# A novel approach for forecasting non-stationary time series: Utilization of a variational autoencoder reflecting seasonal patterns

Young Eun Jeon[a] • Seung Ho Ryu[a] • Jung-In Seo[a]

[a] Department of Data Science, Gyeongkuk National University, Andong, Korea

## 1. Abstract


### Backgrounds

- Synthetic time series generation has become increasingly important in addressing **challenges posed by limited data**, particularly in domains where data collection is costly, privacy-sensitive, or event-sparse.
- Among various generative techniques, a variational autoencoder (VAE) has gained considerable attention, but **a traditional VAE approach often falls short when modeling complex temporal dependencies such as seasonality and periodicity.**

### Purpose

- This study develops a **generating and forecasting framework based on a VAE that incorporates a seasonal prior designed using Fourier terms.**
- Unlike a traditional variational autoencoder, **the proposed framework is designed to more effectively capture and replicate seasonal dynamics inherent in real-world time series** by embedding seasonal information into the latent space.
- **This integration not only facilitates the faithful reconstruction of seasonal patterns of the original time series but also enables the generation of structurally consistent synthetic data.**


## 2. Methodology

### Architecture

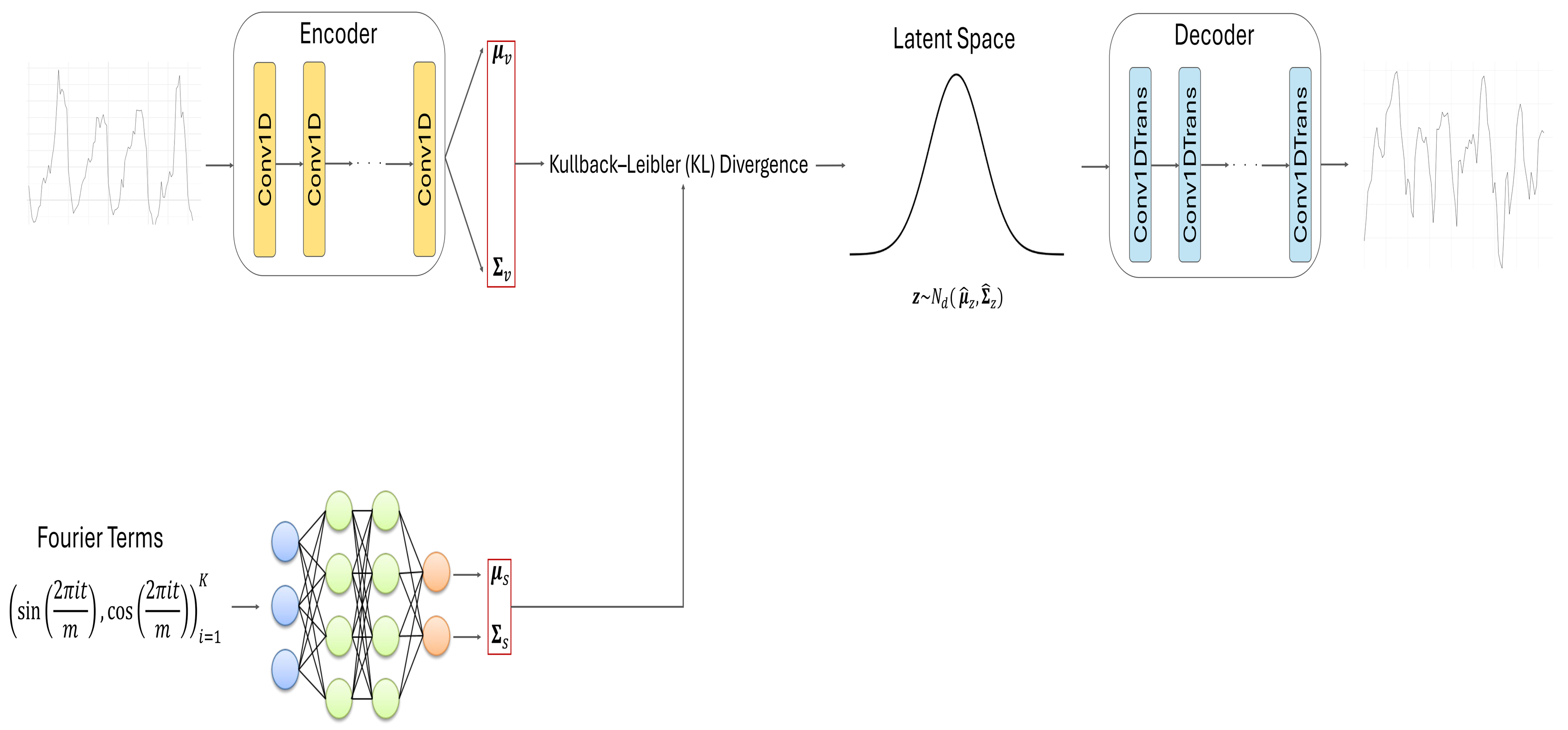



**Fig 1.** Architecture of capturing seasonality in the proposed VAE model

### Loss Function

Let $\mathbf{x}^{(i)} = \left\{x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\right\}$ denote the $i$th input time series sequence of length $n$. Then,

$$\mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}^{(i)}\right) = \mathbb{E}_{q_\phi\left(\mathbf{z}|\mathbf{x}^{(i)}\right)}\left[\log p_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)} \mid \mathbf{z}\right)\right] - D_{KL}\left(q_\phi\left(\mathbf{z} \mid \mathbf{x}^{(i)}\right) \parallel p_{\boldsymbol{\theta}}\left(\mathbf{z}\right)\right)$$

- $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})$: Approximate posterior distribution over the latent variable **z** conditioned on the input sequence $\mathbf{x}^{(i)}$

$$\mathbf{z}|\mathbf{x}^{(i)} \sim N_d(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v)$$

- $p_{\boldsymbol{\theta}}(\mathbf{z})$: Prior distribution over the latent variable z

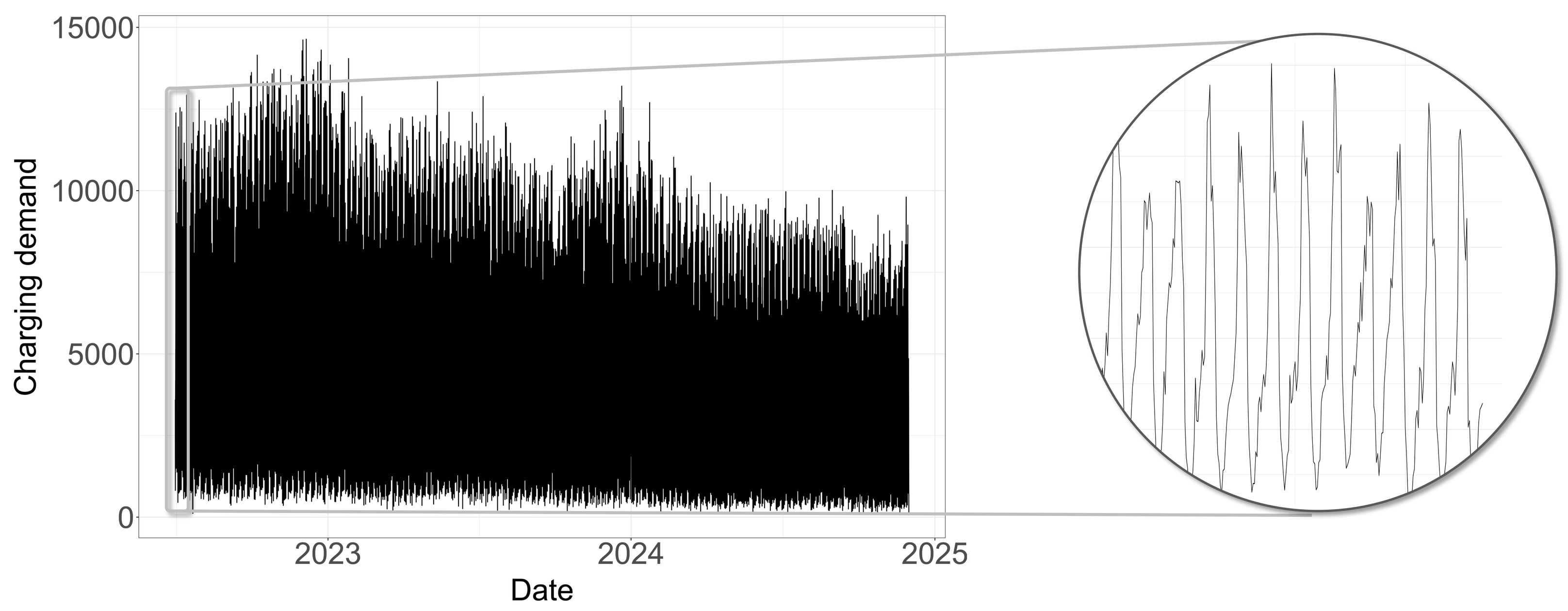
$$\mathbf{z} \sim N_d(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$$

- $p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})$: Conditional likelihood of the input sequence $\mathbf{x}^{(i)}$ given the latent variable $\boldsymbol{z}$

## 3. Application

### Dataset

- Hourly time series of nationwide electric vehicle (EV) charging demand in Korea from July 1, 2022, to November 30, 2024
  (Source: Korea's official public data portal)




### Flowchart

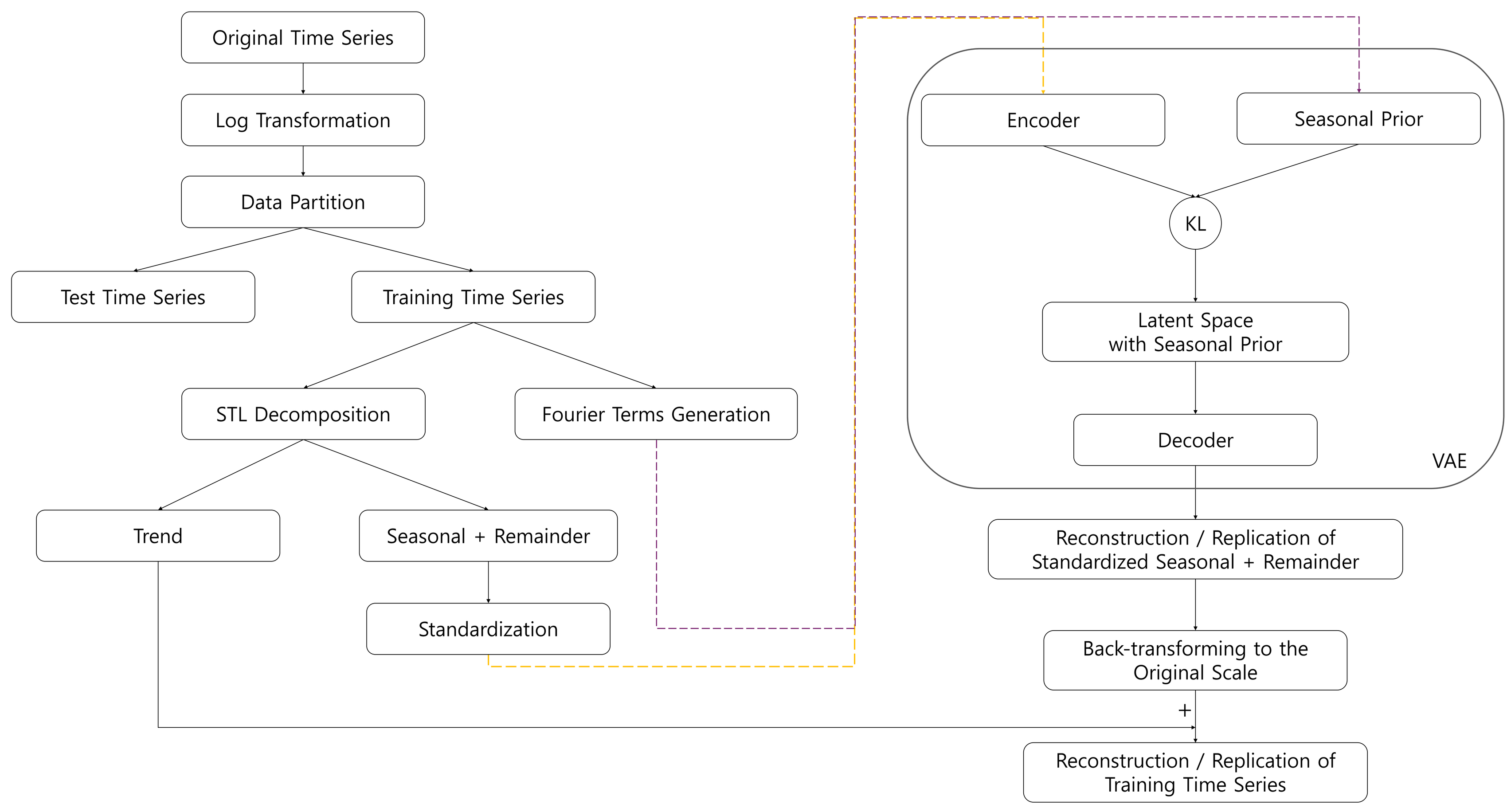



**Fig 2.** Flowchart for reconstruction/replication in the EV dataset

### Results

#### Reconstruction Accuracy for for Test Time Series

| Model | MAE | RMSE | SMAPE | Seasonal MAE | Seasonal RMSE |
|---|---|---|---|---|---|
| $\mathcal{M}_b$ | 0.131 | 0.169 | 1.717 | 0.069 | 0.098 |
| $\mathcal{M}_p$ | 0.112 | 0.141 | 1.449 | 0.055 | 0.075 |

$\mathcal{M}_b$: Vanilla VAE
$\mathcal{M}_p$: Proposed VAE

**Table 1.** Evaluation results for reconstruction accuracy

#### Replication for Training Time Series

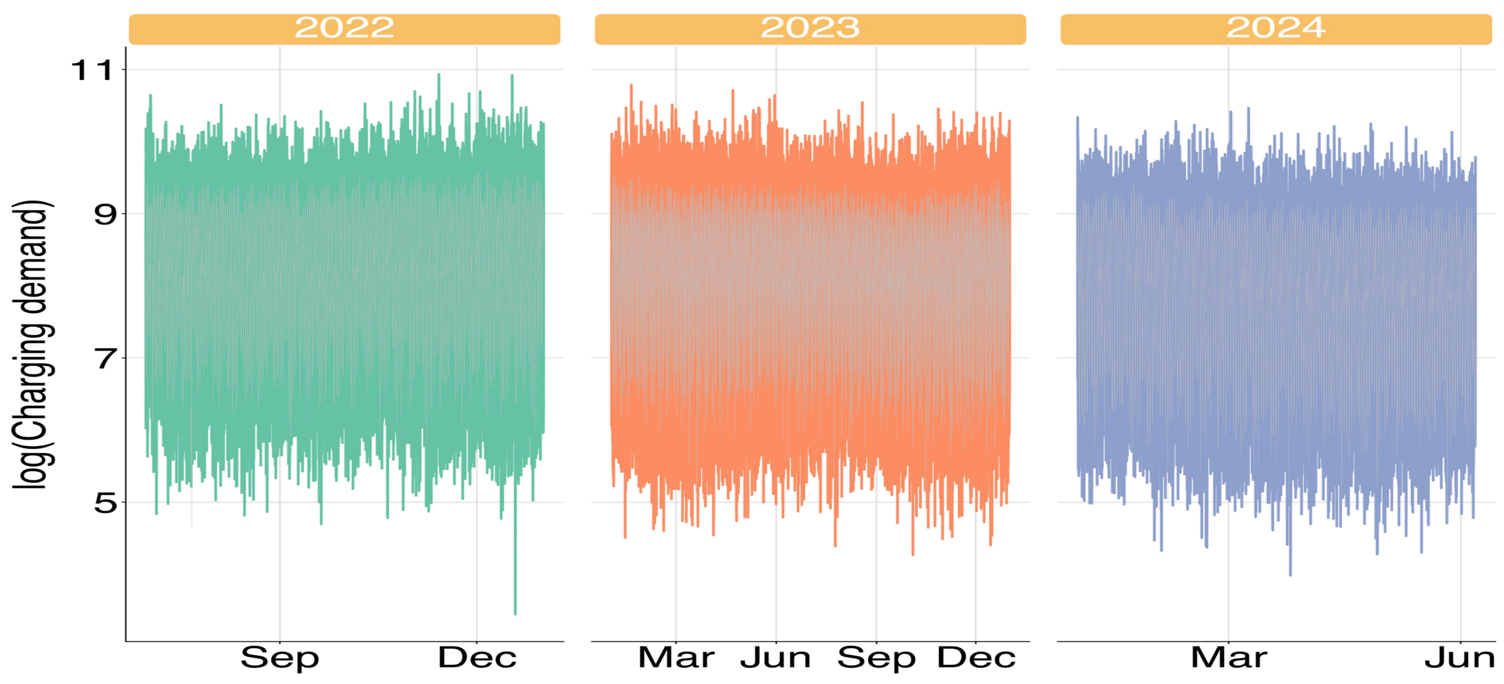



**Fig 3.** 100 replications for training time series (Grey line: Observation)

#### Forecasts for Test Time Series

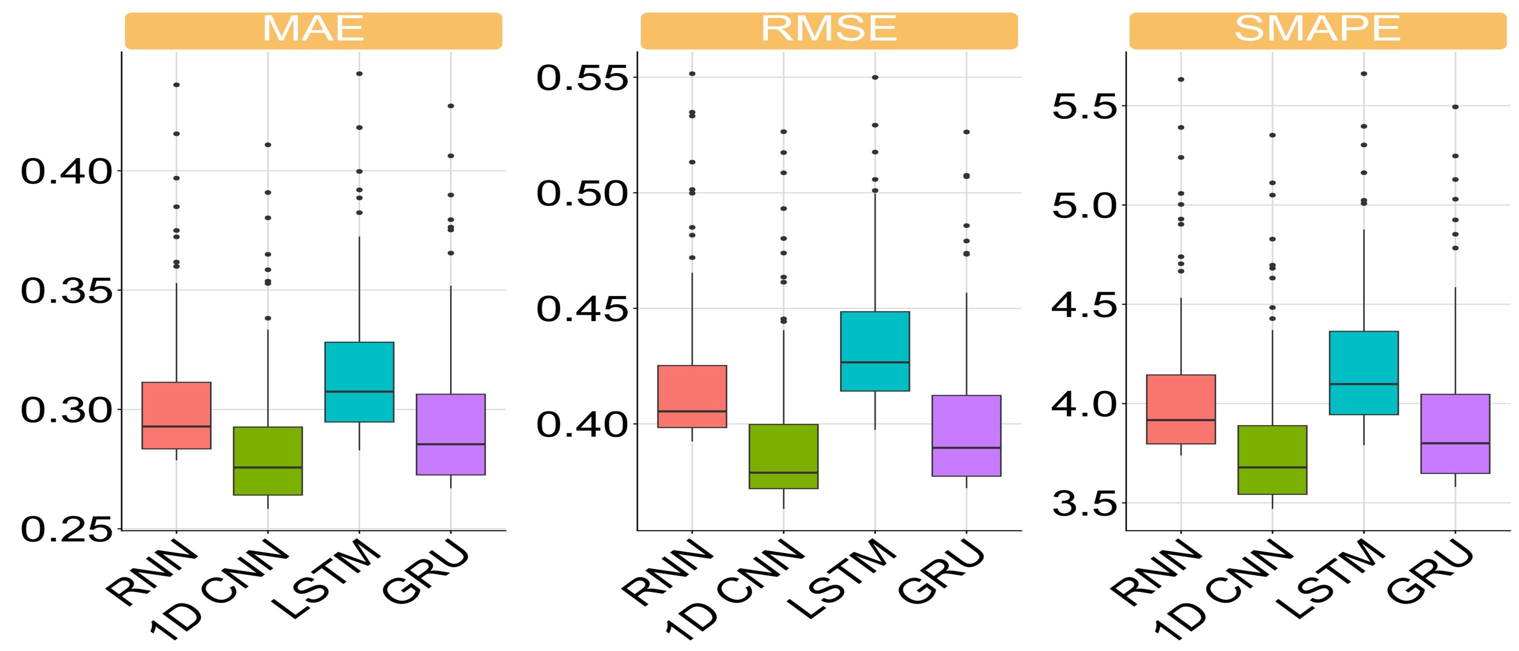



**Fig 4.** Boxplots for forecasting evaluation of test time series using 100 replications

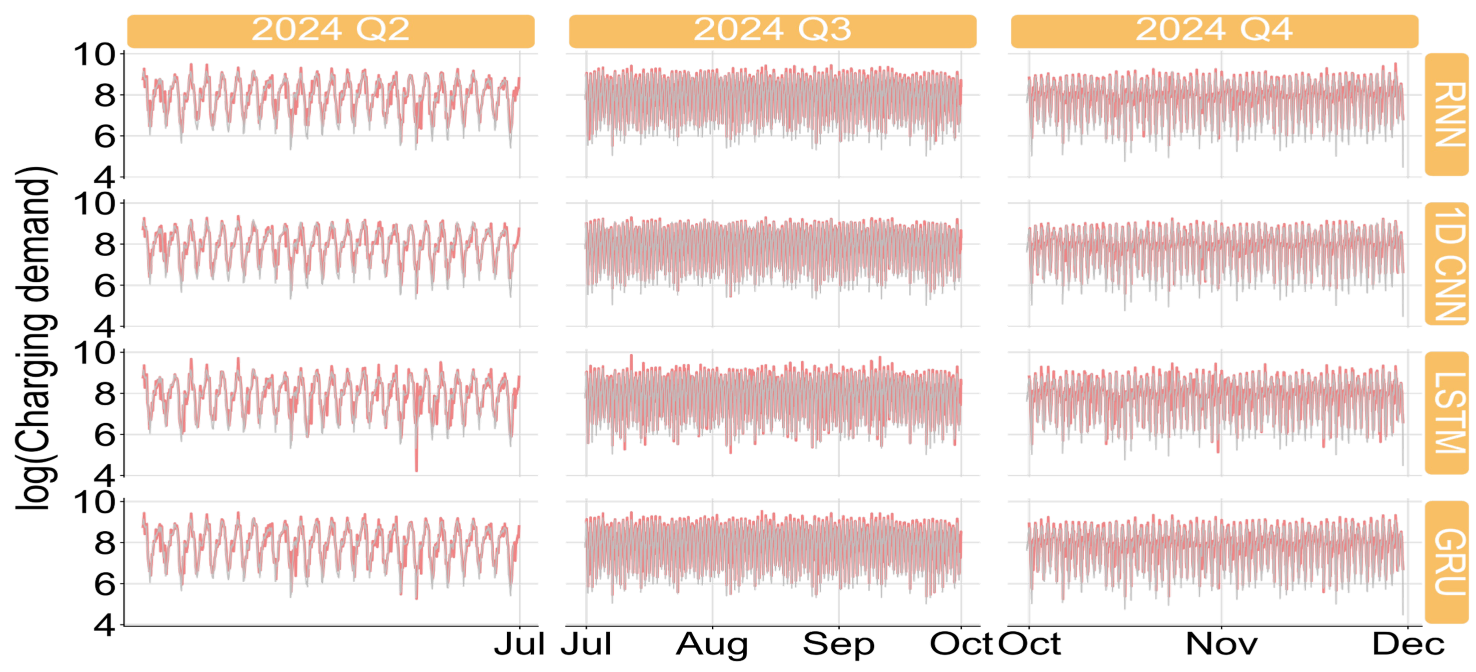



**Fig 5.** Forecasting plots for test time series (Grey line: Observation, Red line: Median)

## 4. Discussion and Conclusion

- The proposed VAE model not only reconstructs time series data more accurately but also preserves seasonal dynamics better, compared to the traditional VAE model.
- 1D CNN achieves the lowest median values across all three metrics, which reveals that it offers more reliable forecasts compared to the other models.
- The proposed framework **enables more effective preservation of seasonal patterns in time series and supports the generation of high-quality synthetic data**, which can be reliably used to train forecasting models.