operational patterns [17,28].

For each cluster, an independent LUBE model is trained, enabling the prediction intervals to adapt to localised patterns of uncertainty. This cluster-specific modelling improves interval sharpness and relevance by accounting for intra-group variability in production behavior.

3.5. Phase 3: Priority-Weighted Optimised Interval Modelling

While statistical accuracy is central to prediction interval modelling, real-world energy systems often require prioritisation based on factors such as infrastructure scale, contribution to the grid, and operational risk [35]. In biogas systems, clusters associated with large-scale output or higher operational stability may require greater emphasis during interval evaluation. Ignoring such differences can lead to suboptimal interval construction, even when standard statistical metrics appear satisfactory [36].

To address this, we introduce a priority-weighted extension to the LUBE objective function, incorporating cluster-level importance into the priority-weighted CWC criterion. Let each cluster i be assigned a weight $w_i \in [0, 1]$, representing its relative importance in the forecasting task. These weights may be derived from domain-specific indicators such

as total electricity contribution, grid connectivity, or operational reliability [37]. Higher values of w_i correspond to clusters of greater strategic relevance [38].

In this study, total electricity contribution was selected as the weighting criterion because it directly reflects the relative importance of each cluster to overall energy production and potential grid contribution. Therefore, the priority weights should be interpreted as target-derived cluster-level importance indicators for policy and operational weighting, rather than as purely input-driven prediction features.

This choice enables cluster-level interval evaluation to reflect the relative operational importance of different groups within the energy system. Nevertheless, alternative weighting schemes may be adopted depending on the application context, including weights based on grid connectivity, carbon reduction potential, economic value, operational stability, or policy-driven priorities.

To avoid the use of test information, the cluster-level priority weights are computed from the training set only and then kept fixed during validation and testing. For an unseen farm, the SOM assigns the farm to a cluster using available operational covariates, and the corresponding pre-computed cluster weight is applied. In this manner, the weighting mechanism does not require access to unseen target values during deployment.

Building upon the modified multi-objective formulation introduced in Section 3.3, the proposed priority-weighted CWC criterion is defined as:

$$\text{CWC}_{\text{priority}} = \text{PINAW} + \rho \cdot \text{PINAfD} + \beta \cdot \left((1 - \alpha + \delta - \text{PICP})^2 \right) + \gamma(1 - w_i) \quad (6)$$

where ρ controls the contribution of failure distance, β enforces coverage reliability, α is the target non-coverage rate, and δ is a tolerance parameter. The additional term $\gamma(1 - w_i)$ introduces a cluster-dependent penalty, where γ regulates the strength of prioritisation.

It should be noted that the priority term $\gamma(1 - w_i)$ operates at the cluster level once the priority weight w_i has been assigned. Therefore, this term should be interpreted as a cluster-level priority adjustment to the CWC criterion rather than as a sample-level adaptive gradient term. The primary interval construction remains governed by the standard interval-quality components, namely PICP, PINAW, and PINAFD, while the priority term allows the final criterion to reflect differences in operational or strategic relevance across heterogeneous clusters.

By integrating statistical objectives (coverage, width, and failure magnitude) with domain-driven prioritisation, the proposed framework supports priority-aware evaluation of interval performance across heterogeneous clusters. The priority term operates at the cluster level and allows the CWC criterion to reflect operational significance, while the underlying interval construction remains governed by PICP, PINAW, and PINAFD. Unlike post-hoc reweighting [39] or rule-based approaches [40], the proposed formulation provides a cluster-level priority adjustment within the CWC criterion. Future work will investigate direct sample-level weighting of these components to further strengthen the optimisation role of priority information.

The next section evaluates the proposed framework across three configurations: baseline LUBE modelling, SOM-based cluster-specific modelling, and the proposed priority-weighted CWC criterion. This design enables a comprehensive comparison of both statistical performance and contextual relevance.

4. Experimental Results and Discussion

The experiments were conducted across three configurations: (i) Baseline LUBE applied to the full dataset, (ii) cluster-specific LUBE using two clusters generated by SOM, and (iii) priority-weighted LUBE, where cluster-level importance weights were incorpo-

rated into the CWC criterion. Each configuration was evaluated at two confidence levels: 90% and 95%. Performance was assessed using standard prediction interval metrics, PICP, PINAW, CWC, and Root Mean Squared Error (RMSE). The objective was to evaluate not only statistical reliability but also contextual alignment with operational relevance.

4.1. Performance Comparison Across Modelling Strategies

Table 2 summarises the evaluation metrics across the three modelling phases: Baseline (no clustering or weighting), clustered (with behavioral segmentation via SOM), and priority-weighted (clustered models evaluated with cluster-level importance weights). The baseline phase provides a reference for interval performance on the full dataset. Cluster-specific models capture local operational patterns and show different coverage-width behaviour across farm groups. The priority-weighted configuration reduced PINAW and CWC in several cases; however, some configurations did not reach the nominal coverage level. Therefore, these results should be interpreted as a trade-off among interval width, coverage probability, and the priority-adjusted CWC criterion rather than as an unconditional improvement in reliability.

For interval-based models, RMSE was computed using the midpoint of the predicted interval as the representative point estimate. Specifically, the point prediction was calculated as $(L_i + U_i)/2$, where L_i and U_i denote the lower and upper prediction bounds, respectively. This midpoint estimate was then used to calculate RMSE against the observed electricity generation values.

Table 2. Summary of performance across experimental phases.

Phase	Confidence	PICP (%)	PINAW	CWC	RMSE
Baseline	90%	90.14	27.39	65.98	36.83
Baseline	95%	95.22	44.44	79.78	36.88
Clustered (C1)	90%	87.80	36.30	78.98	24.94
Clustered (C1)	95%	95.12	49.87	49.67	25.21
Clustered (C2)	90%	83.33	48.71	86.59	54.06
Clustered (C2)	95%	94.44	69.71	74.00	84.29
Priority-Weighted (C1)	90%	85.37	28.66	33.11	20.93
Priority-Weighted (C1)	95%	92.68	36.53	40.98	25.02
Priority-Weighted (C2)	90%	94.44	28.82	33.27	52.61
Priority-Weighted (C2)	95%	93.54	59.06	63.51	55.98

Although priority weighting reduced CWC and PINAW in several cases, some configurations did not fully achieve the nominal coverage level. For example, the Priority-Weighted C1 model achieved PICP values of 85.37% and 92.68% at the 90% and 95% confidence levels, respectively, while the Priority-Weighted C2 model achieved 93.54% at the 95% confidence level. Therefore, the results should be interpreted as demonstrating a trade-off between sharper prediction intervals and coverage reliability rather than an unconditional improvement across all evaluation metrics.

Although priority weighting reduced CWC and PINAW in several cases, some configurations did not fully achieve the nominal coverage level. For example, the Priority-Weighted C1 model achieved PICP values of 85.37% and 92.68% at the 90% and 95% confidence levels, respectively, while the Priority-Weighted C2 model achieved 93.54% PICP at the 95% confidence level. These values indicate that the priority-weighted results should be interpreted as a coverage-width trade-off rather than as a general improvement in reliability. In particular, reduced CWC and narrower intervals must be considered jointly with whether the achieved PICP satisfies the intended nominal coverage.

Figure 5 illustrates the coverage-width behaviour of the priority-weighted configuration by comparing PICP and CWC across clusters and confidence levels.

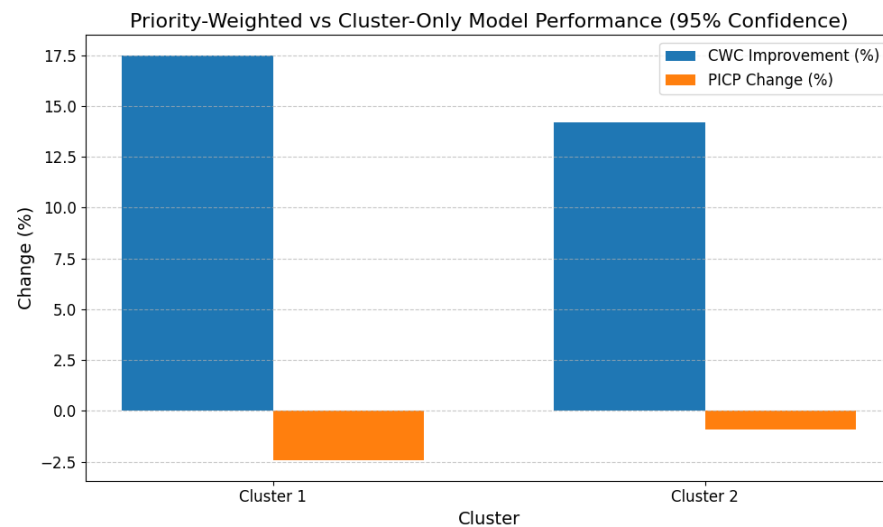


Figure 5. Cluster-wise PICP and CWC behaviour under the priority-weighted configuration.

4.2. Prediction Interval Comparison Across Modelling Phases

Figure 6 presents a comprehensive comparison of the prediction intervals generated under three modelling configurations: baseline LUBE (no clustering), cluster-based LUBE (with SOM segmentation), and priority-weighted LUBE (cluster-specific modelling with importance-based cost adjustments).

Each row of the figure corresponds to a different modelling strategy, while the two columns compare intervals at 90% and 95% confidence levels. The baseline configuration exhibits wider intervals and signs of under-coverage in high-variance regions. Cluster-based LUBE improves calibration and sharpness by tailoring the prediction intervals to behavioural groups.

Finally, the priority-weighted LUBE model incorporates cluster-level importance weights into the CWC criterion. The resulting intervals should be interpreted in terms of the trade-off between interval compactness and achieved coverage, particularly because some priority-weighted configurations do not fully reach the nominal coverage level.

In the priority-weighted modelling phase, cluster-level importance was incorporated into the priority-weighted CWC criterion using a normalised weighting scheme. The priority weight w_i for each cluster was computed based on its contribution to the total electricity output, ensuring:

$$\sum_{i=1}^K w_i = 1, \quad \text{where} \quad w_i = \frac{\text{Electricity from Cluster } i}{\text{Total Electricity from All Clusters}}$$

This formulation incorporates output-driven importance into the priority-weighted CWC criterion. By assigning lower priority penalties to higher-output clusters (i.e., higher w_i), the criterion reflects the greater operational relevance of these clusters. The resulting interval performance should be interpreted jointly using PICP, PINAW, PINAFD, and CWC.

For example, Cluster 1 contributed approximately 75% of the total electricity in the dataset, while Cluster 2 accounted for the remaining 25%. Accordingly, the priority weights were derived as $w_1 = 0.75$ and $w_2 = 0.25$, normalised to satisfy $\sum w_i = 1$. These weights were applied in the priority-weighted CWC criterion (see Equation (6)), with the penalty scaling factor γ set to 10. This setup assigns a lower cluster-level priority penalty to high-

yield clusters such as Cluster 1, while lower-impact clusters such as Cluster 2 receive a proportionally stronger priority adjustment.

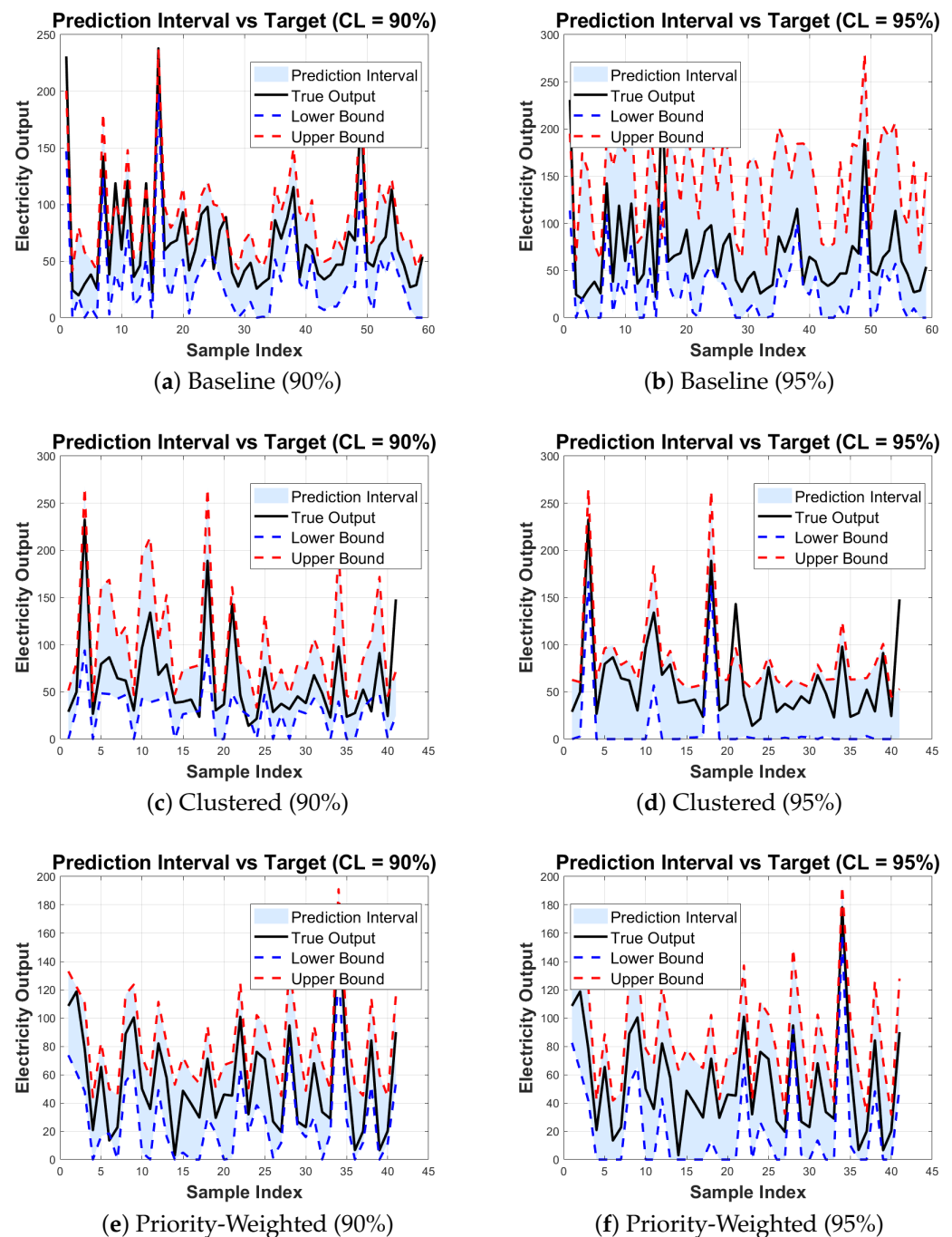


Figure 6. Prediction interval visualisation across baseline, clustered, and priority-weighted modelling strategies under 90% and 95% confidence levels.

4.3. Discussion

This study addressed the dual challenge of constructing statistically meaningful prediction intervals (PIs) while incorporating operational relevance through data-driven prioritisation. The proposed multi-phase framework combines Self-Organising Map (SOM) clustering with a priority-weighted CWC criterion to account for heterogeneity among farm-level biogas systems.

The SOM-based clustering identified two operational groups: cattle-only farms with lower output characteristics and larger systems associated with co-digestion practices. By training specialised models for each cluster, the framework produced cluster-specific prediction intervals with different coverage-width characteristics. The introduction of priority-based weighting allowed the evaluation criterion to reflect the relative operational importance of different clusters, while the resulting performance was interpreted jointly using PICP, PINAW, and CWC.

Cluster 1 ($w = 0.75$): After applying priority weighting, the Coverage Width Criterion (CWC) decreased from 78.98 in the unweighted cluster model to 33.11 at the 90% confidence level, while PINAW narrowed from 36.30 to 28.66 and RMSE decreased to 20.93. However, the corresponding PICP was 85.37%, which is below the nominal 90% level. This indicates improved interval compactness, but with a coverage trade-off.

Cluster 2 ($w = 0.25$): At the 90% confidence level, CWC declined from 86.59 to 33.27, while PICP increased from 83.33% to 94.44% and PINAW decreased from 48.71 to 28.82. This suggests a favourable balance between interval width and coverage for this configuration. However, at the 95% confidence level, the priority-weighted C2 model achieved 93.54% PICP, which remains slightly below the nominal level. Overall, the priority-weighted results indicate that interval compactness and CWC can improve in selected configurations, but these gains should be interpreted together with the achieved coverage level.

From a practical perspective, the proposed framework and findings provide a decision-support perspective for annual farm-level biogas electricity-output prediction under uncertainty. These results may help inform preliminary planning and policy analysis, particularly where farm-level operational characteristics and uncertainty profiles differ across production settings. The use of electricity-output-based weights further provides a way to incorporate cluster-level operational relevance into the interpretation of prediction interval performance. However, because these weights are derived from cluster-level electricity-output contribution, they should be interpreted as policy-oriented importance weights rather than as input-only predictive variables.

Despite these findings, several limitations should be acknowledged. First, the study is based on a static one-year dataset and therefore may not fully capture long-term environmental variations, policy changes, or technological degradation effects that can influence biogas energy production over multiple years. Future validation using multi-year datasets would provide additional insights into the long-term robustness and generalisability of the proposed framework.

Second, the dataset is relatively small and the feature space is limited, relying mainly on cattle population and co-digestion status as available operational covariates. Therefore, the proposed framework should be interpreted as an uncertainty-aware farm-level prediction interval approach under limited-data conditions rather than as a fully general AI-driven forecasting system. Although a train/validation/test split was used for model evaluation, future work should include repeated random splits, cross-validation, and multi-run statistical reporting to further assess robustness across different data partitions.

Third, the present comparison is limited to baseline, clustered, and priority-weighted variants of the LUBE framework. Therefore, the reported findings should be interpreted as improvements within this neural prediction interval modelling family rather than as evidence of superiority over all uncertainty-quantification methods. Broader comparisons with quantile regression, conformal prediction, conformalised quantile regression, quantile random forest or gradient boosting, neural quantile regression, and bootstrap-based intervals are left for future work, as these methods require a separate calibration and validation design, particularly for small farm-level datasets with limited feature space.

Finally, the current two-cluster framework may be less effective for edge-case farms that undergo rapid operational changes, such as substantial herd expansion or the adoption of co-digestion practices during the observation period, as such transitions may alter their cluster membership over time. Large-scale deployment may also introduce additional computational overhead associated with updating cluster assignments, retraining cluster-specific models, and maintaining interval evaluation across distributed energy networks. Future research should therefore investigate scalable implementations suitable for real-time operational environments.

In summary, the proposed framework provides a priority-sensitive and context-aware approach for annual farm-level biogas electricity-output prediction under uncertainty. By combining operational clustering with priority-weighted interval evaluation, the framework offers useful decision-support potential for biogas planning contexts where data availability is limited and uncertainty remains high.

5. Conclusions and Future Direction

This study introduced an uncertainty-aware prediction interval framework for annual farm-level biogas electricity generation estimation, combining unsupervised clustering, prediction interval modelling, and priority-weighted CWC evaluation. By leveraging SOM to identify behavioral subgroups within the dataset and training specialised LUBE models for each cluster, the framework produced cluster-specific prediction intervals with different coverage-width characteristics. A key enhancement was the introduction of a priority-weighted CWC criterion, which incorporated output-driven cluster significance into the evaluation of interval performance. Experimentally, the proposed priority-weighted framework reduced the CWC from 78.98 to 33.11 in Cluster 1 and from 86.59 to 33.27 in Cluster 2 at the 90% confidence level, while improving the PICP of Cluster 2 from 83.33% to 94.44%. However, the Cluster 1 priority-weighted model did not reach the nominal 90% coverage level. These results indicate that the proposed framework can improve interval compactness and CWC in selected configurations, but its benefits should be interpreted together with achieved coverage.

Future research should investigate the use of richer temporal-resolution datasets and suitable sequential modelling approaches to capture time-dependent variation in biogas production. Additional operational data, such as IoT or smart-metering measurements, may also support more adaptive analysis in future studies. Additionally, extending the priority-weighted evaluation criterion to incorporate environmental or economic metrics, such as carbon offsets or grid stability indicators, could provide richer insights for sustainability-aligned prediction. A limitation of the current priority-weighted formulation is that the priority term is applied at the cluster level after the priority weight is assigned. Future work will investigate direct sample-level or component-level weighting of PICP, PINAW, and PINAFD to further strengthen the optimisation role of priority information. Although the present empirical evaluation is based on annual farm-level electricity-output data, the proposed clustering-based prediction interval framework can be extended to weekly or monthly biogas output datasets when such temporal-resolution data are available.

Future studies may also explore spatial grid integration and land-use constrained planning for biogas deployment [41], advanced uncertainty modelling through kernel density estimation and stochastic optimisation techniques [42], and stochastic network planning approaches for evaluating the long-term value of biogas resources within smart energy systems under uncertainty [43]. Overall, this work provides a flexible and extensible platform for annual farm-level biogas electricity-output prediction under uncertainty, supporting informed planning in data-scarce and uncertain environments.

Author Contributions: Conceptualisation, N.S. and M.A.H.; methodology, N.S. and B.K.; software, N.S.; validation, M.A.H., B.K., and M.J.; formal analysis, N.S.; investigation, N.S.; data curation, N.S.; writing-original draft preparation, N.S.; writing-review and editing, M.A.H., M.J., B.K., and D.C.; visualization, N.S.; supervision, M.A.H., M.J., and D.C.; project administration, M.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study were obtained from previously published and publicly available biogas production datasets. Processed data and analysis results are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia, for providing the research environment and support necessary for conducting this study.

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

References

- Pilarski, K.; Pilarska, A.A.; Pietrzak, M.B. Biogas production in agriculture: Technological, environmental, and socio-economic aspects. *Energies* **2025**, *18*, 5844.
- Sakib, N.; Hosen, M.A.; Khan, B.; Gunn, B.; Johnstone, M. Enhancing Biogas Energy Production Forecasting with Box-Cox Transformation and Prediction Interval Methods. In *Proceedings of the 2025 IEEE International Systems Conference (SysCon), Montreal, QC, Canada*; IEEE, 2025; pp. 1–8.
- Salvador, R.; Barros, M.V.; Rosário, J.G.D.P.D.; Piekarski, C.M.; da Luz, L.M.; de Francisco, A.C. Life cycle assessment of electricity from biogas: A systematic literature review. *Environ. Prog. Sustain. Energy* **2019**, *38*, 13133.
- Feng, Y.; Shoaib, M.; Akram, R.; Alnafra, I.; Ai, F.; Irfan, M. Assessing and prioritizing biogas energy barriers: A sustainable roadmap for energy security. *Renew. Energy* **2024**, *223*, 120053.
- Wu, Y.; Xu, Y. A comprehensive evaluation of an environmentally friendly multi-heat recovery system using biogas energy source from thermodynamic, economic, and environmental perspectives. *J. Environ. Chem. Eng.* **2024**, *12*, 112783.
- Tufaner, F.; Demirci, Y. Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models. *Clean Technol. Environ. Policy* **2020**, *22*, 713–724.
- Bórawski, P.; Bědycka-Bórawska, A.; Kapsdorferová, Z.; Rokicki, T.; Parzonko, A.; Holden, L. Perspectives of Electricity Production from Biogas in the European Union. *Energies* **2024**, *17*, 1169.
- Zeunert, J. Bridging Rural and Urban Disconnections: Spatial Graphic Explorations of Australia's Livestock Landscapes. In *Australia and China Perspectives on Urban Regeneration and Rural Revitalization*; Routledge, 2024; pp. 110–143.
- Allen, J. *Development Strategies for Genetic Evaluation for Beef Production in Developing Countries*; Country Report: Australia; Australian Centre for International Agricultural Research, Canberra, Australia, 2002.
- Australian Renewable Energy Agency. *Biogas Opportunities for Australia*; Technical Report; Australian Renewable Energy Agency: Canberra, Australia, 2019.
- Was, A.; Sulewski, P.; Krupin, V.; Popadynets, N.; Malak-Rawlikowska, A.; Szymańska, M.; Skorokhod, I.; Wysokiński, M. The potential of agricultural biogas production in Ukraine—impact on GHG emissions and energy production. *Energies* **2020**, *13*, 5755.
- Bouramdane, A.A. Natural hazards in electricity grids: From landscape dynamics to optimal mitigation and adaptation approaches. *Emerg. Manag. Sci. Technol.* **2024**, *4*, e007.
- Singh, T.; Uppaluri, R.V. Feed-forward ANN and traditional machine learning-based prediction of biogas generation rate from meteorological and organic waste parameters. *J. Supercomput.* **2024**, *80*, 2538–2571.
- Sappl, J.; Harders, M.; Rauch, W. Machine learning for quantile regression of biogas production rates in anaerobic digesters. *Sci. Total Environ.* **2023**, *872*, 161923.
- Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799.
- Jameel, S.S.; Raazi, S.M.K.u.R.; Jameel, S.M. Multimodal ML Strategies for Wind Turbine Condition Monitoring in Heterogeneous IoT Data Environments. In *Proceedings of the International Conference on Forthcoming Networks and Sustainability in the AIoT Era*; Springer: Cham, Switzerland, 2024; pp. 216–228.

17. Varathan, G.; Belwin Edward, J. A review of uncertainty management approaches for active distribution system planning. *Renew. Sustain. Energy Rev.* **2024**, *205*, 114808.
18. Dong, Z.; Li, X.; Luan, F.; Ding, J.; Zhang, D. Point and interval prediction of the effective length of hot-rolled plates based on IBES-XGBoost. *Measurement* **2023**, *214*, 112857.
19. Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Trans. Neural Netw.* **2010**, *22*, 337–346.
20. Hosen, M.A.; Khosravi, A.; Kabir, H.D.; Johnstone, M.; Creighton, D.; Nahavandi, S.; Shi, P. NN-based prediction interval for nonlinear processes controller. *Int. J. Control Autom. Syst.* **2021**, *19*, 3239–3252.
21. Khosravi, A.; Nahavandi, S. Combined nonparametric prediction intervals for wind power generation. *IEEE Trans. Sustain. Energy* **2013**, *4*, 849–856.
22. Nasirzadeh, F.; Kabir, H.D.; Akbari, M.; Khosravi, A.; Nahavandi, S.; Carmichael, D.G. ANN-based prediction intervals to forecast labour productivity. *Eng. Constr. Archit. Manag.* **2020**, *27*, 2335–2351.
23. Wang, Y.; Wang, S.; Jia, X.; Wang, F. Bayesian Network-Based Failure Risk Assessment and Inference Modeling for Biomethane Supply Chain. *Safety* **2026**, *12*, 9.
24. Quan, H.; Khosravi, A.; Yang, D.; Srinivasan, D. A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 4582–4599.
25. Zong, Z.; Guan, Y. AI-driven intelligent data analytics and predictive analysis in Industry 4.0: Transforming knowledge, innovation, and efficiency. *J. Knowl. Econ.* **2025**, *16*, 864–903.
26. Bukhtoyarov, V.; Tynchenko, V.; Bashmur, K.; Kolenchukov, O.; Kukartsev, V.; Malashin, I. Fuzzy Neural Network Applications in Biomass Gasification and Pyrolysis for Biofuel Production: A Review. *Energies* **2024**, *18*, 16.
27. Sakib, N.; Hosen, M.A.; Khan, B.; Gunn, B.; Johnstone, M. Prediction Interval in Renewable Energy Forecasting: A Comprehensive Review of Uncertainty Quantification Methods. *IEEE Access* **2025**, *13*, 185466–185492.
28. Khosravi, A.; Koury, R.N.N.; Machado, L.; Pabon, J.J.G. Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sustain. Energy Technol. Assessments* **2018**, *25*, 146–160.
29. Pantaleo, A.; De Gennaro, B.; Shah, N. Assessment of optimal size of anaerobic co-digestion plants: An application to cattle farms in the province of Bari (Italy). *Renew. Sustain. Energy Rev.* **2013**, *20*, 57–70.
30. Abdelaziz, A.; Santos, V.; Dias, M.S.; Mahmoud, A.N. A hybrid model of self-organizing map and deep learning with genetic algorithm for managing energy consumption in public buildings. *J. Clean. Prod.* **2024**, *434*, 140040.
31. Vlaović, Ž.D.; Stepanov, B.L.; Anđelković, A.S.; Rajs, V.M.; Čepić, Z.M.; Tomić, M.A. Mapping energy sustainability using the kohonen self-organizing maps-case study. *J. Clean. Prod.* **2023**, *412*, 137351.
32. Holubar, P.; Zani, L.; Hagar, M.; Fröschl, W.; Radak, Z.; Braun, R. Modelling of anaerobic digestion using self-organizing maps and artificial neural networks. *Water Sci. Technol.* **2000**, *41*, 149–156.
33. Sakib, N.; Gope, N.; Hosen, M.A.; Rahim, M.A. Repair Before Predict: A Unified Cross-Domain Preprocessing Framework for Improving Renewable Energy Forecasting. *Arab. J. Sci. Eng.* **2026**, 1–16. <https://doi.org/10.1007/s13369-026-11362-3>.
34. Nleya, Y.; Young, B.; Noorae, E.; Baroutian, S. Anaerobic digestion of dairy cow and goat manure: Comparative assessment of biodegradability and greenhouse gas mitigation. *Fuel* **2025**, *381*, 133458.
35. Safaei Pirooz, A.A.; Sanjari, M.J.; Kim, Y.J.; Moore, S.; Turner, R.; Weaver, W.W.; Srinivasan, D.; Guerrero, J.M.; Shahidehpour, M. Adaptation of high spatio-temporal resolution weather/load forecast in real-world distributed energy-system operation. *Energies* **2023**, *16*, 3477.
36. Soussi, A.; Zero, E.; Bozzi, A.; Sacile, R. Enhancing energy systems and rural communities through a system of systems approach: A comprehensive review. *Energies* **2024**, *17*, 4988.
37. Efkarpidis, N.; Goranović, A.; Yang, C.W.; Geidl, M.; Herbst, I.; Wilker, S.; Sauter, T. A generic framework for the definition of key performance indicators for smart energy systems at different scales. *Energies* **2022**, *15*, 1289.
38. Arabzadeh, V.; Frank, R. A Four-Dimensional Analysis of Explainable AI in Energy Forecasting: A Domain-Specific Systematic Review. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 153.
39. Zhu, Y.; Ma, H.; Zhang, C.; Wu, B.; Fu, H.; Zhou, J.T.; Hu, Q. Rethinking the Reliability of Post-hoc Calibration Methods Under Subpopulation Shift. In *Proceedings of the Pacific Rim International Conference on Artificial Intelligence*; Springer: Singapore, Singapore, 2024; pp. 16–28.
40. Zhou, X.; Du, H.; Sun, Y.; Ren, H.; Cui, P.; Ma, Z. A new framework integrating reinforcement learning, a rule-based expert system, and decision tree analysis to improve building energy flexibility. *J. Build. Eng.* **2023**, *71*, 106536.
41. Koutsandreas, D.; Ardehali, A.; Giannelos, S.; Pudjianto, D. Quantifying the Trade-Offs Between Clean-Energy Expansion and Land Requirements: Evidence from Greece. *Energies* **2026**, *19*, 2261.

42. Giannelos, S.; Pudjianto, D.; Strbac, G. Smart home economic operation under uncertainty: Comparing Monte Carlo and stochastic optimization using Gaussian and KDE-based data. *Oper. Res. Perspect.* **2025**, *15*, 100348.
43. Giannelos, S. Option Valuation of Smart Grid Technology Projects Under Endogenous and Exogenous Uncertainty. Ph.D. Thesis, Imperial College London, London, UK, 2016.

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