Generative Modeling and Data Augmentation for Power System Production Simulation

Linna Xu, Yongli Zhu

Sun Yat-sen University

Gunagzhou，China

NeurIPS 2024 Workshop

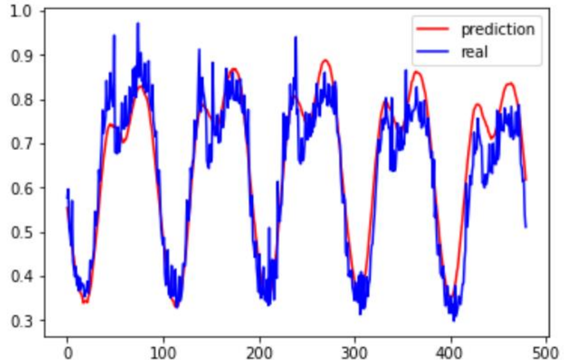
Data-driven and Differentiable Simulations, Surrogates, and Solvers

Training data issues of load forecasting

Load forecasting is critical for the stable operation of power systems and machine learning methods prevail in this field.

Communication failures, device malfunctions, and newly built communities with limited data can impede accurate load forecasting.

Machine learning models usually assume *abundant, high-quality load demand data*, which is often unavailable in the power industry.

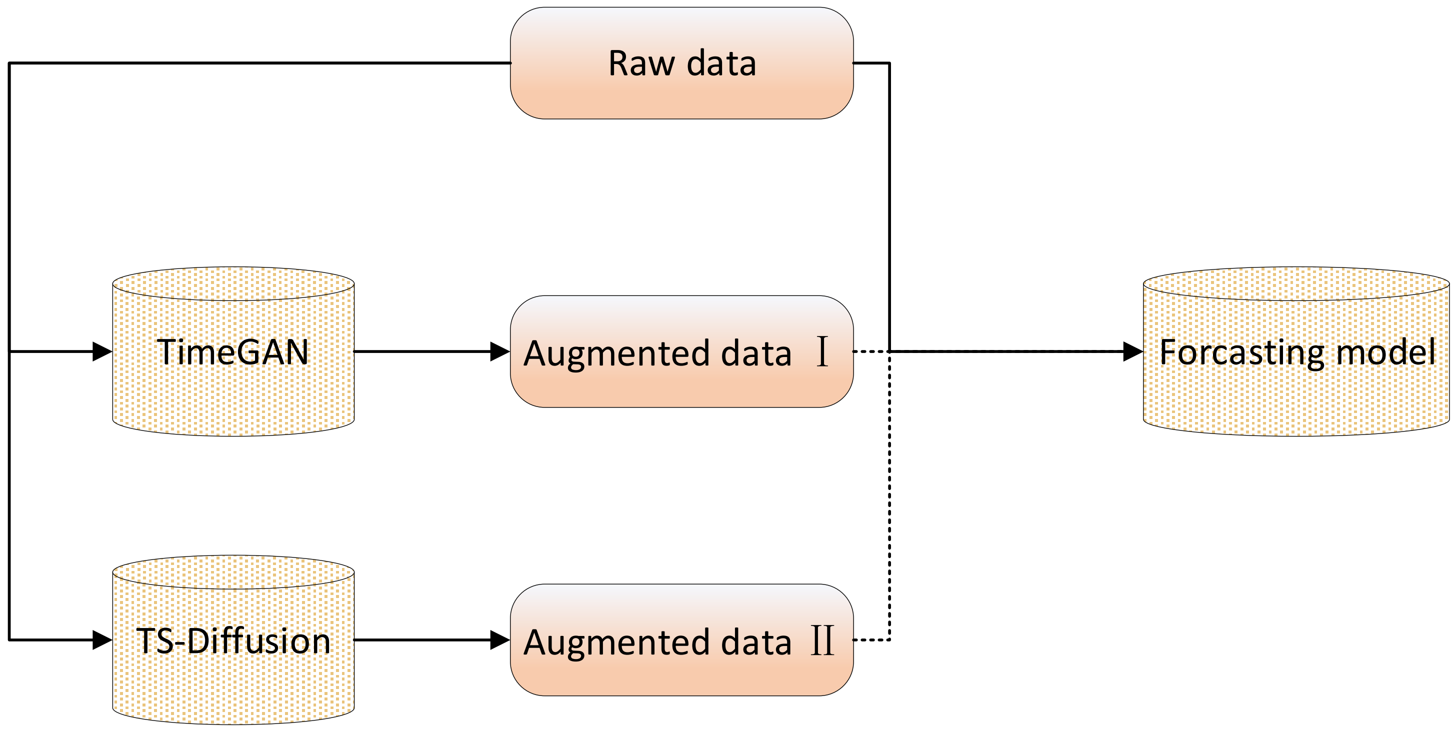


Load forecasting model

Dataset Augmentation for Load Forecasting

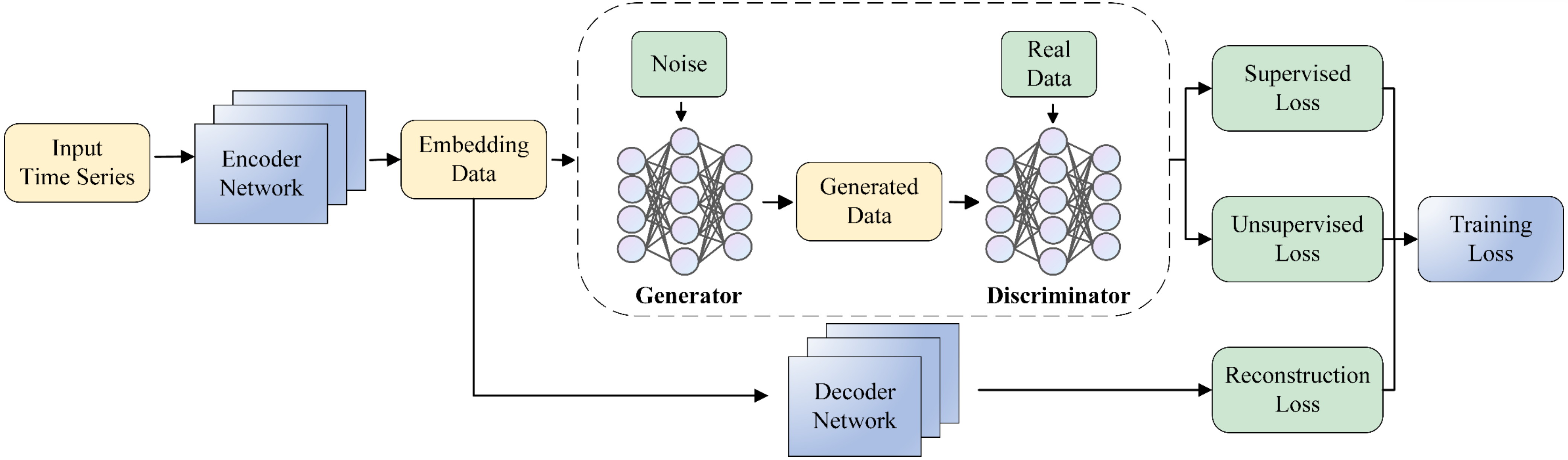
We explore the effectiveness of load data augmentation for power system production simulation using TimeGAN and TS-Diffusion, respectively.

- A data augmentation framework



Dataset Augmentation for Load Forecasting

➢ TimeGAN

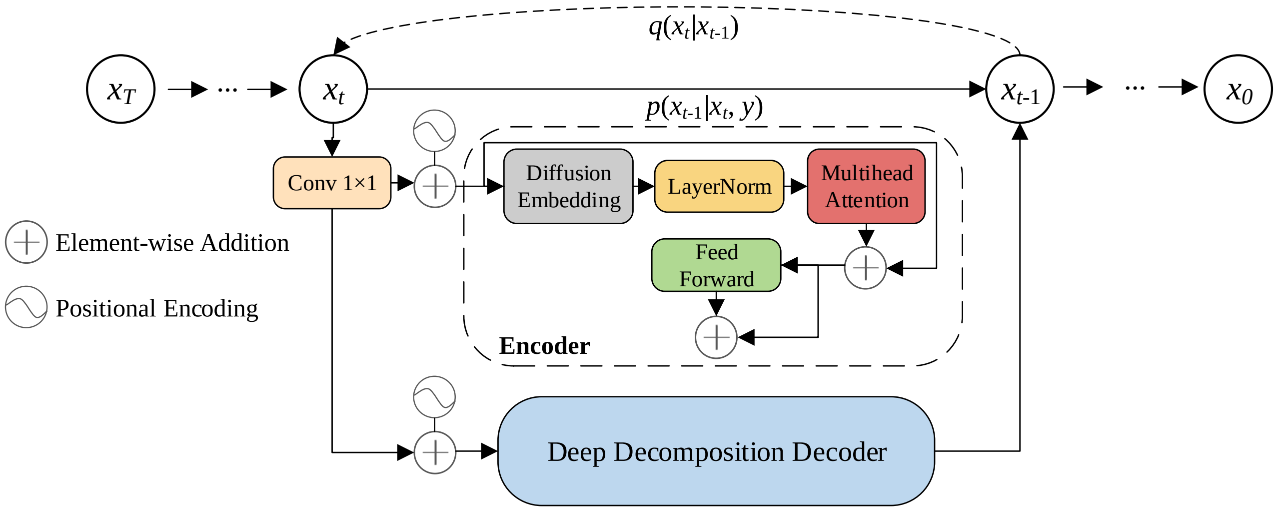


- ***Self-supervised learning:*** It integrates GAN with self-supervised learning to capture complex temporal patterns.
- ***Auto-encoder:*** It maps time series data to latent representations and a decoder that reconstructs the original data.

Dataset Augmentation for Load Forecasting

➢ TS-Diffusion

- A diffusion model for time series data generation
- **Encoder:** processes the input time series using a multi-head attention mechanism and a feed-forward neural network
- **Decoder:** uses multi-head attention and feed-forward layers, plus a deep decomposition design to capture the trend and seasonality components of the time series



Data-driven and Differentiable Simulations, Surrogates, and Solvers Workshop @ NeurIPS '24

Experiment Results

➢ Comparison of indicators on different models

Original vs. TimeGAN augmented vs. TS-Diffusion augmented vs. replicated

Model	dataset	RMSE	MAE
XGBoost	original	0.05774	0.04276
	replicated	0.06485	0.04427
	augmented	0.01526	0.00249
CatBoost	original	0.04389	0.03323
	replicated	0.04536	0.03243
	augmented	0.00236	0.00098
RandomForest	original	0.04183	0.02952
	replicated	0.05846	0.03968
	augmented	0.00153	0.00013
ExtraTree	original	0.04467	0.03209
	replicated	0.04495	0.03229
	augmented	**0.00023**	**0.00004**

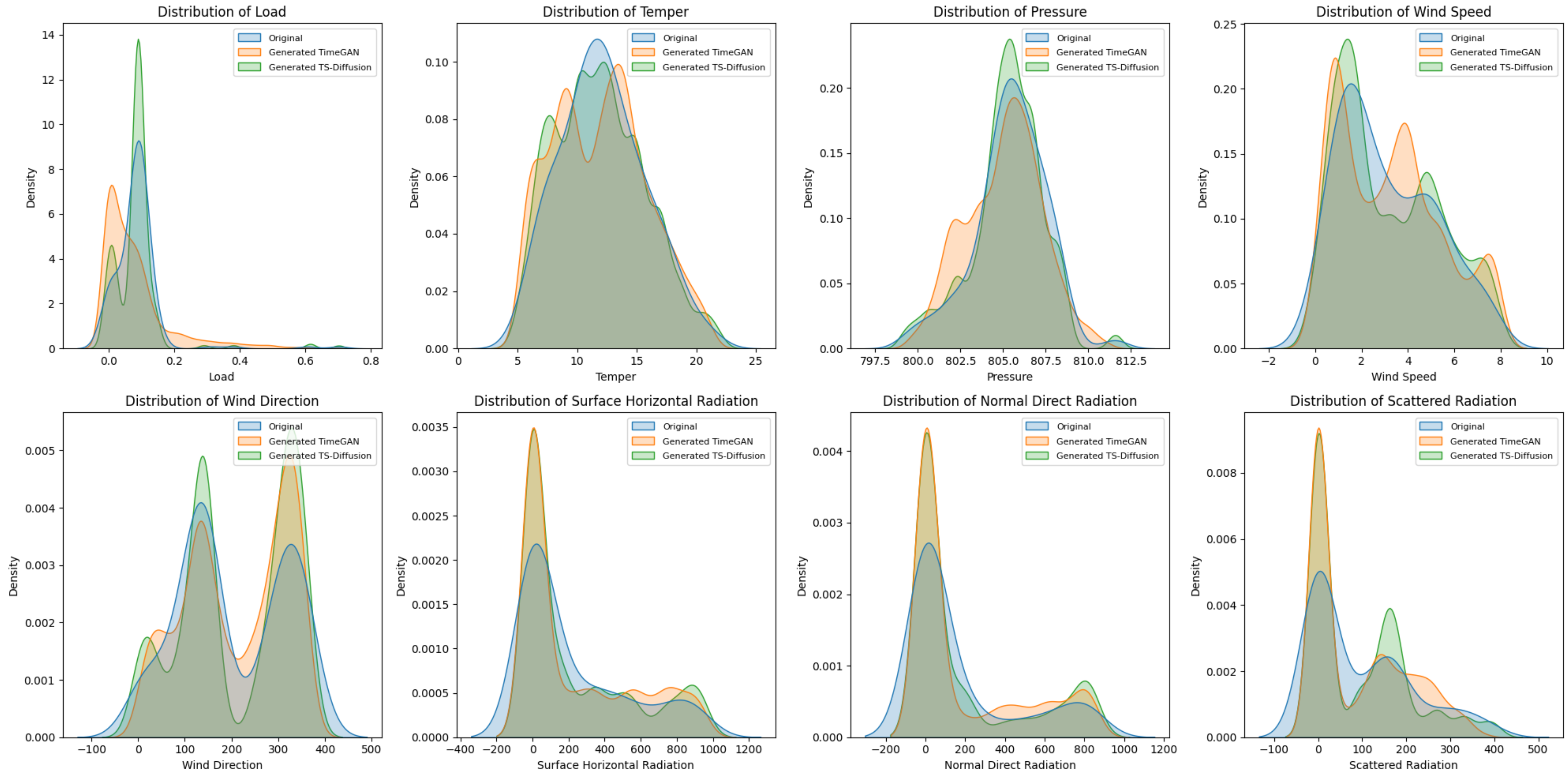
(TS-Diffusion results)

Model	dataset	RMSE	MAE
XGBoost	augmented	0.06398	0.03445
CatBoost	augmented	0.05637	0.02761
RandomForest	augmented	0.06130	0.02949
ExtraTree	augmented	**0.05356**	**0.02395**

(TimeGAN results)

Experiment Results

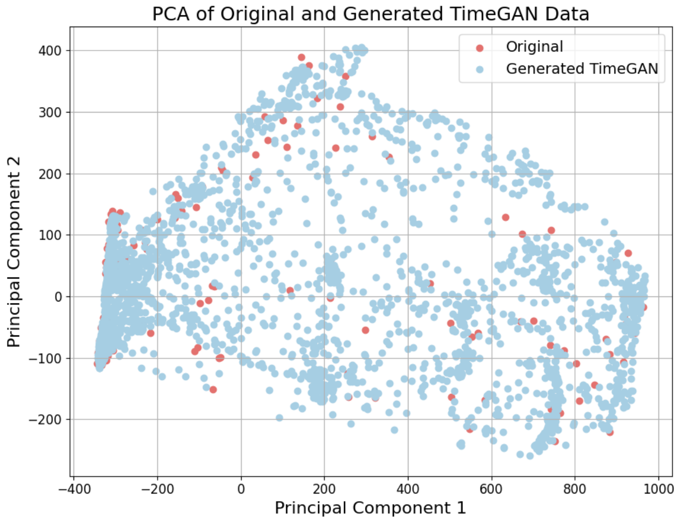
➢ Statistic characteristics of generated data and raw data

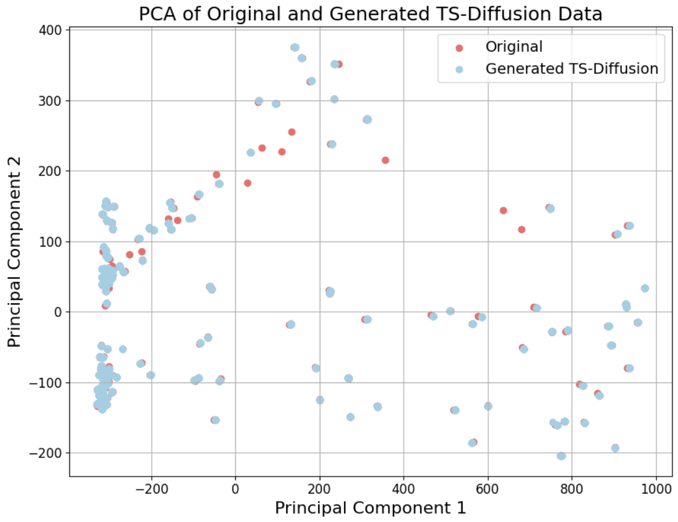


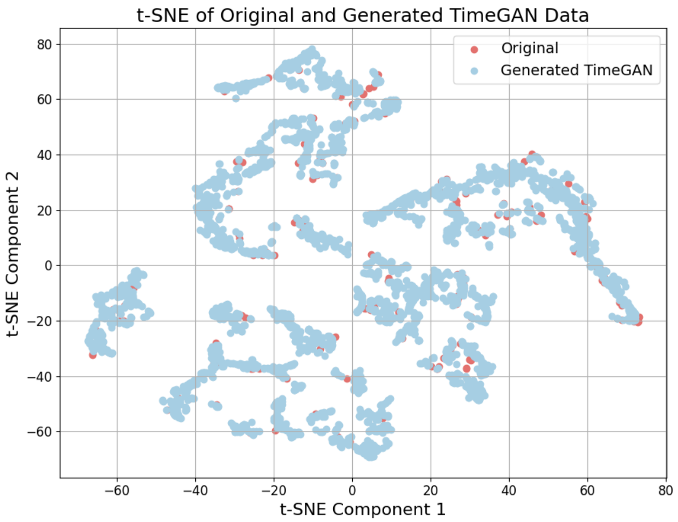
Experiment Results

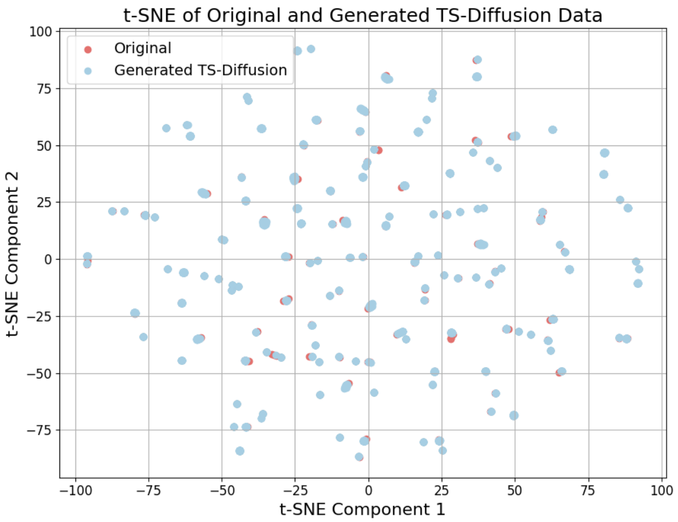
➢ Statistic characteristics of generated data and raw data

- PCA plot





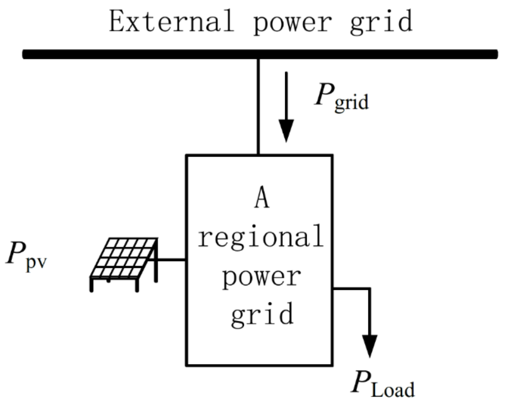




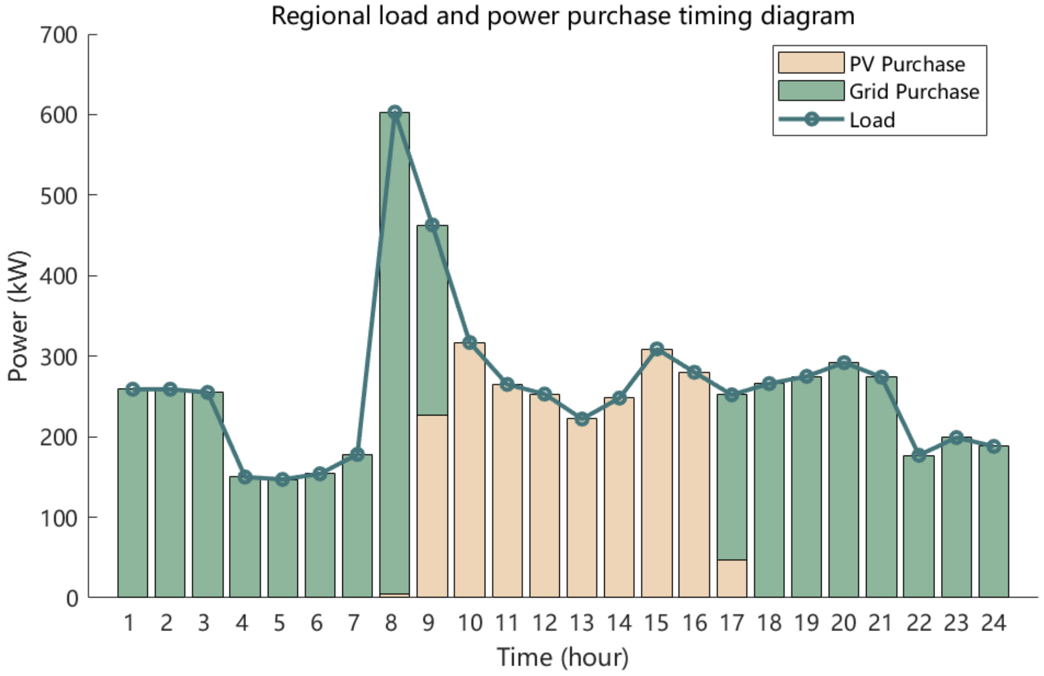
- t-SNE plot

A Simple Showcase for Power System Production Simulation

The previously predicted load data can be utilized in a standard power-system-production-simulation procedure



$$
\begin{aligned}
\text{minimize} \quad & cost_{grid} \sum_{t=1}^{T=24} P_{grid}(t) + cost_{pv} \sum_{t=1}^{T=24} P_{pv}(t) & \\
\text{subject to} \quad & P_{load}(t) = P_{grid}(t) + P_{pv}(t), & t = 1, ..., T \\
& P_{grid}(t) \geq 0, & t = 1, ..., T \\
& 0 \leq P_{pv}(t) \leq P_{pv,max}(t), & t = 1, ..., T
\end{aligned}
$$



Conclusion & Future Work

- Propose a framework to improve the accuracy of load forecasting models using generative machine learning under small samples
- The quality of the generated load data (especially by the diffusion model) significantly improves the load-forecasting accuracy
- Future work includes fine-tuning the generative models for better data quality and conducting additional comparisons
- Applying transfer learning to enhance the model's generalizability will be the next step

Thanks!

xuln6@sysu.edu.cn,
yzhu16@vols.utk.edu