Empirically Estimated Uncertainty-Guided Iterative Decoding Strategy of Diffusion Models for Time Series Forecasting

Koki Ueno, **Kohei Makino**, **Yutaka Sasaki**
Computational Intelligence Lab, Toyota Technological Institute
uenokouki.soccer@gmail.com, {sd21505, yutaka.sasaki}@toyota-ti.ac.jp,

Abstract

Diffusion models have made significant progress in diverse data generation; however, their performance in time series forecasting, which is a task to forecast future series from past series, is insufficient compared to other existing methods. Therefore, this paper aims to improve the performance of diffusion model-based time series forecasting. In our pilot experiment, we observed the problem of the outputs of the diffusion models degrading away from the gold series. To solve this problem, it is necessary to explore forecast series from various series that do not deviate from the gold series. Therefore, inspired by the uncertainty sampling used in active learning, this paper proposes EUGID (Empirically estimated Uncertainty-Guided Iterative Decoding strategy), a novel decoding strategy for diffusion models that samples forecast series from estimated uncertainty distribution. Specifically, the sampling is achieved by iteratively reproducing the uncertainty distribution by a forward diffusion process and generating a forecast series by a reverse diffusion process. EUGID improves forecasting performance on seven out of eight time series forecasting benchmarks compared to existing diffusion model-based methods.

1 Introduction

Time Series Forecasting (TSF) is the task of predicting future sequences based on past time series data. TSF has widely studied because of its usefulness in various applications, such as in economics [Henrique *et al.*, 2019], meteorology [Li *et al.*, 2022], and energy systems [Lai *et al.*, 2018]. Various forecasting models have been proposed to improve forecasting performance. Recently, deep learning-based models have become mainstream [Liu *et al.*, 2024] while classical engineering-based methods [Box *et al.*, 2015], including statistical approaches such as moving averages and autoregressive models, were conventionally widely used.

Diffusion models have been used for time series interpolation [Tashiro *et al.*, 2021; Alcaraz and Strodthoff, 2023] and generation [Narasimhan *et al.*, 2024]. Since the diffusion models can represent data distribution and generate diverse and high-quality data, they demonstrate remarkable potential in generating data, enabling significant advancements in various domains, such as image [Rombach *et al.*, 2021], text [Gong *et al.*, 2023], and audio generation [Chen *et al.*, 2021]. These methods generally involve two primary processes: (1) the forward diffusion process, in which noise is gradually added to the input data, such as an image, until it becomes Gaussian noise; (2) the reverse denoising process, in which a denoising network incrementally removes the noise to reconstruct the original input. Given that TSF can be viewed as a conditional generation task, diffusion models naturally lend themselves to the application of predicting future time series. Such diffusion models perform well in modeling time series such as time series generation [Narasimhan *et al.*, 2024] and missing value imputation [Tashiro *et al.*, 2021; Alcaraz and Strodthoff, 2023].

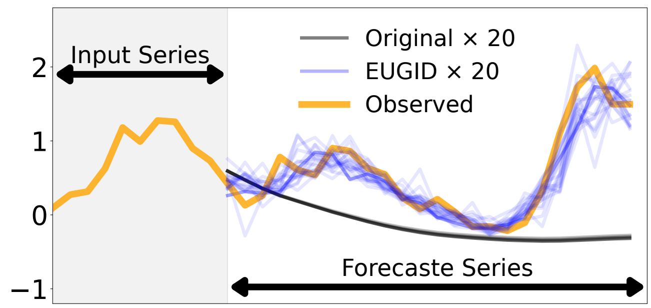

Figure 1: 20 forecast series using different random seeds. The forecast series of the existing method CSDI degenerates to a single series, whereas CSDI with the proposed method EUGID produces diverse forecast series around the correct observed data.

Existing time series diffusion models have not achieved SotA performance for TSF [Rasul *et al.*, 2021; Tashiro *et al.*, 2021; Alcaraz and Strodthoff, 2023; Kollovieh *et al.*, 2023; Liu *et al.*, 2024; Yang *et al.*, 2024]. The performance in other tasks implies a higher level of prediction performance for TSF as well. Although the diffusion models are capable of generating a wide variety of data, the variation in the data generated deviates from gold observed data. As shown in Fig. 1, which compares the diversity of the generation of the forecast series, we observed that the predicted series of an existing method (i.e., black lines) deviated from the observed gold series (i.e., orange line) and the same predictions are output for 20-time trials. In this paper, we describe Empirically estimated Uncertainty-Guided Iterative Decoding strategy (EUGID), which generate diverse predictions (i.e., blue lines) while the predictions are close to the gold observation data. We found that the ways to improve performance for TSF might be to provide a clue to exploring the potential use of the diffusion model for forecasting tasks.

We considered how to improve the performance of diffusion model-based TSF by avoiding deviated generation. Although the diffusion models are known to model the data distribution well, we believe the deviated generation is the decoding in the reverse diffusion process. [Sadat *et al.*, 2024] enabled diffusion models to enhance generation diversity by adding a perturbation to the reverse diffusion process. This implies that the output of the existing diffusion model selects and generates data with extremely high probability in the distribution in the data distribution diffusion models have.

Therefore, we aim to replace the decoding strategy not to generate deviated forecast series and improve the performance of diffusion models for TSF. However, simply making generation diverse does not result in high performance in TSF and accurate prediction of future series is required. In TSF, the incompleteness of given information and the limitation of the model results in uncertainty in the forecast series [Smith, 2013; Simon *et al.*, 2018; Möller and Reuter, 2007; Sullivan, 2015]. Therefore, this study takes an approach to focus on improving the more uncertain parts, such as uncertainty sampling in active learning [Raj and Bach, 2022].

We propose a novel decoding strategy, EUGID, that iteratively samples from the uncertain part in decoding the diffusion model using empirically estimated uncertainties in order to reduce the deviation in the uncertain part of the model's predictions. Our proposed strategy allows exploring the uncertainty part, resulting in a diffusion model that can generate more diverse outputs than existing methods without deviating, as shown in Fig. 1. The Conditional Score-based Diffusion models for Imputation (CSDI) we used as a base method produces degenerated outputs that are far from the correct answer, resulting in insufficient prediction performance. On the other hand, our proposal generates diverse outputs around the gold observed series.

A technical challenge in resampling the forecast series based on uncertainty is that direct sampling from the estimated uncertainty distribution can lead to deviant forecasts depending heavily on the accuracy of the estimated uncertainty distribution. To address this, we treat uncertainty as noise introduced by the forward diffusion process in the diffusion model. By applying the forward diffusion process, we diffuse the uncertain parts of the series so that the reverse diffusion process can remake the new prediction. This approach mitigates the impact of uncertainty.

Our contributions are following points:

- This paper proposes a method to guide the decoding strategy of the diffusion model with uncertainty for TSF, EUGID, which realizes sampling from an uncertainty distribution without deviating from the forecast series.
- Our experiments show that applying EUGID improves the forecasting performance of CSDI, a diffusion model-based TSF, significantly reducing the forecast error on seven of the eight TSF benchmark datasets.

2 Background

2.1 Time Series Forecasting

TSF, which is the task of predicting future data series based on past time series patterns. Let an input series of H time steps and the number of variates D be $x \in \mathbb{R}^{H \times D}$, TSF model $f_\theta(x)$ predicts a future series of following L time steps $\hat{y} \in \mathbb{R}^{L \times D}$. Since the objective of TSF is to predict gold series with low error, typical models employ Mean Squared Error (MSE) loss for the objective represented as Eq. (1).

$$\mathcal{L} = \mathbb{E}\left[|| y_{\text{true}} - \hat{y} || \right] \tag{1}$$

Since given input for TSF are incomplete because of things like inaccuracies in measuring and missing data and limitations of TSF modeling, the forecast series have uncertainty [Smith, 2013; Simon *et al.*, 2018; Sullivan, 2015; Erlygin *et al.*, 2023; Möller and Reuter, 2007]. Uncertainty can be expressed as the variance of the forecast series and, thus, as the distribution of the forecast series.

2.2 Diffusion Models

Diffusion models [Ho *et al.*, 2020] reverse the forward diffusion process, which is the process of information diffusion to random noise. Let the original data be z_0, Eq. (2) shows τ-th step diffused data z_τ in the T steps forward diffusion process.

$$z_\tau = \sqrt{1 - \beta_\tau} z_{\tau-1} + \sqrt{\beta_\tau} \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{2}$$

For efficient sampling of τ-th diffused data, Eq. (2) is convertible to Eq. (3), where $\alpha_\tau = \prod_{i=1}^{\tau} 1 - \beta_\tau$.

$$z_\tau = \sqrt{\alpha_\tau} z_0 + \sqrt{1 - \alpha_\tau} \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{3}$$

α_τ denotes the ratio between the original data and the noise added during the forward diffusion process. T and β_τ are set to converge z_T to Gaussian noise.

The diffusion model sequentially reverses this forward diffusion process as in Eq. (5) with modeling noise added by the forward diffusion process using noise decoder $\epsilon_\theta(z_\tau, \tau)$ from Gaussian noise, which means from step T [Ho *et al.*, 2020; Song *et al.*, 2021].

$$\hat{z}_T \sim \mathcal{N}(0, I) \tag{4}$$

$$\hat{z}_{\tau-1} = \frac{1}{\sqrt{\alpha_\tau}} \left(z_\tau - \frac{\beta_\tau}{\sqrt{1 - \alpha_\tau}} \epsilon_\theta(z_\tau, \tau) \right) + \sqrt{\beta_\tau} \epsilon \tag{5}$$

The objective to train the noise decoder is described in Eq. (6).

$$\mathcal{L} = \mathop{\mathbb{E}}_{z_0 \sim \text{Pop}, \epsilon \sim \mathcal{N}(0, I), \tau \sim \text{Uniform}(1, T)} \| \epsilon - \epsilon_\theta(z_\tau, \tau) \|^2 \tag{6}$$

The decoding strategy of diffusion models refers to the procedures used to obtain the final generated data. General diffusion models simply apply the reverse diffusion processes sequentially. This paper modifies the procedure for generating data.

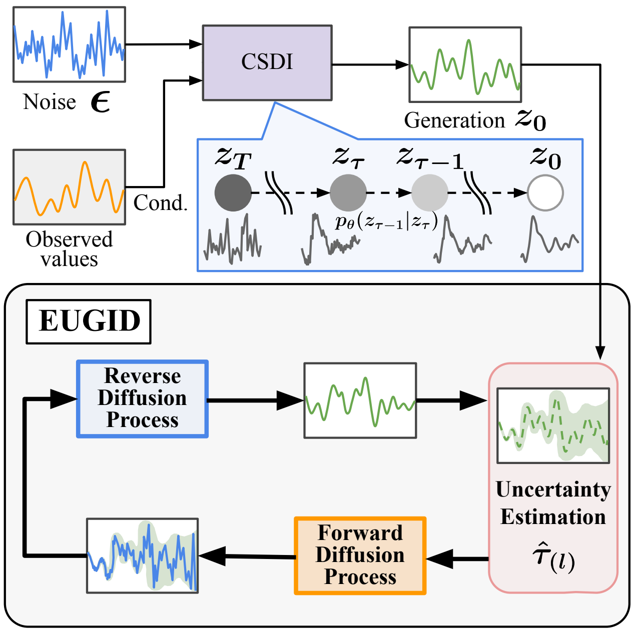

Figure 2: EUGID schema

2.3 Conditional Score-based Diffusion models for Imputation

TSF can be considered as conditional generation of future series y conditioning with input of past series x. Therefore, CSDI [Tashiro *et al.*, 2021] formulates diffusion model-based TSF as generating future series y with the condition of x. In other words, CSDI is a diffusion model with $z_0 = y \in \mathbb{R}^{L\times D}$ and a conditional noise decoder $\epsilon_\theta(z_\tau, \tau|x)$ instead of the unconditional decoder. CSDI uses the Transformer model as its model architecture of the encoder of condition and the noise decoder.

The decoding strategy for TSF is a procedure used to obtain the forecast series. CSDI employs a simple strategy applying the reverse diffusion process sequentially. The first step is to prepare a Gaussian noise $z_T \in \mathbb{R}^{L\times D}$ of the same size as the forecast series following Eq. (4). The outputs are then computed according to Eq. (5) until the final output z_0 is obtained.

3 Proposed Method

This section describes our proposal EUGID, which is the decoding strategy of the diffusion model for TSF to explore forecast series by iterating forward diffusion processes and reverse diffusion processes with the guide of empirically estimated uncertainty, as shown in Fig. 2. This method uses a trained model according to CSDI and replaces the decoding strategy used to generate the forecast series. To achieve better predictive exploration by sampling from the estimated uncertainty distribution like uncertainty sampling, EUGID reproduces the uncertainty distribution by the forward diffusion process, which means we estimate uncertainty in Gaussian, and forecasts the series again by a reverse diffusion process to output different series. This approach changes more uncertain points to a greater degree.

Algorithm 1 Reverse diffusion process for series diffused to different diffusion steps for each time step.

Input: Estimated diffusion time steps corresponding to the uncertainty: $\hat{\tau}^{(c)} \in \mathbb{N}^L$, Diffused series: $\hat{z}^{(c+1)}$
Output: Forecast series of next iteration: $z_0^{(c+1)}$

```
z_0^(c+1) ← z^(c+1)
τ̂_max ← max_l τ̂_l^(c)
τ' ← τ̂_max
while 1 ≤ τ' ≤ τ̂_max do
    o ← (1/√α_τ') ( z_τ' − ((1−β_τ')/√(1−α_τ')) ε_θ(z_τ', τ') )
    for 1 ≤ l ≤ L do
        if τ̂_l ≤ τ' then
            z_0^{l,(c+1)} ← e_l
        end if
    end for
    τ' ← τ' − 1
end while
```

3.1 Structure of EUGID

We will explain the core of EUGID, which is the iterative uncertainty estimation and the generation of forecast series with guiding uncertainty. EUGID expresses the uncertainty using the forward diffusion process. In other words, uncertainty estimation in EUGID estimates which diffusion step corresponds to uncertainty $\hat{\tau}$. Let the forecast series at c-th iteration be $y_0^{(c)}$, uncertainty for the c-th iteration at l-th time step is $\hat{\tau}_l^{(c)} = \mathrm{UncertaintyEstimation}(y_0^{(c)}, z^{(c)}, l)$.

For uncertainty sampling-like forecasting, the time series is diffused with the estimated diffusion steps, and the reverse diffusion process generates the forecast series for the next iteration. Eq. (7) represents the diffused series. Because the time series is diffused to a different diffusion step for each time step, a uniform reverse diffusion process is inappropriate.

$$\hat{z}^{(c+1)} = \left[y^{(c)}_{1,\hat{\tau}_1^{(c)}}, y^{(c)}_{2,\hat{\tau}_2^{(c)}}, \ldots, y^{(c)}_{L,\hat{\tau}_L^{(c)}} \right] \tag{7}$$

Therefore, we propose the reverse diffusion process starting from the most diffused time step as shown in Algorithm 1. This process achieves a reverse diffusion process at different levels at different time steps by removing noise from the time step of noisy parts and then gradually aligning the diffusion levels. Although this process uses a noise decoder in a different situation than training, we use this process because the diffusion model is empirically robust and worked well in the experiments in Section 4, so that it can work even if the classifier is used to guide the generation [Dhariwal and Nichol, 2021].

The original reverse diffusion process introduces a term of a small amount of noise $\sqrt{\beta_\tau}\epsilon$, ensuring the generated series's diversity. However, in CSDI, this approach failed to produce predictions with sufficient diversity. To address this, we introduced noise guided by uncertainty to enable more diverse predictions instead of a small amount of noise.

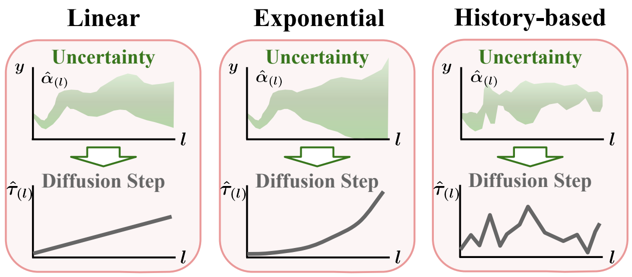

Figure 3: Overviews of our three empirical uncertainty estimators

Before the first iteration, similar to the existing model, we prepare the forecast series with a normal reverse diffusion process. The starting point of the generation is a Gaussian noise $z_T^{(0)} \sim \mathcal{N}(0, I)$. Then, the normal reverse diffusion process is applied to obtain $z_0^{(0)}$. Note that $z_0^{(0)}$ is equivalent to CSDI's output. After that, the forward diffusion processes and the reverse diffusion processes of Algorithm 1 using estimated uncertainty are iteratively applied to $z_0^{(0)}$. The iterations finish when the maximum number of iterations $c_{\max}$ is reached or when the estimated uncertainty is sufficiently small.

3.2 Empirical Uncertainty Estimation

This section describes how to estimate $\hat{\tau}$ as uncertainty, i.e., we explain the contents of $\mathrm{UncertaintyEstimation}(y_0^{(c)}, z^{(c)}, l)$. We propose three different uncertainty estimation methods to verify which estimation is appropriate. The uncertainty required here arises when forecasting using the diffusion model, and it is difficult to estimate it rigorously under these conditions. Furthermore, since the output of the diffusion model for a time series does not yield forecast distribution, uncertainty cannot be obtained by a simple method such as using the variance of the distribution. Therefore, an uncertainty estimation method suitable for EUGID is needed.

Therefore, we propose methods for empirically estimating uncertainty. The methods estimate uncertainty in the form of variance of the forecast series for each time step $\hat{\alpha}_l$. We then convert the variance $\hat{\alpha}_l$ to forward diffusion process steps by mapping it to α_τ, which represents the variance of the τ-th step in the forward diffusion process, as in Eq. (8).

$$\hat{\tau}_l = \operatorname*{argmin}_i |\hat{\alpha}_l - \alpha_i| \tag{8}$$

Specifically, we propose three methods: linear, exponential, and wavered reverse diffusion history-based uncertainty estimation as shown in Fig. 3. Informally, for the first two approaches, we assume that a greater degree of uncertainty be concomitant with the prediction of a longer-term future. For another approach, we assume the uncertainty could be estimated based on the diffusion history. These three approaches are experimented in iteratively reducing the future uncertainty in the decoding process, like simulated annealing.

The first two methods consider that the uncertainty of the forecast increases as the time step advances, which is a characteristic of TSF. They are independent of the forecast series. Eqs. (9) and (10) show their specific formulations, which assume linear and exponential increases in uncertainty with time steps, respectively.

$$\hat{\alpha}_l = \max\left(0, \min\left(1, al + \delta c + b\right)\right) \tag{9}$$

$$\hat{\alpha}_l = \max\left(0, \min\left(1, -(\gamma^c) \cdot e^{(pl+q)} + r\right)\right) \tag{10}$$

The wavered reverse diffusion history-based estimator formulates uncertainty as how much the model wavers in its predictions by weighting the number of times the increase and decrease switched in the reverse diffusion process. During the reverse diffusion process, the diffusion model explores the forecast series by moving it up or down as the diffusion step progresses. Assuming that the ups and downs indicate how much the diffusion model wavers in forecasting, we use the number of times it does so to estimate the uncertainty. Let the number of the switches be $s \in \mathbb{N}^{L \times D}$ computed by Eq. (11), Eq. (12) formulates this method.

$$s_l = \sum_{\tau=2}^{T} \frac{\mathrm{sign}(z_{\tau-2} - z_{\tau-1}) - \mathrm{sign}(z_{\tau-1} - z_\tau)}{2} \tag{11}$$

$$\hat{\alpha}_l = \max\left(0, \min\left(1, \frac{s_l}{\max_l s_l} + \lambda c\right)\right) \tag{12}$$

Linear and exponential estimators do not depend on the time series and assume uncertainty monotonically increases. On the other hand, the wavered reverse diffusion history-based estimator is not monotonic, and the uncertainty is not proportional to time because it depends on the transition of the reverse diffusion process.

4 Experiments

4.1 Experimental settings

We applied EUGID to CSDI and compared it with the original CSDI in our evaluation. In other words, the CSDI diffusion model is used as is, and the uncertainty estimation and forecasting using the uncertainty are repeated using CSDI model. We manually tuned the hyper-parameters for each decoding strategy as follows: for the linear strategy, $a = -0.0208, b = 0.1, \delta = 0.01$; for the exponential strategy, $p = 0.03, q = 0, r = 2, e = 2.7, \gamma = 0.99$; and for the reverse diffusion history-based strategy, $\lambda = 0.05$. The maximum value of iteration $c_{\max}$ in our experiments was set to 200. CSDI ensembles 100 models by averaging 100 forecast series. The hyper-parameters for the CSDI model were set following the settings described in previous research, where the details are described in supplementary materials.

We evaluate our proposed decoding strategy on eight real-world time series datasets of standard benchmarks for multivariate TSF, Electricity dataset [Rasul *et al.*, 2021], four types of Electricity Transformer Temperature (ETT) datasets [Zhou *et al.*, 2021] ETTm1, ETTm2, ETTh1 and ETTh2, Weather dataset [Zhou *et al.*, 2021], Wind dataset [Li *et al.*, 2022], and Solar dataset [Lai *et al.*, 2018]. We used the datasets and preprocessed them following existing research [Nguyen and Quanz, 2021; Salinas *et al.*, 2019; Tashiro *et al.*, 2021].

Model	CSDI		CSDI + EUGID					
Decoding Method	Original		Linear		Exponential		History-based	
Evaluation Metric	MSE	CRPS sum	MSE	CRPS sum	MSE	CRPS sum	MSE	CRPS sum
Electricity	**0.143**	**0.036**	**0.267**	**0.111**	0.268	0.117	0.468	0.204
ETTm1	0.147	0.150	0.110	0.117	**0.103**	**0.108**	**0.066**	**0.084**
ETTm2	0.055	0.058	**0.032**	**0.038**	0.031	0.038	**0.018**	**0.027**
ETTh1	0.242	0.208	0.185	**0.181**	**0.172**	0.182	**0.134**	**0.142**
ETTh2	0.093	0.085	**0.050**	**0.056**	0.051	0.057	**0.036**	**0.041**
Weather	0.265	**0.232**	**0.255**	0.246	0.261	0.255	**0.248**	**0.239**
Wind	0.287	0.022	0.028	**0.006**	**0.021**	**0.006**	**0.006**	**0.002**
Solar	0.091	0.959	0.014	0.225	**0.011**	**0.151**	**0.005**	**0.056**

Table 1: Forecasting performance of CSDI with EUGID and original CSDI for eight benchmark datasets

The datasets were split at a ratio of 7:1 for training and validation after taking seven samples for testing following the setting of CSDI [Tashiro *et al.*, 2021]. All features were normalized to have a mean of 0 and a variance of 1 for all experiments. We predict future 24 time steps using the past 168 time steps for all features of datasets as in previous studies [Nguyen and Quanz, 2021; Salinas *et al.*, 2019; Tashiro *et al.*, 2021]. Further details of the dataset are described in supplementary materials.

To evaluate prediction performance, we adopted two metrics to test set: MSE to evaluate forecasting performance and Continuous Ranked Probability Score (CRPS)-sum [Matheson and Winkler, 1976] to measure the distance between the distributions of gold data and forecast data. The CRPS-sum is the CRPS computed for the distribution of the sum of all time series across K features, capturing their joint effect. Both MSE and CRPS indicate a low error when their values are low. We report the averaged values of three-run results with different seeds.

4.2 Main Results

Table 1 shows the results of evaluation, in which EUGID improves CSDI's prediction performance. Specifically, EUGID outperformed the conventional method in terms of both MSE and CRPS-sum on seven out of eight datasets, except the Electricity dataset. Notably, a substantial improvement in MSE was observed for ETTm2, with approximately a 67% reduction.

All uncertainty estimation methods reduce prediction errors, especially when using uncertainty obtained with the wavered reverse diffusion history-based estimator. Since the monotone linear and exponential estimators show little difference, neither changes much. The wavered reverse diffusion history-based estimator performed further improvement than them, indicating that this method is suitable for modeling uncertainty in EUGID.

On the other hand, EUGID increased the errors in the forecasting performance of the Electricity dataset. This aspect confirms that there are datasets for which the EUGID is ineffective. When we looked at the method of estimating uncertainty, the history-based estimator, which performed the best on the other data sets, performed the worst, indicating that EUGID is counterproductive for the Electricity dataset. However, since the cause of poor performance for the Electricity dataset cannot be determined from these results, we will further investigate the cause in the analysis in the following Section 5.

Dataset	Kurtosis
Electricity	**3.987**
ETTm1	0.413
ETTm2	-0.746
ETTh1	-1.591
ETTh2	0.257
Weather	-1.118
Wind	-0.340
Solar	-1.306

Table 2: Fisher's kurtosis for each dataset

5 Discussion

This section visualizes the prediction results and EUGID iterations and analyzes the behavior of EUGID to analyze how EUGID contributes to improved prediction performance and in what cases it degrades performance. For this purpose, Fig. 4 visualizes forecast series in the test set in which EUGID improved the forecast performance and did not. While CSDI predictions lack diversity and multiple predictions with different seeds degenerate into similar predictions in Fig. 1, EUGID generates different predictions in Fig. 4.

Comparing the final predictions with the gold data, the series with better predictions were smoother, while those with worse predictions were noisier and changed more abruptly. In particular, the Electricity dataset, in which EUGID performed the poorest forecast performance, showed a series of rapid and repeated changes.

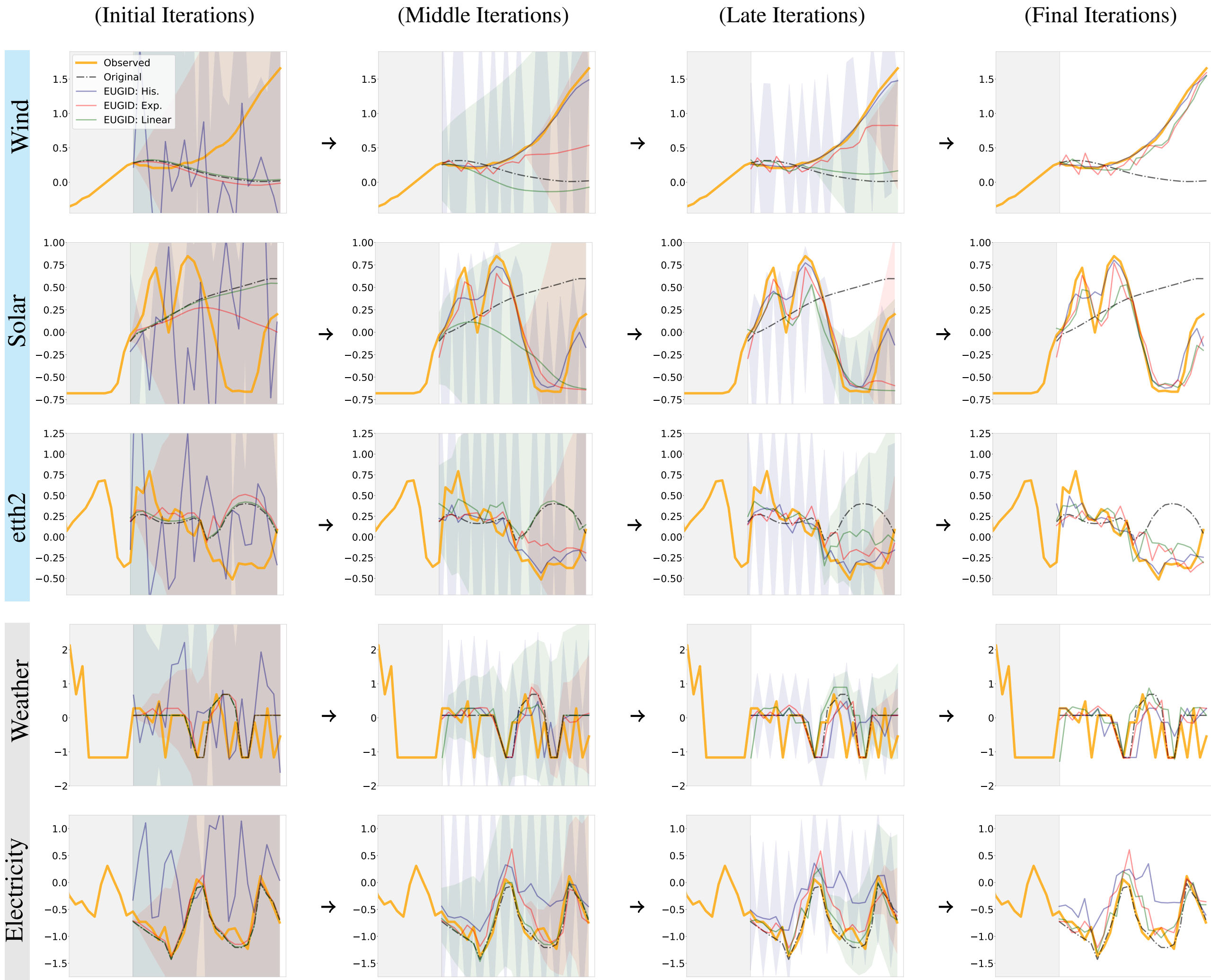

Figure 4: Visualization of forecast series and uncertainty distribution estimated with each uncertainty estimator of EUGID for manually selected data and variates in five datasets. Each column indicates the iteration of EUGID, where the left and right columns correspond to the initial and final iterations. The painted areas indicate the uncertainty distributions in the $\pm 5\sigma$ range.

In order to see how sharply the characteristics of each dataset change, Table 2 shows the average kurtosis of the samples in the dataset, where we computed Fisher's kurtosis using SciPy [Virtanen *et al.*, 2020]. The Electricity dataset has particularly high kurtosis, indicating that it is a dataset with many peaks. EUGID can degrade performance for data with such characteristics.

Next, we observe how the predictions evolve with each iteration. When we look at the improving cases, we can see that the forecast series approaches the gold data for all uncertainty estimation methods as the iterations progress. Comparing the uncertainty estimation methods, the series for the exponential estimator reaches closer to the gold data earlier in the iterations than one for the linear estimator. Moreover, for the wavered reverse diffusion history-based estimator, the uncertainty is higher at the point with a high rate of change, resulting in a prediction series closer to the gold data even earlier, indicating the estimator is able to model the uncertainty suitable for use in EUGID.

6 Related Work

6.1 Time Series Forecasting

TSF is the task of predicting future sequences based on past sequences. TSF has various applications, such as predicting future electricity demand [Blum and Riedmiller, 2013] and optimizing resource planning and inventory management in warehouses [Gómez-Losada and Duch-Brown, 2019]. Given its critical role across diverse domains,TSF continues to be an active area of research.

The state-of-the-art performance is shown when formulated as a regression task where the output of the model is directly a forecast series. In particular, transformer-based models [Yang *et al.*, 2024; Liu *et al.*, 2024] show state-of-the-art performance with engineering to capture time-series-specific characteristics.

On the other hand, generation-based TSF is a recently hot topic to make it applicable to other tasks, such as time series generation, or use metadata as conditions. However, the forecasting performance for TSF has been limited. In particular, TSF based on diffusion model generation [Rasul *et al.*, 2021; Tashiro *et al.*, 2021; Alcaraz and Strodthoff, 2023; Shen and Kwok, 2023], which has been successful in other fields, does not perform as well as other models, and there is much room for improvement.

6.2 Uncertainty Sampling

Uncertainty Sampling is one of the common strategies in active learning [Zhu *et al.*, 2008; Raj and Bach, 2022], where the data deemed to have the highest uncertainty in the current prediction model are selected for the next labeling. Recently, uncertainties of tokens have been used as a penalty for diverse text generation, successfully achieving text decoding with improved quality of generation [Garces Arias *et al.*, 2024]. This study empirically estimates the uncertainty of forecast series generated by the model and applies the mechanism of uncertainty sampling to TSF by using the diffusion model.

7 Conclusion and Future Work

In this study, we proposed a novel decoding strategy EUGID for diffusion model-based TSF, which iterates forward and reverse diffusion processes with a guide of empirically estimated uncertainty distribution. EUGID enables exploring series in the distribution without deviating from forecast series at the previous iteration. We conducted experiments using eight standard benchmark datasets for TSF to evaluate its effectiveness. EUGID improved the forecasting performance of the base model of CSDI in seven out of eight datasets. The analysis with visualization of the forecast series for each iteration confirmed that the forecast series closed to gold data as the iterations progressed. Furthermore, as a limitation, we confirmed that it is difficult for EUGID to cope with noisy data with many peaks.

In the future, we will tackle several remaining challenges: (1) evaluating more comprehensively by applying to other diffusion models [Rasul *et al.*, 2021; Alcaraz and Strodthoff, 2023; Shen and Kwok, 2023] to reveal EUGID generalizability; (2) evaluating on the recently mainstreamed long-term TSF [Zhou *et al.*, 2021]; (3) improving the applicability of the noisy data, which is limitation revealed in our analysis.

References

[Alcaraz and Strodthoff, 2023] Juan Lopez Alcaraz and Nils Strodthoff. Diffusion-based time series imputation and forecasting with structured state space models. *Transactions on Machine Learning Research*, 2023.

[Blum and Riedmiller, 2013] Manuel Blum and Martin Riedmiller. Electricity demand forecasting using gaussian processes. In *Proceedings of the 15th AAAI Conference on Trading Agent Design and Analysis*, AAAIWS'13-15, page 10–13. AAAI Press, 2013.

[Box *et al.*, 2015] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley, 2015.

[Chen *et al.*, 2021] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation. In *International Conference on Learning Representations*, 2021.

[Dhariwal and Nichol, 2021] Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc.

[Erlygin *et al.*, 2023] Leonid Erlygin, Vladimir Zholobov, Valeriia Baklanova, Evgeny Sokolovskiy, and Alexey Zaytsev. Surrogate uncertainty estimation for your time series forecasting black-box: learn when to trust. In *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1247–1258, 2023.

[Garces Arias *et al.*, 2024] Esteban Garces Arias, Julian Rodemann, Meimingwei Li, Christian Heumann, and Matthias Aßenmacher. Adaptive contrastive search: Uncertainty-guided decoding for open-ended text generation. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 15060–15080, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[Gómez-Losada and Duch-Brown, 2019] Álvaro Gómez-Losada and Néstor Duch-Brown. Time series forecasting by recommendation: An empirical analysis on amazon marketplace. In Witold Abramowicz and Rafael Corchuelo, editors, *Business Information Systems*, pages 45–54. Springer International Publishing, 2019.

[Gong *et al.*, 2023] Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and Lingpeng Kong. Diffuseq: Sequence to sequence text generation with diffusion models. In *The Eleventh International Conference on Learning Representations*, 2023.

[Henrique *et al.*, 2019] Bruno Miranda Henrique, Vinicius Amorim Sobreiro, and Herbert Kimura. Literature review: Machine learning techniques applied to financial market prediction. *Expert Syst. Appl.*, 124(C):226–251, June 2019.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

[Kollovieh *et al.*, 2023] Marcel Kollovieh, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Yuyang (Bernie) Wang. Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting. In A. Oh, T. Naumann,

A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 28341–28364. Curran Associates, Inc., 2023.

[Lai *et al.*, 2018] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, page 95–104, New York, NY, USA, 2018. Association for Computing Machinery.

[Li *et al.*, 2022] Yan Li, Xinjiang Lu, Yaqing Wang, and Dejing Dou. Generative time series forecasting with diffusion, denoise, and disentanglement. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 23009–23022. Curran Associates, Inc., 2022.

[Liu *et al.*, 2024] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

[Matheson and Winkler, 1976] James E. Matheson and Robert L. Winkler. Scoring Rules for Continuous Probability Distributions. *Management Science*, 22(10):1087–1096, 1976.

[Möller and Reuter, 2007] Bernd Möller and Uwe Reuter. *Uncertainty Forecasting in Engineering*. Springer, 2007.

[Narasimhan *et al.*, 2024] Sai Shankar Narasimhan, Shubhankar Agarwal, Oguzhan Akcin, Sujay Sanghavi, and Sandeep P. Chinchali. Time weaver: A conditional time series generation model. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 37293–37320. PMLR, 21–27 Jul 2024.

[Nguyen and Quanz, 2021] Nam Nguyen and Brian Quanz. Temporal latent auto-encoder: A method for probabilistic multivariate time series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10):9117–9125, 2021.

[Raj and Bach, 2022] Anant Raj and Francis Bach. Convergence of uncertainty sampling for active learning. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 18310–18331. PMLR, 17–23 Jul 2022.

[Rasul *et al.*, 2021] Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8857–8868. PMLR, 18–24 Jul 2021.

[Rombach *et al.*, 2021] Robin Rombach, A. Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685, 2021.

[Sadat *et al.*, 2024] Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann M. Weber. CADS: Unleashing the diversity of diffusion models through condition-annealed sampling. In *The Twelfth International Conference on Learning Representations*, 2024.

[Salinas *et al.*, 2019] David Salinas, Michael Bohlke-Schneider, Laurent Callot, Roberto Medico, and Jan Gasthaus. High-dimensional multivariate forecasting with low-rank gaussian copula processes. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[Shen and Kwok, 2023] Lifeng Shen and James T. Kwok. Non-autoregressive conditional diffusion models for time series prediction. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

[Simon *et al.*, 2018] Christophe Simon, Philippe Weber, and Mohanmed Sallak. *Data Uncertainty and Important Measures (Volume 3)*. Wiley, 2018.

[Smith, 2013] Ralph C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, 2013.

[Song *et al.*, 2021] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[Sullivan, 2015] T.J. Sullivan. *Introduction to Uncertainty Quantification*. Springe, 2015.

[Tashiro *et al.*, 2021] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: conditional score-based diffusion models for probabilistic time series imputation. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc.

[Virtanen *et al.*, 2020] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef

Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[Yang *et al.*, 2024] Yingnan Yang, Qingling Zhu, and Jianyong Chen. Vcformer: Variable correlation transformer with inherent lagged correlation for multivariate time series forecasting. In Kate Larson, editor, *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 5335–5343. International Joint Conferences on Artificial Intelligence Organization, 2024.

[Zhou *et al.*, 2021] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):11106–11115, May 2021.

[Zhu *et al.*, 2008] Jingbo Zhu, Huizhen Wang, Tianshun Yao, and Benjamin K Tsou. Active learning with sampling by uncertainty and density for word sense disambiguation and text classification. In Donia Scott and Hans Uszkoreit, editors, *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 1137–1144, Manchester, UK, August 2008. Coling 2008 Organizing Committee.