2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 5-8, 2025. Vienna, Austria

2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 979-8-3315-3358-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/SMC58881.2025.11342512

Synthetic Data Generation for Wind Energy Forecasting: Comparison Between Statistical and Deep Learning Models

Olga Klyagina[1], Weijie Xia[2], José R. Andrade[1], Pedro P. Vergara[2], Ricardo J. Bessa[1]

***Abstract*— This paper examines the effectiveness of various synthetic data generation methods for deterministic wind power forecasting. Specifically, this work evaluates four approaches—Gaussian Mixture Models (GMMs), t-Copula, DoppelGANger, and FCPFlow—by comparing the forecasting performance, measured using Mean Absolute Error and Root Mean Squared Error, of models trained on synthetic versus real datasets. Our findings indicate that statistical methods (such as GMM and t-Copula) achieve notably better performance under limited data availability. However, the deep generative model FCPFlow yields superior results when sufficient training data is available. These findings suggest that the choice of synthetic data generation method should be informed by the specific data availability context.**

I. Introduction

The transition toward sustainable energy is driving profound changes in modern energy systems. With the integration of renewable energy sources, which rely on rather unpredictable weather conditions, the uncertainty within energy systems has significantly increased over the past decades [1]. Wind power output forecasting remains a persistent challenge in the research field [2]. With the recent advancements in artificial intelligence (AI), various methods have been applied to face this problem, yielding remarkable results. However, a major challenge remains: the limited availability of necessary data. Since power output data can only be collected gradually over time, this constraint hinders the broader adoption of data-driven forecasting methods [3].

Synthetic data generation offers a potential solution by creating more complete and well-balanced datasets [4]. However, despite the widespread adoption of synthetic data generation methods, there remains a lack of quantitative analyses on the performance of these data generation methods and their impact on wind power forecasting. There are various approaches to generating synthetic data, which can be broadly categorized into two major groups based on the generation method. The first group consists of methods that replicate real-world physical properties through mathematical modeling. These approaches are commonly used in climatic and weather modeling [5] and in electrical system simulations based on the digital twin principle [6]. The second group comprises methods that generate time series data by leveraging the statistical properties of existing datasets. These approaches, primarily driven by machine learning and AI techniques, are the focus of this study.

The energy and AI domains have proposed numerous synthetic data generation methods. To augment datasets with rare or underrepresented conditions, Li et al. [7] introduced a controllable VAE-GAN model, which enables the generation of unseen scenarios by incorporating controllable vectors alongside random noise in the generator input. A similar problem is addressed in [8], which focuses on electricity consumption scenarios. The proposed CENTS framework enhances embedding informativeness through context encoding and normalization by introducing a context reconstruction task trained jointly with any generative model. In [9], the authors propose EnergyDiff, a novel framework that leverages the strengths of diffusion models for more effective energy data generation. To improve generation quality and mitigate KL vanishing, a controllable VAE framework was introduced in [10], incorporating a PI controller to dynamically adjust the KL term during training. Additionally, to model long-range dependencies in time series, TTS-GAN was proposed in [11], employing a pure transformer-based architecture for both the generator and discriminator to generate realistic synthetic sequences.

Although numerous synthetic data generation methods have been proposed, to the best of our knowledge, there is a lack of comprehensive quantitative analysis regarding their impact on predictive tasks, particularly in the context of wind power forecasting. Existing studies predominantly emphasize data fidelity or realism, while the utility of synthetic data in downstream tasks remains largely underexplored.

This paper presents a systematic evaluation of how the synthetic data generated by different models - FCPFlow [12], Gaussian Mixture Models (GMMs) [13], t-Copula [14], and DoppelGANger [15] - affect the performance of forecasting at varying levels of data availability. By understanding the relationship between synthetic data quality and forecasting accuracy, our study provides practical insights into how data augmentation can be effectively used to support a downstream task of deterministic wind power forecasting.

II. Generative models

In this section, we introduce the synthetic data generation methods: FCPFlow [12], GMMs [13], t-Copula [14], and

[1]Olga Klyagina, Ricardo Andrade and Ricardo Bessa are with the Center for Power and Energy Systems (CPES) in the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal (e-mail:{olga.klyagina, jose.r.andrade, ricardo.j.bessa}@inesctec.pt)

[2]Weijie Xia and Pedro P. Vergara are with the Intelligent Electrical Power Grids (IEPG) Group, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail:{W.xia, P.P.VergaraBarrios}@tudelft.nl).

This work is supported by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project ATE, with reference 56.

DoppelGANger [15], which we examined in this research. FCPFlow and DoppelGANger are neural network-based methods, while GMMs and t-Copula are statistical methods.

A. Gaussian Mixture Models

GMMs is a probabilistic model frequently used for data generation in the energy domain [13]. Formally, a GMM represents the probability density function as a weighted sum of K Gaussian components

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{1}$$

where π_k are the mixture weights satisfying $\sum_{k=1}^{K} \pi_k = 1$ and $\pi_k \geq 0$, $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the Gaussian distribution with mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$.

B. t-Copula

The t-Copula is another widely used dependency model in energy-related data generation [14]. Formally, the t-Copula defines a multivariate dependence structure using the multivariate Student's t-distribution. Given a d-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ with marginal cumulative distribution functions (CDFs) $F_i(x)$, the t-Copula function $C(u_1, \ldots, u_d)$ is given by

$$C(u_1, \ldots, u_d) = T_{\nu,\boldsymbol{R}}(T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d)), \tag{2}$$

where $T_{\nu,\boldsymbol{R}}$ is the CDF of the multivariate Student's t-distribution with degrees of freedom ν and correlation matrix $\boldsymbol{R}$, and $T_\nu^{-1}(\cdot)$ is the inverse CDF (quantile function) of the univariate Student's t-distribution with ν degrees of freedom.

C. FCPFlow

FCPFlow is a flow-based generative model designed for the generation of energy profiles [12]. The FCPFlow model transforms input data $\mathbf{x}$ into a latent variable $\mathbf{z}$ following a standard Gaussian distribution using a sequence of invertible transformations

$$\mathbf{z} = f^{-1}(\mathbf{x}), \tag{3}$$

where f^{-1} is a bijective function. The probability density of $\mathbf{x}$ can then be expressed using the change of variables theorem

$$p(\mathbf{x}) = p_{\mathbf{z}}(f^{-1}(\mathbf{x})) \left| \det \frac{\partial f^{-1}(\mathbf{x})}{\partial \mathbf{x}} \right|. \tag{4}$$

D. DoppelGANger

DoppelGANger is a generative adversarial networks (GAN) - based model introduced in [15] for time series generation with high fidelity. GANs were introduced in [16] and became a widely adopted approach for the generation of synthetic data, particularly in the domains of image and time series generation. GANs utilize a principle of minmax game between two networks: generator G, which creates new data samples from noise, and discriminator D that distinguishes between the real and generated samples. The models are trained simultaneously, and G maximizes the loss of D.

In the DoppelGANger, the generator is based on Long Short-Term Memory (LSTM) architecture, allowing it for better performance with time series compared to the multilayer perceptron basement that is used in the original GAN, as well as to track autocorrelation in the dataset. To tackle mode collapse - a common problem for GANs when a generator starts producing very similar samples only - the authors normalize each time series signal individually, using the min-max range as metadata.

III. Forecasting Models

In this experiment, we employ two deterministic forecasting models: LightGBM [17], representing classical statistical models, and Feedforward Neural Network (FNN) [18], representing deep learning-based models.

A. LightGBM

LightGBM is a gradient boost-based model that can be used for both regression and classification tasks. The method is scalable and suitable for large datasets. The model relies on Gradient-based One-Side Sampling and Exclusive Feature Bundling techniques that allow it to use the samples with high gradients only and reduce the number of important features, thus speeding up the training process compared to the conventional gradient boosting trees, still being able to provide accurate results [17].

B. Feedforward Neural Network (FNN)

FNN is a deep learning model that maps input features $\mathbf{x} \in \mathbb{R}^d$ to output targets using multiple layers of neurons [18]. The output of an L-layer FNN can be expressed as:

$$\mathbf{h}^{(l)} = \sigma(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad l = 1, \ldots, L, \tag{5}$$

where $\mathbf{h}^{(0)} = \mathbf{x}$ represents the input layer, $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the weight matrix and bias vector of layer l, and $\sigma(\cdot)$ is an activation function.

The final output layer produces the predicted value

$$\hat{y} = \mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}. \tag{6}$$

The network parameters $\Theta = \{\mathbf{W}^{(l)}, \mathbf{b}^{(l)}\}_{l=1}^{L}$ are learned by minimizing the mean squared error (MSE) loss

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2. \tag{7}$$

Optimization is performed using gradient-based methods.

IV. Experimental Setting

A. Dataset

We use the dataset available at [19], which includes three years of wind and solar power generation data at 30-minute resolution, along with hourly numerical weather predictions (NWPs). In the experiments of Sections V-A and V-B, we focus on wind power forecasting using relevant

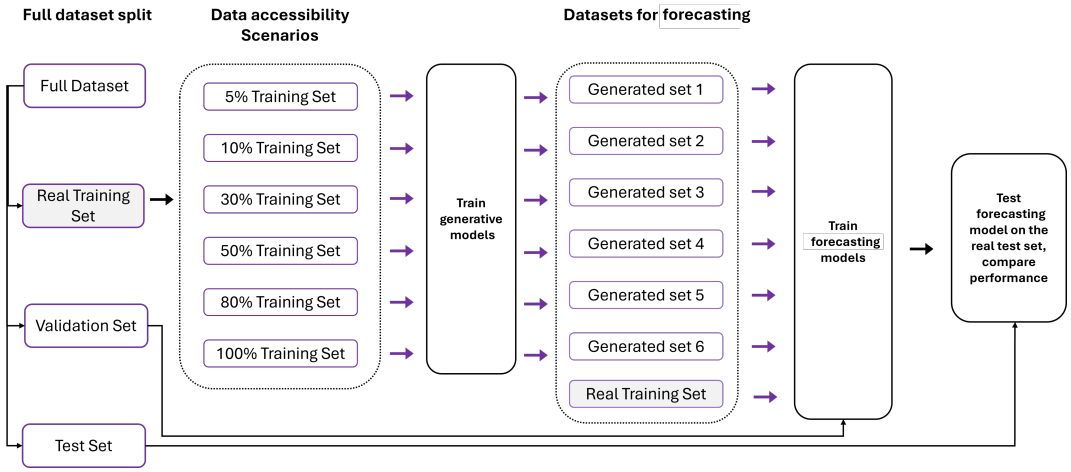

Fig. 1. Overview of the experimental pipeline for evaluating forecasting models under different levels of data accessibility. The full dataset is first split into training, validation, and test sets. The training set is further divided into six subsets with varying proportions of the original data. Synthetic datasets are generated for each subset, and forecasting models are trained and evaluated to analyze the impact of real and generated data on forecasting performance.

meteorological variables: wind speed and direction (at 10m and 100m), temperature, and relative humidity. NWPs are taken from 36 surrounding points using the DWD ICON EU model [20] together with the farm's total generated energy. To evaluate the impact of data generation methods under varying levels of access to real, we first split the dataset into training (70%), validation (10%), and test parts (20%). The full training set contains 978 daily samples, each with seven input features measured twice an hour, resulting in a shape of (978, 48, 7). From this training set, we construct six subsets containing 5%, 10%, 30%, 50%, 80%, and 100% of the training data to simulate varying data availability.

In the experiments of Section V-C, we use the same original dataset but split it to half-hourly samples instead of daily; therefore, a training dataset's size increases to more than 40000 samples with 7 input features[1].

B. Evaluation of Synthetic Datasets

There is still no universal way to evaluate the quality of synthetic data, so the methods of analyzing it depend on the dataset itself and its application [21]. In this work, since we are focusing on the forecasting task, the natural properties of the dataset that are used to estimate the fidelity of the synthetic dataset are:

1) *Pairwise Pearson correlation* between available variables: shows the correlation between different variables in a dataset;
2) *Autocorrelation*: indicates the correlation of a variable with respect to itself in the past;
3) *Power curve*: a dependency of generated power on a wind farm from a wind speed.

Finally, we also estimate *utility* of the synthetic datasets by training the same model on synthetic, testing on real datasets, and comparing the resulting errors.

[1]The code for the model, experiments, and data processing is available at Project Repository and TU Delft Repository.

C. Evaluation Metrics for Forecasting

To assess the performance of the deterministic forecasting models, we employ two commonly used metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) [22]. These metrics quantify the discrepancy between the predicted values $\hat{y}_i$ and the true values y_i over a dataset of size N.

D. Overall Experiment Pipeline

The experimental pipeline, illustrated in Fig. 1, evaluates the impact of the data generation method under different levels of data accessibility on forecasting models. As mentioned in Section IV-A, the full dataset is first split into training, validation, and test sets, with the training set further divided into six subsets, which are used to generate synthetic datasets, producing six generated datasets per generative model that simulate different levels of data availability. The number of samples in synthetic datasets is the same as in the original. Forecasting models are then trained on real and generated datasets, and their performance is assessed using the validation and test sets. This setup enables a systematic analysis of how varying access to real data influences prediction accuracy and whether generated datasets can effectively replace real-world data in predictive modeling.

V. Results and Discussion

A. Fidelity of Synthetic Datasets

In this subsection, we analyze the fidelity metrics introduced in Section IV-B.

- The ***autocorrelation*** plots in Fig. 2 reveal clear differences across model outputs. The reduced autocorrelation comes from independently sampled 24-hour segments (48-time steps), leading to no temporal continuity between samples. All models—except FCPFlow—capture daily temporal patterns. However, even within the first 48 lags, synthetic sequences show a faster decay in autocorrelation than the real data,

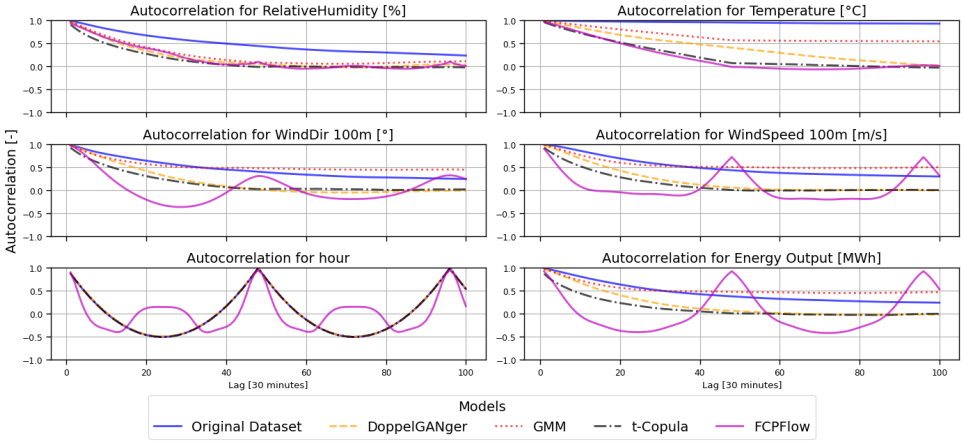

Fig. 2. Auto-correlations of the selected variables in the dataset. Horizontal axis: time lag; vertical axis: correlation value.

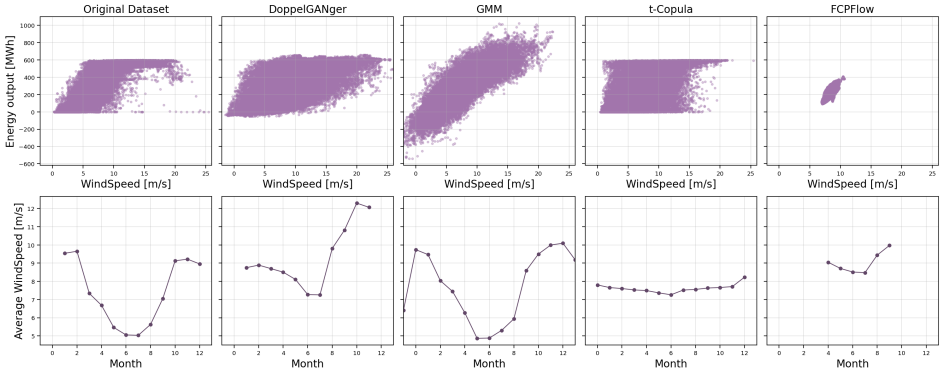

Fig. 3. Power curves (top) and seasonality (bottom) of the synthetic data compared to real data.

Fig. 4. Pairwise Pearson correlations between features in the synthetic and real datasets.

indicating higher noise levels and weaker temporal consistency in the generated datasets.

- ***Power curves*** and ***seasonality*** The plots in Fig.3 show that not all models accurately capture feature relationships in the dataset. DoppelGANger and t-Copula best replicate the power curve but fail to preserve seasonal patterns. In contrast, GMMs capture seasonal boundaries well but occasionally generate unrealistic values, such as negative wind speeds. FCPFlow produces outputs with limited variability, which may suggest undertraining with the current amount of data – a point we further examined in SectionV-C.
- ***Pearson correlations*** on Fig. 4 are maintained for all four considered models. FCPFlow model generated slightly higher correlations between Relative Humidity, Temperature and Wind Speed values compared to the real dataset.

B. Forecasting Performance With Synthetic Datasets

Table I presents a comparative evaluation of forecasting errors—measured by MAE and RMSE—for models trained on synthetic data generated by various methods: DoppelGANger, GMM, t-Copula, and FCPFlow, as well as on real data. Fig. 5 shows a prediction example. From Table I, we observe that the performance of models trained on DoppelGANger-generated data deteriorates as the train-

TABLE I

FORECASTING PERFORMANCE OF FNN AND LIGHTGBM TRAINING ON SYNTHETIC AND REAL DATASETS (DAILY SAMPLES)

Generator	Data	FNN		LightGBM	
		MAE	RMSE	MAE	RMSE
DoppelGANger	5%	92.13	129.13	77.06	108.46
	10%	91.86	117.55	85.13	110.56
	30%	97.70	123.29	103.69	131.72
	50%	108.75	141.36	113.36	138.95
	80%	104.59	128.08	104.82	126.43
	100%	106.91	137.85	109.16	136.56
GMMs	5%	103.00	148.50	85.95	114.86
	10%	84.32	111.28	76.92	107.38
	30%	82.16	116.65	80.68	112.57
	50%	80.55	112.07	80.92	110.30
	80%	80.60	111.57	77.53	107.94
	100%	75.74	106.08	73.63	103.08
t-Copula	5%	92.63	129.61	95.51	120.96
	10%	76.59	111.62	86.93	113.00
	30%	74.84	103.01	79.03	104.79
	50%	77.82	104.43	80.99	106.89
	80%	**72.85**	**102.72**	80.05	105.94
	100%	79.36	108.17	82.00	106.99
FCPFlow	5%	211.71	249.20	189.30	209.95
	10%	212.47	252.17	185.52	205.96
	30%	215.02	254.62	192.45	213.72
	50%	213.48	251.57	187.65	208.44
	80%	225.37	269.21	189.14	210.13
	100%	168.77	193.98	134.85	157.93
Real	5%	82.16	117.48	75.74	110.60
	10%	84.32	117.34	71.54	105.77
	30%	86.80	117.74	58.15	94.09
	50%	91.03	117.16	58.18	94.27
	80%	78.82	109.68	60.25	94.56
	100%	81.54	111.65	**59.79**	**93.94**

TABLE II

FORECASTING PERFORMANCE UNDER NEW EXPERIMENTAL SETTING (HALF-HOURLY SAMPLES)

Data Source	Model	MAE	RMSE
FCPFlow	LightGBM	**58.89**	**93.53**
GMMs	LightGBM	65.68	97.99
t-Copula	LightGBM	65.94	97.79
Real	LightGBM	60.17	93.81
FCPFlow	FNN	**60.34**	**95.41**
GMMs	FNN	65.68	98.41
t-Copula	FNN	65.69	102.62
Real	FNN	62.06	97.12

ing data volume increases, suggesting potential underfitting or suboptimal hyperparameter choice. Across both evaluation metrics and model types, synthetic data produced by FCPFlow consistently yields the highest prediction errors. This finding is consistent with the discussion in Section V-A, which concluded that FCPFlow fails to adequately capture daily temporal patterns. In contrast, both t-Copula and GMM demonstrate strong and stable performance, often achieving error levels comparable to those of models trained on real data. Notably, in some cases, the FNN models trained on synthetic data generated by these two statistical methods even outperform those trained on real data.

Overall, these results indicate that synthetic data generated by statistical models—particularly t-Copula and GMM—tend to be more robust and reliable for downstream prediction tasks than those produced by deep generative models.

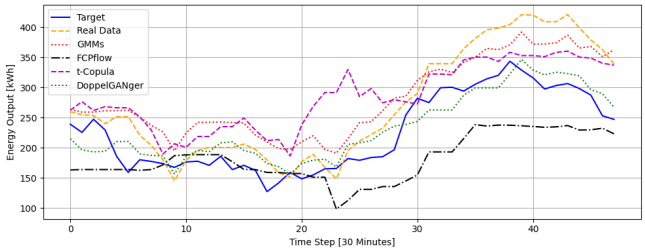

Fig. 5. Forecasting example with synthetic and real datasets for 100% data using LightGBM predictor

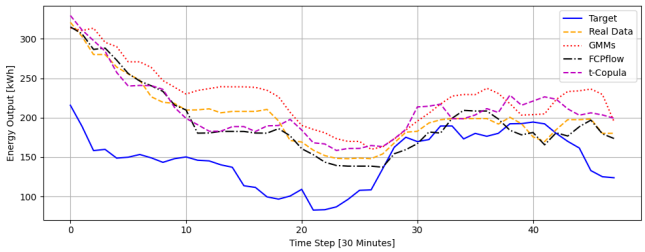

Fig. 6. Forecasting example for models with sufficient training data using LightGBM predictor

C. Can Deep Generative Model outperform Statistical Model?

Table I reveals that deep learning–based generative models generally do not outperform statistical approaches. However, FCPFlow shows a notable decrease in prediction error as data availability increases from 80% to 100%, prompting further investigation into whether neural methods can surpass statistical ones when trained on sufficient data.

To explore this, we retrain GMM, t-Copula, and FCPFlow on the dataset with half-hourly sampling introduced in Section IV-A, using the full training set. Synthetic data generated by each model is then used to train LightGBM and FNN models, with prediction performance summarized in Table II and a prediction example demonstrated in Fig. 6.

Under this new setting, FCPFlow outperforms the statistical models in most cases. For LightGBM, it achieves the lowest RMSE (93.52), even better than the model trained on real data (93.81). In the FNN setting, FCPFlow again leads, with the lowest MAE (60.34) and RMSE (93.81) among all generators. These results suggest that with sufficient training data, neural network–based methods like FCPFlow can better capture complex data distributions and outperform traditional statistical approaches and even real data in downstream forecasting tasks.

VI. DISCUSSION

Our results offer valuable insights into when and why deep learning–based generative methods can underperform or outperform statistical ones in synthetic data generation for forecasting.

A. Why Neural Models Initially Underperform

In the first experimental setting, deep learning-based models—especially DoppelGANger and FCPFlow—do not

outperform statistical methods. This is likely due to undertuning in DoppelGANger and data inefficiency in FCPFlow. FCPFlow requires large datasets to learn complex distributions. With only 978 samples of 48×7 features, FCPFlow is unlikely to be sufficiently trained, leading to poor performance.

B. Improved Performance with Sufficient Training Data

In the second setting, with more samples available, FCPFlow clearly outperforms statistical models. This highlights the high expressiveness of deep learning-based models, which can capture complex dependencies more effectively when trained with sufficient data. Unlike statistical models with rigid assumptions, neural methods adapt better to intricate patterns.

C. Synthetic Data Surpassing Real Data: A Dual Perspective

Interestingly, predictive models trained on FCPFlow-generated data surpass those trained on real data in Section V-C. This can be understood in two ways: (1) synthetic data may "fill in the gaps" of real data, enhancing generalization; (2) from a Bayesian view, flow-based models map a Gaussian prior to a complex posterior, effectively smoothing the learned distribution and improving generalization.

VII. Conclusion

In this study, we conducted a comprehensive evaluation of statistical and deep learning–based generative models for synthetic data generation in energy prediction tasks. Our findings show that while statistical models like GMM and t-Copula perform robustly under limited data, deep generative models such as FCPFlow can outperform them—and even real data—when sufficient training data is available. This highlights the potential of deep generative models to enhance data-driven prediction, especially in data-rich settings. Our work underscores the importance of data availability and model capacity in selecting appropriate generative approaches for reliable and generalizable performance in downstream forecasting tasks. Future research will explore the generalizability of these methods across a broader spectrum of real-world energy datasets.

References

[1] L. E. Jones, *Renewable energy integration: practical management of variability, uncertainty, and flexibility in power grids*. Academic press, 2017.

[2] I. K. Bazionis and P. S. Georgilakis, "Review of deterministic and probabilistic wind power forecasting: Models, methods, and future research," *Electricity*, vol. 2, no. 1, pp. 13–47, 2021.

[3] Y. Zhang, X. Kong, J. Wang, H. Wang, and X. Cheng, "Wind power forecasting system with data enhancement and algorithm improvement," *Renewable and Sustainable Energy Reviews*, vol. 196, p. 114349, 2024.

[4] A. Figueira and B. Vaz, "Survey on synthetic data generation, evaluation methods and gans," *Mathematics*, vol. 10, no. 15, p. 2733, 2022.

[5] P. Veers, K. Dykes, E. Lantz, S. Barth, C. L. Bottasso, O. Carlson, A. Clifton, J. Green, P. Green, H. Holttinen *et al.*, "Grand challenges in the science of wind energy," *Science*, vol. 366, no. 6464, p. eaau2027, 2019.

[6] M. Grieves, *Virtually Intelligent Product Systems: Digital and Physical Twins*. American Institute of Aeronautics and Astronautics, 07 2019, pp. 175–200.

[7] Z. Li, X. Peng, W. Cui, Y. Xu, J. Liu, H. Yuan, C. S. Lai, and L. L. Lai, "A novel scenario generation method of renewable energy using improved VAEGAN with controllable interpretable features," *Applied Energy*, vol. 363, p. 122905, Jun. 2024.

[8] M. Fuest, A. Cuesta, and K. Veeramachaneni, "CENTS: Generating synthetic electricity consumption time series for rare and unseen scenarios," Jan. 2025.

[9] N. Lin, P. Palensky, and P. P. Vergara, "Energydiff: Universal time-series energy data generation using diffusion models," 2025. [Online]. Available: https://arxiv.org/abs/2407.13538

[10] H. Shao, S. Yao, D. Sun, A. Zhang, S. Liu, D. Liu, J. Wang, and T. Abdelzaher, "Controlvae: Controllable variational autoencoder," in *International conference on machine learning*. PMLR, 2020, pp. 8655–8664.

[11] X. Li, V. Metsis, H. Wang, and A. H. H. Ngu, "Tts-gan: A transformer-based time-series generative adversarial network," in *International conference on artificial intelligence in medicine*. Springer, 2022, pp. 133–143.

[12] W. Xia, C. Wang, P. Palensky, and P. P. Vergara, "A flow-based model for conditional and probabilistic electricity consumption profile generation and prediction," *arXiv preprint arXiv:2405.02180*, 2024.

[13] W. Xia, H. Huang, E. M. S. Duque, S. Hou, P. Palensky, and P. P. Vergara, "Comparative assessment of generative models for transformer-and consumer-level load profiles generation," *Sustainable Energy, Grids and Networks*, vol. 38, p. 101338, 2024.

[14] E. M. S. Duque, P. P. Vergara, P. H. Nguyen, A. Van Der Molen, and J. G. Slootweg, "Conditional multivariate elliptical copulas to model residential load profiles from smart meter data," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4280–4294, 2021.

[15] Z. Lin, A. Jain, C. Wang, G. Fanti, and V. Sekar, "Using GANs for Sharing Networked Time Series Data: Challenges, Initial Promise, and Open Questions," in *Proceedings of the ACM Internet Measurement Conference*, Oct. 2020, pp. 464–483, arXiv:1909.13403 [cs]. [Online]. Available: http://arxiv.org/abs/1909.13403

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, p. 139–144, Oct. 2020. [Online]. Available: https://doi.org/10.1145/3422622

[17] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: a highly efficient gradient boosting decision tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 3149–3157.

[18] K. Bhaskar and S. N. Singh, "Awnn-assisted wind power forecasting using feed-forward neural network," *IEEE transactions on sustainable energy*, vol. 3, no. 2, pp. 306–315, 2012.

[19] J. Browell, S. Haglund, H. Kälvegren, E. Simioni, R. Bessa, Y. Wang, and D. van der Meer, "Hybrid energy forecasting and trading competition," 2023. [Online]. Available: https://dx.doi.org/10.21227/5hn0-8091

[20] Deutscher Wetterdienst, "Numerical weather prediction forecast data (icon)," 2024, accessed: April 2, 2025. [Online]. Available: https://www.dwd.de/EN/ourservices/nwp_forecast_data/nwp_forecast_data.html

[21] M. Stenger, R. Leppich, I. Foster, S. Kounev, and A. Bauer, "Evaluation is key: A survey on evaluation measures for synthetic time series," *Journal of Big Data*, vol. 11, 09 2024.

[22] N. Lin, D. Yun, W. Xia, P. Palensky, and P. P. Vergara, "Comparative analysis of zero-shot capability of time-series foundation models in short-term load prediction," *arXiv preprint arXiv:2412.12834*, 2024.