

TimeDP: Learning to Generate Multi-Domain Time Series with Domain Prompts

Yu-Hao Huang
huangyh@smail.nju.edu.cn
National Key Laboratory for Novel
Software Technology,
School of Computer Science, Nanjing
University
Nanjing, China

Chang Xu*
chanx@microsoft.com
Microsoft Research Asia
Beijing, China

Yueying Wu
wuyueying@stu.pku.edu.cn
Peking University
Beijing, China

Wu-Jun Li
liwujun@nju.edu.cn
National Key Laboratory for Novel
Software Technology,
School of Computer Science, Nanjing
University
Nanjing, China

Jiang Bian
jiang.bian@microsoft.com
Microsoft Research Asia
Beijing, China

Abstract

Time series generation models are crucial for applications like data augmentation and privacy preservation. Most existing time series generation models are typically designed to generate data from one specified domain. While leveraging data from other domain for better generalization is proved to work in other application areas, this approach remains challenging for time series modeling due to the large divergence in patterns among different real world time series categories. In this paper, we propose a multi-domain time series diffusion model with domain prompts, named **TimeDP**. In TimeDP, we utilize a time series semantic prototype module which defines time series prototypes to represent time series basis, each prototype vector serving as “word” representing some elementary time series feature. A prototype assignment module is applied to extract the domain specific prototype weights, for learning domain prompts as generation condition. During sampling, we extract “domain prompt” with few-shot samples from the target domain and use the domain prompts as condition to generate time series samples. Experiments demonstrate that our method outperforms baselines to provide the state-of-the-art in-domain generation quality and strong unseen domain generation capability.

CCS Concepts

• **Computing methodologies** → **Machine learning algorithms**.

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/18/06
<https://doi.org/XXXXXXX.XXXXXXX>

Keywords

Time Series Generation, Cross-Domain

ACM Reference Format:

Yu-Hao Huang, Chang Xu, Yueying Wu, Wu-Jun Li, and Jiang Bian. 2018. TimeDP: Learning to Generate Multi-Domain Time Series with Domain Prompts. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 15 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

In the landscape of large models and advanced machine learning techniques, time series foundation models [3, 6] have garnered increasing attention. These models, typically trained on extensive datasets spanning various domains [23], have predominantly emphasized forecasting tasks rather than the generation of new data. However, the accurate and meaningful generation of time series is critical for applications such as medical record synthetic [17] and financial scenario simulations [2, 9], as well as for augmenting datasets where historical records are limited or incomplete [13].

Although some research has been conducted on time series generation, most efforts have been confined to the development of generation model for single-domain data. In contrast, cross-domain time series generation presents a significantly more complex challenge, as it requires the creation of new data across various domains without relying on existing historical records. This stands out as a gap, underscoring a substantial opportunity for further advancements in multi-domain time series generation.

One straightforward approach to multi-domain time series generation involves the use of predefined domain labels during the training process [16]. This method relies on the availability of domain labels to formulate the conditional generation process. However, this approach may struggle generalizing to large number of domains or unseen domains. Moreover, the challenge intensifies when domain labels are not explicitly available.

An alternative approach frames cross-domain time series generation as a conditional generation task by describing the domain using natural language [11, 12]. However, the use of natural language descriptions introduces significant challenges. Domain-specific nuances are often difficult to articulate precisely, leading to noisy, incomplete, or ambiguous prompts. Moreover, for entirely new or evolving domains, crafting these domain descriptors can be impractical. This has underscored a critical need for a more systematic and robust way to represent and utilize domain-specific information in time series generation.

To address these challenges, we propose a label-free, text-free method that learns time series prototypes as basic elements to construct domain prompts for generating time series with a diffusion model, named **TimeDP**. Through training, the prototypes learn to represent time series basis, serving as “word” with time series semantics. A prototype assignment module is applied for each training samples to construct the specific “prompt” for generating this sample. During sampling, we extract “prompt” with few-shot samples from the target domain to construct the population of domain prompts and use the domain prompts as condition to generate time series samples.

To summarize, the main contributions of this paper are listed as follows:

- We propose **TimeDP**, a multi-domain time series generation model by learning a set of time series prototypes and prototype assignment module to construct domain prompts, where the domain prompts serve as condition for a time series diffusion model.
- We are the first to propose a multi-domain time series generation model using label-free, text-free conditioning mechanism.
- Experiments demonstrate that our method outperforms baselines with the state-of-the-art in-domain generation quality, and strong unseen domain generation capability.

2 Related Work and Backgrounds

2.1 Time Series Generation

Existing time series generation models has based on various foundational type of generative models. GAN-based methods has been introduced to encourage the network to consider temporal dynamic by jointly optimize both supervised and adversarial objectives for a learned embedding space [25]. VAE-based methods have designed specific decoder structure for temporal data considering trend and seasonal decomposition [4], and first introduces vector quantization technique together with bidirectional transformers to better capture temporal consistency [16]. Another category is considered as mixed-type methods, combining GANs, flows and ODEs [10]. Different from these methods, we utilize denoising diffusion probabilistic models (DDPM) as our generation backbone.

Existing diffusion-based time series generation methods leverage both unconditional and conditional diffusion models for generating time series data with various denoising network backbones [14, 24]. Researchers have also considered combining diffusion models with the constrained generation problem [1] and the extraction of time series intrinsic such as seasonal-trend decomposition techniques [26]. Compared with these single-domain methods, we first propose to

utilize label-free, text-free domain prompts as condition for generating time series.

2.2 Cross-Domain Time Series Model

There have been several recent work consider utilizing multiple-domain time series data for training time series foundation models. These works can be divided into two branches. The first branch builds two-stage models. Kraus et al. [15] pretrains a representation learning model on 75 datasets for the first stage and finetuning to task specific models at the second stage. Gao et al. [6] conducts masked reconstruction pretraining on 38 multi-domain datasets for the first stage and a multi-task supervised learning for downstream tasks. The second branch pretrains end-to-end transformer models with patch tokenizers for time series forecasting. Woo et al. [23] pretrains on a dataset with over 2B observations and Das et al. [3] pretrains on a dataset with 100B time-points, both using patching and instance normalizing to unify across different data scale, frequencies and lengths. These methods employ instance normalization to generate forecasts based on historical data without explicitly addressing domain differences. Compared with these approaches, we propose to use time series prototypes, constructing domain prompts to explicitly distinguish domains as well as bridge them.

2.3 Denoising Diffusion Probabilistic Models (DDPMs)

A diffusion probabilistic model [22] learns to reverse the transitions of a Markov chain which is known as the diffusion process that gradually adds noise to data, ultimately destroying the signal.

Let $\mathbf{x}_0 \in \mathbb{R}^d \sim q(\mathbf{x}_0)$ be real data of dimension d from space \mathcal{X} . The diffusion process generates $\mathbf{x}_1, \dots, \mathbf{x}_N$ from the same space with the same shape as \mathbf{x}_0 , using a Markov chain that adds Gaussian noise over N time steps: $q(\mathbf{x}_1, \dots, \mathbf{x}_N | \mathbf{x}_0) := \prod_{n=1}^N q(\mathbf{x}_n | \mathbf{x}_{n-1})$. The transition kernel is commonly defined as:

$$q(\mathbf{x}_n | \mathbf{x}_{n-1}) := \mathcal{N}(\mathbf{x}_n; \sqrt{1 - \beta_n} \mathbf{x}_{n-1}, \beta_n \mathbf{I}), \quad (1)$$

where $\{\beta_n \in (0, 1)\}_{n=1, \dots, N}$ defines the variance schedule. Note that \mathbf{x}_n at any arbitrary time step n can be derived in a closed form $q(\mathbf{x}_n | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_n; \sqrt{\bar{\alpha}_n} \mathbf{x}_0, (1 - \bar{\alpha}_n) \mathbf{I})$, where $\alpha_n := 1 - \beta_n$ and $\bar{\alpha}_n := \prod_{s=1}^n \alpha_s$. For the reverse process, the diffusion model, parameterized by θ , yields:

$$p_\theta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N) := p(\mathbf{x}_N) \prod_{n=1}^N p_\theta(\mathbf{x}_{n-1} | \mathbf{x}_n), \quad (2)$$

where $p_\theta(\mathbf{x}_{n-1} | \mathbf{x}_n) := \mathcal{N}(\mathbf{x}_{n-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_n, n), \Sigma_\theta(\mathbf{x}_n, n))$ and the transitions start at $p(\mathbf{x}_N) = \mathcal{N}(\mathbf{x}_N; \mathbf{0}, \mathbf{I})$. The optimization objective is derived into the maximizing of an approximation of the evidence lower bound (ELBO) of the log-likelihood $\log p_\theta(\mathbf{x}_0)$. With a widely adopted parameterization:

$$\boldsymbol{\mu}_\theta(\mathbf{x}_n, n) = \frac{1}{\sqrt{\alpha_n}} (\mathbf{x}_n - \frac{\beta_n}{\sqrt{1 - \bar{\alpha}_n}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_n, n)), \quad (3)$$

the training is performed to predict the noise term added in the forward process which simplifies the objective to:

$$L_{simple} := \mathbb{E}_{\mathbf{x}_0, \boldsymbol{\epsilon}, n} [\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_n} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_n} \boldsymbol{\epsilon}, n)\|^2]. \quad (4)$$

On sampling, $\mathbf{x}_{n-1} = \frac{1}{\sqrt{\alpha_n}}(\mathbf{x}_n - \frac{1-\alpha_n}{\sqrt{1-\alpha_n}}\epsilon_\theta(\mathbf{x}_n, n)) + \sigma_n \mathbf{z}$, where $\sigma_n = \sqrt{\beta_n}$ and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ [7].

Typical time series diffusion models for forecasting task [5, 20, 21] encodes history context into c as condition and make use of the conditional form of DDPMs [8] for generating future time series:

$$p_\theta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N | c) := p(\mathbf{x}_N) \prod_{n=1}^N p_\theta(\mathbf{x}_{n-1} | \mathbf{x}_n, c), \quad (5)$$

In the problem setting of time series generation, the generation process does not rely on time series history. We explore the use of term c to provide domain semantics for time series generation model in this work.

2.4 Problem Formulation

Let $D_i^T = \{\mathbf{x} \in \mathbb{R}^T\}^{N_i}$, $\mathbf{x} = (x_1, x_2, \dots, x_T)$ denote a time series dataset domain i with N_i time series samples, where each sample contains T sequential values. A straightforward single-domain time series generation model fits the joint distribution over time steps $p(x_1, x_2, \dots, x_T)$ of each dataset with a separate model parameterized by θ_i , namely $p_{\theta_i}(x_1, x_2, \dots, x_T)$ for all \mathbf{x} in D_i^T .

In this paper, we explore a domain-unified setting where the mixture of M domain datasets with sequence length T is denoted by $D^T = \bigcup_{i=1}^M D_i$ and we aim to build one model for the mixed dataset parameterized by θ , namely $p_\theta(x_1, x_2, \dots, x_T | i)$ for all \mathbf{x} in D^T .

Adhering to the channel-independent setting [19] that is widely accepted by recent researches, we formulate the problem studied in this paper in a uni-variate time series generation manner to handle the heterogeneity of time series in terms of dimension [23].

3 Methodology

With sequences from all data domains mixed together during training, all time series features within latent representation are entangled without a explicit way for distinguishing specific time series data domain. Although utilizing domain labels as class labels for training the time series generation model can provide instruction for identifying specific domain, this approach implies an assumption that all domains are independent from each other, neglecting the different similarity levels among domain pairs. Therefore, it is challenging to equip the model with the ability of generating time series in selected domain while considering the inter-relationship among domains. To overcome this challenge, the key is to build a triggering mechanism for cross-domain time series model that can control the model for generating time series data from specific domain. Motivated by the recent advancements in controllable content generation with prompting technique, we propose to construct domain prompts for controlling a cross-domain model.

In the rest of this section, we first describe the model architecture design. Then, we describe the optimization objective and training algorithm of the proposed model. Finally, we discuss the procedure for in-domain sampling and unseen domain generation using the proposed model.

3.1 Domain Prompts

Different from text and image modality where the generation target can be expressed by natural language or categorized into discrete classes, it is difficult to obtain explicit representation of time series with words or class labels. Inspired by the widely adopted technique to extract ‘‘basis’’ which are the elementary features of time series [18], these basis can be utilized as the shared ‘‘dictionary’’ among different domains, each of which encodes different semantic feature for time series.

3.1.1 Semantic Prototype Module. Each basis represents certain elementary time series feature like trend and seasonality, that may exist in time series data samples. Different individual time series samples are assumed to share the same collection of basis but reflect distinct subset of the collection. As a result, each time series gets unique realization of these underlying features, similar to variable weighted allocations to all the basis. Based on this assumption, a set of latent arrays is introduced as time series prototypes $P \in \mathbb{R}^{N_p \times d}$ for representing cross-domain time series common knowledge, where each prototype vector $p \in \mathbb{R}^{1 \times d}$ serves as the representation of a time series basis. In practice, the time series prototypes P are initialized with random orthogonal vectors and are frozen afterwards.

3.1.2 Prototype Assignment Module (PAM). Given the assumption that each time series sample corresponds to a distinct allocation of all the basis, the mapping from time series samples to the allocations needs to be established for explicitly identifying important prototypes for each time series instance as well as distinguishing among domains. We propose to extract a prototype assignment for each time series instance as the importance weights of each time series on each prototype, and the prototype assignments then serve as conditions for the generation model.

Specifically, each input sequence \mathbf{x} is mapped into a weight vector whose dimension equals to the number of prototypes using weight extractor ϕ , which is a neural network. The vector $\phi(\mathbf{x})$ represents the weight of each vector inside P , and the weights are utilized to modify the attention weight within the cross-attention mechanism so that the predicted noises are only conditioned on the assigned prototypes. Therefore, sequences from different domains are represented by different \mathbf{m} weighted combinations of the shared same set of time series prototypes. For ensuring sparsity on prototype assignments, all negative weights are discarded when conducting prototype assignment. Formally, the prototype assignments \mathbf{m} is extracted with the following formula:

$$\mathbf{m} = \phi(\mathbf{x}_0) - \mathbf{I}_{\phi(\mathbf{x}_0) < 0} \cdot \infty, \quad (6)$$

where $\mathbf{I}_{<0}$ is the indicator function of negative elements.

3.2 Domain-Unified Training

Instead of training individual model for each specific dataset, we train one model with data from multiple datasets at the same time for generating different domain data. Here, we treat each dataset as a separate domain. While data from each domain only represent limited fraction of possible data distribution, leveraging data from other domain can help model capture a more diverse time series data distribution.

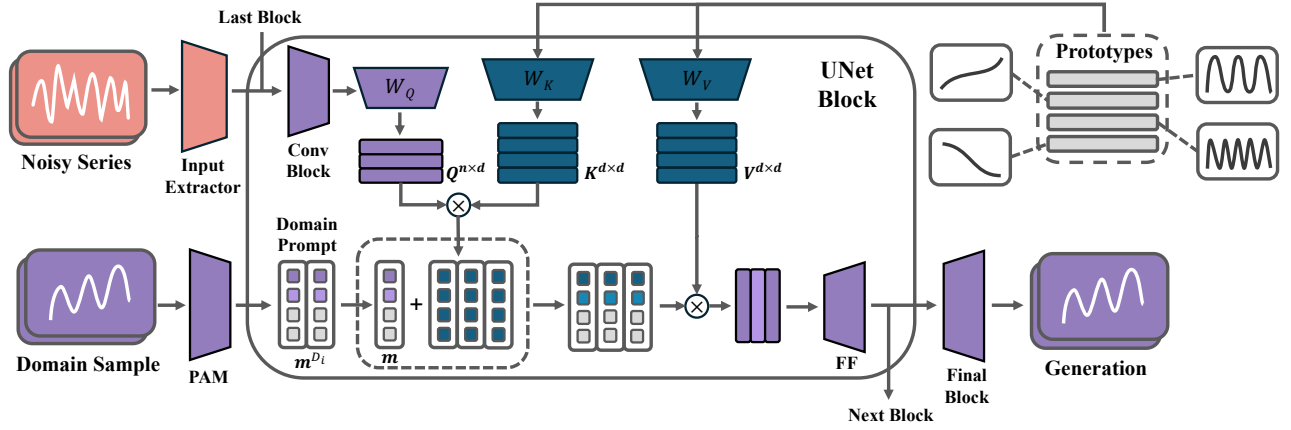


Figure 1: Overview of TimeDP model.

Other than taking the unconditional denoising diffusion objective stated in eq. (4), we employ the conditional denoising objective using c from eq. (5) as condition to make use of the encoded semantic context for denoising process, where the conditions are incorporated into the intermediate layers of noise prediction network by spatial attention.

$$Q^{(i)} = z^{(i-1)} \cdot W_Q^{(i)}, \quad (7)$$

$$K^{(i)} = P \cdot W_K^{(i)}, \quad V^{(i)} = P \cdot W_V^{(i)}, \quad (8)$$

$$z^{(i)} = \text{FF}(\text{softmax}(\frac{Q^{(i)} K^{(i)T}}{\sqrt{d}} + m) \cdot V^{(i)}), \quad (9)$$

where $z^{(i)} \in \mathbb{R}^{N \times d}$ denotes the output of the i^{th} last U-Net block. $W_Q^{(i)} \in \mathbb{R}^{d \times d}$, $W_K^{(i)} \in \mathbb{R}^{d \times d}$ and $W_V^{(i)} \in \mathbb{R}^{d \times d}$ are learnable projection matrices applied on the sequence dimension. FF denotes feed forward layer. The attention output z^{final} is followed by another feed forward network to produce final block output $\hat{e} = \text{FF}(z^{\text{final}})$.

With the conditional denoising mechanism described above, the denoising objective using ϵ -parameterization can be written and simplified into:

$$L_{\text{cond}} = \mathbb{E}[\|e - \hat{e}\|^2] \\ = \mathbb{E}_{\mathbf{x}_0 \in D^T, \epsilon \sim \mathcal{N}(0, I), n} [\|e - e_{\theta, P}(\mathbf{x}_n, n, \mathbf{m})\|^2]. \quad (10)$$

Due to the imbalance number of training samples across domain, we adopt a re-weight sampling method for making the probability equal for training on samples from each domain. Let N_i denote the number of sample sequences in dataset i , we set the weight for sampling each sample of this dataset as $w_i = \frac{1}{N_i * |D|}$, such that the probability for sampling sequence from each dataset is balanced. The pseudo code for training algorithm is shown at Algorithm 1.

3.3 Generation with Domain Prompt

To generate time series samples of selected domain after the domain-unified training on multiple datasets, we first extract domain-specific

Algorithm 1 Training algorithm

Require: Sequence sample \mathbf{x}

Ensure: Network parameters ϕ and θ , prototypes P

- 1: Initialize prototypes P
 - 2: **repeat**
 - 3: Sample \mathbf{x}_0 from D^T
 - 4: Extract prototype assignments \mathbf{m} according to eq. (6)
 - 5: Randomly set P as unconditional identifier \mathbf{p}_u
 - 6: Randomly sample time step $n \sim \mathcal{U}(1, N)$
 - 7: Randomly sample noise $\epsilon \sim \mathcal{N}(0, I)$
 - 8: Corrupt data $\mathbf{x}_n = \sqrt{\alpha_n} \mathbf{x}_0 + \sqrt{1 - \alpha_n} \epsilon$
 - 9: Predict step noise with $\tilde{\epsilon} = \tilde{\epsilon}_{\theta, P}(\mathbf{x}_n, n, \mathbf{m})$
 - 10: Compute loss with Equation (10) and take gradient step.
 - 11: **until** maximum training step
-

prototype assignments of a small random subset of training samples for the selected domain and group them into a distribution of domain prompt representing the selected domain. Let K denotes the number of selected samples from dataset i , the domain prompt is denoted by $\mathbf{m}^{D_i} = \{\mathbf{m}_1^i, \dots, \mathbf{m}_K^i\}$. By constructing the conditioning input, the model generates samples adhere to the selected domain while is not constrained by the general temporal patterns exhibited in the selected samples. When the number of expected generated samples is larger than K , we use a strategy of repeatedly generate with each assignment in K samples until the number of expected samples is satisfied.

The sampling algorithm is described as Algorithm 2.

Algorithm 2 Sampling with domain prompts

Require: K time series prompts \mathbf{x} , prototypes P

Ensure: Generated time series samples $\hat{\mathbf{x}}$

- 1: Extract prototype prompts \mathbf{m} with \mathbf{x} according to eq. (6)
 - 2: Randomly sample noise $\hat{\mathbf{x}}_N \sim \mathcal{N}(0, I)$
 - 3: **for** n from N to 1 **do**
 - 4: Predict step noise with $\tilde{\epsilon}_n = \tilde{\epsilon}_{\theta, P}(\hat{\mathbf{x}}_n, n, \mathbf{m})$
 - 5: Denoise $\hat{\mathbf{x}}_{n-1} = \frac{\hat{\mathbf{x}}_n - \sqrt{1 - \alpha_n} \tilde{\epsilon}_n}{\sqrt{\alpha_n}}$
 - 6: **end for**
 - 7: $\hat{\mathbf{x}} = \hat{\mathbf{x}}_0$
-

3.4 Unseen Domain Generation

Since the prototypes provide representations for time series basis, their representation ability is not restricted to the domains in training sets. Therefore, they can be utilized to represent unseen domain or datasets. For any unseen dataset or domain D_j with respect to the training set, we can use extract “few-shot” samples x_1^j, \dots, x_K^j from the dataset to construct domain prompt $m^{D_j} = \{m_1^j, \dots, m_K^j\}$, and then feed them into the model as condition to generate new samples of the required dataset.

4 Experiments

In this section, we provide empirical experiment results for our method using multiple real-world datasets. The experiment goal is to investigate on the following research questions: (a) How good is the quality of prompted generation on trained domain? (b) Can the learned prototype help the model generalize to unseen domains?

4.1 Experiment Settings

Datasets. The experiments are conducted on 12 datasets across four time series domains: Electricity, Solar and Wind from the energy domain; Traffic, Taxi and Pedestrian from the transport domain; Air Quality, Temperature and Rain from the nature domain; NN5, Fred-MD and Exchange from the economic domain. All datasets are obtained by GluonTS package and Monash Time Series Forecasting Repository. We pre-process all datasets into non-overlapping uni-variate sequence slices with length in $\{24, 96, 168, 336\}$. More details on datasets are described in the technical appendix.

Baselines. We compare our generation results with several representative state-of-the-art time series generation methods, including TimeGAN [25], GT-GAN [10], TimeVAE [4], and TimeVQVAE [16]. Note that in their original implementation, these methods are trained with single dataset from a single domain for each time. To align their implementations with our problem setting and method, we train these models with the multi-domain dataset which mix all of the twelve datasets together. Note that the TimeVQVAE baseline has a class-conditioned variant, which we considered as a conditional time series generation baseline that is able to sample by domain label after being trained with the multi-domain dataset.

Implementation Details. We follow common practice of diffusion models to utilize a U-Net architecture for our denoising model. The architecture details are discussed in the technical appendix. The number of prototypes are set to 16 for all the main evaluations. Models for each sequence length are trained for 50,000 steps using a batch size of 128 and a learning rate of 1×10^{-4} with 1,000 warm-up steps. For all baselines, we take the implementations and recommended hyper-parameter settings from their public codes.

Evaluation Metrics. The major consideration for evaluating performance of time series generation methods are two fold: (1) the similarity between real and synthetic time series distributions; (2) the internal temporal dependency within each instance. In light of these two we apply the following three metrics for evaluating generation performance: (1) **Maximum Mean Discrepancy (MMD)** which maps the data points into a high-dimensional feature space with a kernel function and then compare the means of the two data

distribution; (2) **Kullback-Leibler divergence (K-L)** which is the measure of how one probability distribution diverges from a second, reference probability distribution, indicating the distribution difference on variable level; (3) **Marginal distribution difference (MDD)** which calculates the empirical histogram for each time step and calculate the average absolute difference of real and synthetic data across bins.

4.2 Results

4.2.1 Evaluation of Generation Quality. In this experiment, we evaluate the time series generation performance of our model on all 12 datasets across four domains, comparing with the baselines on MMD, K-L and MDD. The proposed model and baseline models are evaluated after trained with the multi-domain dataset. While the baselines are not designed for learning time series from multiple domains simultaneously, we modify them to align with our setting that treating the multi-domain dataset as a whole. Additionally, while TimeVQVAE can be trained in a class-conditional manner, we also test on this variant, using the source of data domain as class label to obtain a cross-domain version, as shown in column “TimeVQVAE-C”. Each run is repeated 5 times with different random seeds. The average and the standard deviation of MMD and K-L results for the sequence length of 168 are shown in Table 1. The MDD results and other results for different sequence length settings are reported in the appendix.

As shown in Table 1, our approach achieves the best results on most of the twelve in-domain datasets, demonstrating superior performance on generating new time series samples that have the closest distribution to real dataset samples, both jointly and marginally. Most second best scores are obtained with the class-conditional version of TimeVQVAE model. The results indicates that, simply mixing datasets as the multi-domain training approach generally fails to generate samples that are close to real data in distribution, majorly due to the diverse intrinsic pattern among different domains. Although using dataset labels helps greatly on the training with mixed data as shown with TimeVQVAE-C, our model outperforms it without being explicitly supervised with class labels, indicating strong representation disentanglement capability. The results for other sequence length settings are included in the technical appendix. The results show that TimeDP consistently demonstrate the best generation performance on both long and short sequence length scenarios.

4.2.2 Evaluation of Unseen Domain Time Series Synthesis. In this experiment, we evaluate the synthesis performance on unseen domain, where datasets that are not included in the training data. We randomly select a few samples from the new datasets as demonstrations for all models, and ensure a test set that does not overlap with these samples. For TimeDP, these samples are used to extract domain prompts. For baseline methods that are not designed to generate following prompts or can not generate with unseen class labels, we evaluate their unconditional generation outputs. The results for these baseline models are labeled with “unconditional” at the table. Additionally, we conduct a fine-tuning on all baseline models to evaluate their few-shot learning setting. The fine-tuned model are labeled with “Fine-tuned” at the table. We use 3, 10 and 100 samples as “prompt” and fine-tuning data in this experiment.

		TimeDP	TimeGAN	GT-GAN	TimeVAE	TimeVQVAE	TimeVQVAE-C
Maximum Mean Discrepancy	Electricity	0.001 ± 0.001	0.367 ± 0.255	0.254 ± 0.166	0.577 ± 0.006	0.152 ± 0.024	<u>0.002</u> ± 0.001
	Solar	0.041 ± 0.011	0.628 ± 0.053	0.578 ± 0.039	0.353 ± 0.014	0.437 ± 0.020	<u>0.058</u> ± 0.005
	Wind	<u>0.025</u> ± 0.017	0.213 ± 0.017	0.170 ± 0.040	0.170 ± 0.004	0.131 ± 0.014	0.018 ± 0.007
	Traffic	0.083 ± 0.034	0.567 ± 0.057	0.538 ± 0.078	0.218 ± 0.007	0.213 ± 0.016	<u>0.089</u> ± 0.001
	Taxi	0.095 ± 0.023	0.275 ± 0.054	0.319 ± 0.032	0.139 ± 0.007	0.128 ± 0.004	<u>0.109</u> ± 0.014
	Pedestrian	0.044 ± 0.020	0.090 ± 0.030	0.112 ± 0.019	0.065 ± 0.002	0.067 ± 0.007	<u>0.058</u> ± 0.002
	Air	0.011 ± 0.003	0.120 ± 0.045	0.211 ± 0.041	0.089 ± 0.016	<u>0.028</u> ± 0.002	0.041 ± 0.008
	Temperature	0.219 ± 0.022	0.926 ± 0.042	0.809 ± 0.081	1.002 ± 0.014	0.323 ± 0.008	<u>0.259</u> ± 0.043
	Rain	0.057 ± 0.039	0.329 ± 0.285	0.111 ± 0.109	0.292 ± 0.019	<u>0.074</u> ± 0.007	0.080 ± 0.004
	NN5	0.164 ± 0.010	0.874 ± 0.088	0.632 ± 0.074	0.821 ± 0.061	0.327 ± 0.012	<u>0.243</u> ± 0.041
K-L	Fred-MD	0.002 ± 0.001	0.043 ± 0.021	0.133 ± 0.102	0.059 ± 0.008	0.008 ± 0.002	<u>0.005</u> ± 0.002
	Exchange	0.151 ± 0.024	0.530 ± 0.154	0.475 ± 0.116	0.543 ± 0.149	0.342 ± 0.050	<u>0.233</u> ± 0.107
	Electricity	0.012 ± 0.016	0.488 ± 0.175	0.407 ± 0.079	0.734 ± 0.023	0.280 ± 0.051	<u>0.027</u> ± 0.015
	Solar	0.016 ± 0.005	0.612 ± 0.447	<u>0.120</u> ± 0.041	0.260 ± 0.016	0.865 ± 0.108	<u>0.234</u> ± 0.062
	Wind	<u>0.152</u> ± 0.034	1.924 ± 1.233	0.107 ± 0.016	0.484 ± 0.015	0.483 ± 0.066	0.183 ± 0.047
	Traffic	0.009 ± 0.003	1.305 ± 0.320	1.409 ± 0.251	0.211 ± 0.014	0.178 ± 0.026	<u>0.016</u> ± 0.003
	Taxi	0.011 ± 0.004	0.650 ± 0.180	0.950 ± 0.197	0.110 ± 0.020	0.110 ± 0.026	<u>0.038</u> ± 0.010
	Pedestrian	0.014 ± 0.010	0.417 ± 0.181	0.411 ± 0.096	0.065 ± 0.005	0.405 ± 0.051	<u>0.039</u> ± 0.008
	Air	0.027 ± 0.016	0.348 ± 0.093	0.578 ± 0.049	0.164 ± 0.012	<u>0.054</u> ± 0.012	0.093 ± 0.025
	Temperature	0.171 ± 0.073	8.892 ± 2.681	3.174 ± 2.685	2.183 ± 0.110	0.735 ± 0.066	<u>0.379</u> ± 0.110
	Rain	0.013 ± 0.012	0.506 ± 0.174	0.432 ± 0.099	0.160 ± 0.022	<u>0.047</u> ± 0.018	0.065 ± 0.018
	NN5	0.054 ± 0.014	4.928 ± 4.112	1.386 ± 0.520	1.337 ± 0.220	1.063 ± 0.274	<u>0.220</u> ± 0.151
	Fred-MD	0.203 ± 0.035	0.512 ± 0.290	0.380 ± 0.070	<u>0.346</u> ± 0.041	0.831 ± 0.077	1.118 ± 0.276
	Exchange	1.866 ± 0.132	8.861 ± 3.397	7.201 ± 4.380	10.404 ± 1.434	5.052 ± 1.385	8.475 ± 3.056

Table 1: Maximum mean discrepancy (MMD) and K-L divergence (K-L) of in-domain generation for sequence length 168. Best results are highlighted in bold face and second best results are underlined.

The MMD and K-L results of our model and baselines for the generation sequence length 168 are shown in Table 2.

Table 2 shows that our model demonstrates robust zero-shot time series synthesis capability, obtaining the best general MMD and K-L scores compared to baselines. The fine-tuned models do not show consistent performance improvement against the unconditional model, indicating that in the low data regime, fine-tuning is not effective for fitting the data distribution. On the contrary, our model is able to infer the unseen domain distribution using domain prompts without additional tuning, and the performance improves as the number of few-shot samples grows, showing strong unseen domain generation capability.

4.2.3 Ablation Study. In this section, we evaluate the sensitivity of PAM design by modifying the number of prototypes, and evaluate its effectiveness by removing it and using $\phi(x)$ as domain prompt instead, which is shown with “-PAM”. We also test the unconditional version of model, labeled with “-Prompt”. Table 3 shows the results on MMD and K-L scores, and MDD scores are shown in the appendix. Our model’s performance remains consistent when the number is large enough. Removing PAM and removing conditioning mechanism both lead to great drop in MMD score while maintaining stable K-L scores, indicating that diffusion model backbone is strong enough to capture marginal distribution while PAM and domain prompts are essential in capturing sequence-wise time series distribution.

5 Conclusion

In this paper, we propose a multi-domain time series diffusion model with domain prompts, named **TimeDP**. TimeDP learns prototypes to represent time series basis, serving as “word” with time series semantics. A prototype assignment module is applied to extract the domain specific prototype weights, for learning domain prompts as generation condition. During sampling, we construct “domain prompt” with few-shot samples from the target domain and use the domain prompts as condition to generate time series samples. Experiments demonstrate that our method outperforms baselines to provide the state-of-the-art in-domain generation quality and strong unseen domain generation capability.

References

- [1] Andrea Coletta, Sriram Gopalakrishnan, Daniel Borrajo, and Svitlana Vyetrenko. 2023. On the Constrained Time-Series Generation Problem. In *Advances in Neural Information Processing Systems*. 61048–61059.
- [2] Andrea Coletta, Matteo Prata, Michele Conti, Emanuele Mercanti, Novella Bartolini, Aymeric Moulin, Svitlana Vyetrenko, and Tucker Balch. 2021. Towards realistic market simulations: a generative adversarial networks approach. In *Proceedings of the ACM International Conference on AI in Finance*. 1–9.
- [3] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. 2024. A decoder-only foundation model for time-series forecasting. In *Proceedings of the International Conference on Machine Learning*. 10148–10167.
- [4] Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. 2021. TimeVAE: A Variational Auto-Encoder for Multivariate Time Series Generation. *arXiv preprint arXiv:2111.08095* (2021).
- [5] Xinyao Fan, Yueying Wu, Chang Xu, Yuhao Huang, Weiqing Liu, and Jiang Bian. 2024. MG-TSD: Multi-Granularity Time Series Diffusion Models with Guided Learning Process. In *International Conference on Learning Representations*.

		Stock			Web				
		# Shots	3	10	100	3	10	100	
Maximum Mean Discrepancy	Unconditional	TimeGAN	0.517 \pm 0.030	0.823 \pm 0.042	0.944 \pm 0.075	0.197 \pm 0.027	0.089 \pm 0.045	0.126 \pm 0.104	
		GT-GAN	0.466 \pm 0.050	0.790 \pm 0.048	0.692 \pm 0.074	0.109 \pm 0.027	0.117 \pm 0.012	0.112 \pm 0.014	
		TimeVAE	0.351 \pm 0.009	0.870 \pm 0.022	0.898 \pm 0.025	0.203 \pm 0.004	0.146 \pm 0.006	0.099 \pm 0.006	
		TimeVQVAE	0.105 \pm 0.005	0.111 \pm 0.001	0.096 \pm 0.004	0.078 \pm 0.007	0.063 \pm 0.009	0.062 \pm 0.006	
	Fine-tuned	TimeGAN	0.124 \pm 0.081	0.144 \pm 0.034	0.136 \pm 0.093	0.157 \pm 0.034	0.182 \pm 0.054	0.055 \pm 0.018	
		GT-GAN	0.139 \pm 0.076	0.161 \pm 0.113	0.261 \pm 0.132	0.101 \pm 0.053	0.191 \pm 0.037	0.230 \pm 0.068	
		TimeVAE	0.348 \pm 0.007	0.790 \pm 0.028	0.761 \pm 0.032	0.199 \pm 0.013	0.160 \pm 0.045	0.118 \pm 0.019	
		TimeVQVAE	0.205 \pm 0.089	0.176 \pm 0.095	0.127 \pm 0.050	0.278 \pm 0.226	0.198 \pm 0.077	0.085 \pm 0.060	
	Prompted	DP-Diff	0.048 \pm 0.014	0.050 \pm 0.012	0.014 \pm 0.005	0.068 \pm 0.035	0.065 \pm 0.014	0.021 \pm 0.004	
	K-L	Unconditional	TimeGAN	5.077 \pm 1.314	5.213 \pm 1.211	3.717 \pm 1.563	1.485 \pm 0.766	0.432 \pm 0.375	0.485 \pm 0.340
			GT-GAN	2.010 \pm 0.302	2.086 \pm 0.294	1.541 \pm 0.120	0.645 \pm 0.123	1.351 \pm 0.291	1.626 \pm 0.274
			TimeVAE	1.872 \pm 0.094	2.175 \pm 0.106	1.666 \pm 0.075	0.345 \pm 0.021	0.056 \pm 0.007	0.104 \pm 0.007
TimeVQVAE			0.840 \pm 0.117	0.458 \pm 0.084	0.692 \pm 0.125	0.175 \pm 0.035	0.448 \pm 0.034	0.579 \pm 0.045	
Fine-tuned		TimeGAN	3.164 \pm 0.361	1.844 \pm 0.684	1.304 \pm 0.450	1.739 \pm 0.662	0.606 \pm 0.119	0.357 \pm 0.140	
		GT-GAN	1.183 \pm 0.266	0.482 \pm 0.253	0.681 \pm 0.292	0.588 \pm 0.313	1.230 \pm 0.248	1.640 \pm 0.553	
		TimeVAE	1.331 \pm 0.269	1.855 \pm 0.170	1.344 \pm 0.080	0.309 \pm 0.050	0.058 \pm 0.010	0.482 \pm 0.097	
		TimeVQVAE	1.139 \pm 0.416	0.614 \pm 0.303	0.864 \pm 0.688	1.176 \pm 1.338	0.540 \pm 0.318	0.383 \pm 0.295	
Prompted		DP-Diff	1.613 \pm 0.065	0.721 \pm 0.041	0.090 \pm 0.037	0.245 \pm 0.027	0.254 \pm 0.030	0.091 \pm 0.010	
MDD		Unconditional	TimeGAN	0.837 \pm 0.014	0.803 \pm 0.007	0.790 \pm 0.015	0.084 \pm 0.008	0.054 \pm 0.013	0.062 \pm 0.022
			GT-GAN	0.795 \pm 0.007	0.804 \pm 0.002	0.756 \pm 0.006	0.050 \pm 0.005	0.064 \pm 0.002	0.059 \pm 0.002
			TimeVAE	0.814 \pm 0.003	0.858 \pm 0.003	0.859 \pm 0.003	0.076 \pm 0.002	0.057 \pm 0.002	0.045 \pm 0.001
	TimeVQVAE		0.711 \pm 0.005	0.702 \pm 0.003	0.706 \pm 0.006	0.047 \pm 0.005	0.072 \pm 0.003	0.073 \pm 0.002	
	Fine-tuned	TimeGAN	0.750 \pm 0.031	0.770 \pm 0.046	0.760 \pm 0.038	0.066 \pm 0.022	0.079 \pm 0.003	0.062 \pm 0.012	
		GT-GAN	0.717 \pm 0.026	0.711 \pm 0.032	0.731 \pm 0.018	0.048 \pm 0.010	0.072 \pm 0.010	0.080 \pm 0.010	
		TimeVAE	0.801 \pm 0.012	0.847 \pm 0.005	0.840 \pm 0.006	0.077 \pm 0.002	0.056 \pm 0.004	0.071 \pm 0.006	
		TimeVQVAE	0.741 \pm 0.022	0.718 \pm 0.029	0.708 \pm 0.024	0.076 \pm 0.046	0.087 \pm 0.015	0.062 \pm 0.025	
	Prompted	DP-Diff	0.720 \pm 0.012	0.690 \pm 0.009	0.653 \pm 0.002	0.042 \pm 0.006	0.051 \pm 0.005	0.038 \pm 0.002	

Table 2: Maximum mean discrepancy (MMD), K-L divergence (K-L) and Marginal distribution difference (MDD) in unseen domain settings of Stock and Web datasets for generation sequence length 168. Best results are highlighted in bold face.

- [6] Shanghua Gao, Teddy Koker, Owen Queen, Thomas Hartvigsen, Theodoros Tsiligkaridis, and Marinka Zitnik. 2024. UniTS: Building a Unified Time Series Model. *arXiv preprint arXiv:2403.00131* (2024).
- [7] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*. 6840–6851.
- [8] Jonathan Ho and Tim Salimans. 2021. Classifier-Free Diffusion Guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*.
- [9] Yu-Hao Huang, Chang Xu, Yang Liu, Weiqing Liu, Wu-Jun Li, and Jiang Bian. 2024. Controllable Financial Market Generation with Diffusion Guided Meta Agent. *arXiv preprint arXiv:2408.12991* (2024).
- [10] Jinsung Jeon, JEONGHAK KIM, Haryong Song, Seunghyeon Cho, and Noseong Park. 2022. Time-series Generative Adversarial Networks. In *Advances in Neural Information Processing Systems*. 36999–37010.
- [11] Yushan Jiang, Zijie Pan, Xikun Zhang, Sahil Garg, Anderson Schneider, Yuriy Nevmyvaka, and Dongjin Song. 2024. Empowering Time Series Analysis with Large Language Models: A Survey. In *Proceedings of the International Joint Conference on Artificial Intelligence*. 8095–8103.
- [12] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. 2024. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. In *International Conference on Learning Representations*.
- [13] Marcel Kollovich, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Yuyang Wang. 2023. Predict, refine, synthesize: self-guiding diffusion models for probabilistic time series forecasting. In *Advances in Neural Information Processing Systems*. 28341–28364.
- [14] Marcel Kollovich, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Yuyang (Bernie) Wang. 2023. Predict, Refine, Synthesize: Self-Guiding Diffusion Models for Probabilistic Time Series Forecasting. In *Advances in Neural Information Processing Systems*. 28341–28364.
- [15] Maurice Kraus, Felix Divo, David Steinmann, Devendra Singh Dhami, and Kristian Kersting. 2024. United We Pretrain, Divided We Fail! Representation Learning for Time Series by Pretraining on 75 Datasets at Once. *arXiv preprint arXiv:2402.15404* (2024).
- [16] Daesoo Lee, Sara Malacarne, and Erlend Aune. 2023. Vector Quantized Time Series Generation with a Bidirectional Prior Model. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*. 7665–7693.
- [17] Hongming Li, Shujian Yu, and José C. Principe. 2023. Causal Recurrent Variational Autoencoder for Medical Time Series Generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Brian Williams, Yiling Chen, and Jennifer Neville (Eds.). 8562–8570.
- [18] Zelin Ni, Hang Yu, Shizhan Liu, Jianguo Li, and Weiyao Lin. 2023. Basisformer: Attention-based time series forecasting with learnable and interpretable basis. *Advances in Neural Information Processing Systems* (2023), 71222–71241.
- [19] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *International Conference on Learning Representations*.
- [20] Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. 2021. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *Proceedings of the International Conference on Machine Learning*. 8857–8868.
- [21] Lifeng Shen and James T. Kwok. 2023. Non-autoregressive conditional diffusion models for time series prediction. In *Proceedings of the International Conference on Machine Learning*. 31016–31029.
- [22] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In

	# Prototypes	TimeDP					- PAM	- Prompt
		4	8	16	32	64		
Maximum Mean Discrepancy	Electricity	0.004 \pm 0.002	0.001 \pm 0.001	<u>0.001\pm0.001</u>	0.001 \pm 0.001	0.001\pm0.001	0.004 \pm 0.002	0.016 \pm 0.015
	Solar	0.061 \pm 0.025	0.041 \pm 0.018	0.041 \pm 0.011	<u>0.037\pm0.007</u>	0.037\pm0.006	0.545 \pm 0.030	0.520 \pm 0.029
	Wind	0.029 \pm 0.017	0.027 \pm 0.017	<u>0.025\pm0.017</u>	0.025 \pm 0.017	0.025\pm0.016	0.134 \pm 0.028	0.074 \pm 0.047
	Traffic	0.101 \pm 0.025	<u>0.077\pm0.018</u>	0.083 \pm 0.034	0.081 \pm 0.023	0.067\pm0.005	0.270 \pm 0.015	0.288 \pm 0.020
	Taxi	0.111 \pm 0.021	0.095 \pm 0.013	<u>0.095\pm0.023</u>	0.095 \pm 0.020	0.093\pm0.016	0.149 \pm 0.026	0.140 \pm 0.014
	Pedestrian	0.049 \pm 0.021	0.040\pm0.010	0.044 \pm 0.020	0.044 \pm 0.016	<u>0.041\pm0.016</u>	0.052 \pm 0.009	0.057 \pm 0.007
	Air	0.016 \pm 0.003	0.011 \pm 0.001	<u>0.011\pm0.003</u>	0.012 \pm 0.003	0.011\pm0.002	0.061 \pm 0.031	0.017 \pm 0.007
	Temperature	0.221 \pm 0.034	<u>0.209\pm0.014</u>	0.219 \pm 0.022	0.221 \pm 0.023	0.224 \pm 0.027	0.354 \pm 0.011	0.179\pm0.043
	Rain	0.056 \pm 0.020	<u>0.047\pm0.020</u>	0.057 \pm 0.039	0.051 \pm 0.030	0.057 \pm 0.057	0.056 \pm 0.046	0.043\pm0.018
	NN5	0.198 \pm 0.059	<u>0.157\pm0.003</u>	0.164 \pm 0.010	0.155\pm0.008	0.158 \pm 0.005	0.348 \pm 0.021	0.257 \pm 0.021
	Fred-MD	0.001\pm0.000	<u>0.002\pm0.001</u>	0.002 \pm 0.001	0.002 \pm 0.001	0.002 \pm 0.001	0.338 \pm 0.114	0.005 \pm 0.005
	Exchange	<u>0.146\pm0.031</u>	0.146 \pm 0.022	0.151 \pm 0.024	0.150 \pm 0.024	0.152 \pm 0.023	1.059 \pm 0.008	0.117\pm0.028
	Average	0.083	0.071	0.074	0.073	<u>0.072</u>	0.281	0.143
K-L	Electricity	0.003 \pm 0.002	<u>0.011\pm0.014</u>	0.012 \pm 0.016	0.014 \pm 0.019	0.012 \pm 0.018	0.011\pm0.009	0.013 \pm 0.013
	Solar	0.028 \pm 0.021	0.016 \pm 0.007	0.016 \pm 0.005	0.013\pm0.002	<u>0.014\pm0.005</u>	0.019 \pm 0.004	0.042 \pm 0.036
	Wind	<u>0.151\pm0.027</u>	0.153 \pm 0.044	0.152 \pm 0.034	0.156 \pm 0.037	0.155 \pm 0.032	0.072\pm0.010	0.156 \pm 0.073
	Traffic	0.024 \pm 0.030	0.007\pm0.003	<u>0.009\pm0.003</u>	0.012 \pm 0.009	0.011 \pm 0.004	0.014 \pm 0.013	0.047 \pm 0.027
	Taxi	0.051 \pm 0.033	<u>0.008\pm0.003</u>	0.011 \pm 0.004	0.012 \pm 0.010	0.009 \pm 0.005	0.003\pm0.001	0.032 \pm 0.022
	Pedestrian	0.031 \pm 0.028	0.010\pm0.008	0.014 \pm 0.010	0.014 \pm 0.009	<u>0.012\pm0.008</u>	0.021 \pm 0.010	0.041 \pm 0.047
	Air	0.047 \pm 0.027	<u>0.024\pm0.008</u>	0.027 \pm 0.016	0.027 \pm 0.010	0.025 \pm 0.012	0.020\pm0.005	0.036 \pm 0.017
	Temperature	0.304 \pm 0.079	0.176 \pm 0.061	0.171 \pm 0.073	0.179 \pm 0.087	<u>0.166\pm0.074</u>	0.132\pm0.009	0.168 \pm 0.096
	Rain	0.033 \pm 0.009	<u>0.008\pm0.004</u>	0.013 \pm 0.012	0.008\pm0.005	0.013 \pm 0.008	0.009 \pm 0.008	0.009 \pm 0.004
	NN5	0.217 \pm 0.223	0.050 \pm 0.021	0.054 \pm 0.014	0.045\pm0.007	<u>0.046\pm0.006</u>	0.049 \pm 0.016	0.073 \pm 0.031
	Fred-MD	0.188\pm0.012	<u>0.190\pm0.013</u>	0.203 \pm 0.035	0.216 \pm 0.052	0.201 \pm 0.024	0.196 \pm 0.008	0.196 \pm 0.016
	Exchange	1.863 \pm 0.223	1.828 \pm 0.131	1.866 \pm 0.132	2.313 \pm 1.007	1.875 \pm 0.235	<u>1.623\pm0.146</u>	1.506\pm0.086
	Average	0.245	0.207	0.212	0.251	0.212	0.181	<u>0.193</u>

Table 3: Maximum mean discrepancy (MMD) and K-L divergence (K-L) results for ablation study on generation sequence length 168. Best results are highlighted in bold face. Second best results are underlined.

- Proceedings of the International Conference on Machine Learning*. 2256–2265.
- [23] Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. 2024. Unified Training of Universal Time Series Forecasting Transformers. In *Proceedings of the International Conference on Machine Learning*. 53140–53164.
- [24] Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, Jiang Bian, Shirui Pan, and Qingsong Wen. 2024. A Survey on Diffusion Models for Time Series and Spatio-Temporal Data. *arXiv preprint arXiv:2404.18886* (2024).
- [25] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. 2019. Time-series Generative Adversarial Networks. In *Advances in Neural Information Processing Systems*. 5509–5519.
- [26] Xinyu Yuan and Yan Qiao. 2024. Diffusion-TS: Interpretable Diffusion for General Time Series Generation. In *International Conference on Learning Representations*.

A Technical Appendix

A.1 Model Details

A.1.1 Architecture. The denoising network of the diffusion model used in this paper is built with U-Net structure, with 4 up/down sampling blocks. Each up/down sampling block consists of 2 residual blocks and 1 cross-attention block, where each residual block contains two 1-D convolution layers and each cross-attention block uses 1-D convolution as input/output projection layers and uses 8 attention heads. There is a middle block containing two residual blocks and one attention blocks, placed after the down sampling blocks and before the up sampling blocks. There are also additional input and output blocks, each with one 1-D convolution layer. We use SiLU as the non-linear activation function in this module.

For the prototype assignment module, we use two 1-D convolution layers as feature extractor. Then there are another two 1-D convolution layers with residual connection, followed by a linear projection layer, as assignment generator producing final prototype assignment. All non-linear activation functions in this module are chosen to be ReLU.

A.1.2 Description on Datasets. In this section, we provide descriptions into the datasets used in model training this paper:

- **Electricity.** This dataset represents the hourly electricity consumption of 321 clients from 2012 to 2014 in kilowatt(kW). It was originally extracted from UCI.
- **Solar.** This dataset contains 137 time series representing the solar power production recorded every 1 hour in the state of Alabama in 2006.
- **Wind.** This dataset contains a single very long daily time series representing the wind power production in MW recorded per every 4 seconds starting from 01/08/2019. It was downloaded from the Australian Energy Market Operator (AEMO) online platform.
- **Traffic.** This dataset contains 15 months worth of daily data (440 daily records) that describes the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time.
- **Taxi.** This dataset contains spatio-temporal traffic time series of New York taxi rides taken at 1214 locations every 30 minutes in the months of January 2015 and January 2016.
- **Pedestrian.** This dataset contains hourly pedestrian counts captured from 66 sensors in Melbourne city starting from May 2009. The original dataset is regularly updated when a new set of observations become available. The dataset uploaded here contains pedestrian counts up to 2020-04-30.
- **Air Quality.** This dataset was used in the KDD Cup 2018 forecasting competition. It contains long hourly time series representing the air quality levels in 59 stations in 2 cities: Beijing (35 stations) and London (24 stations) from 01/01/2017 to 31/03/2018. The air quality level is represented in multiple measurements such as PM2.5, PM10, NO2, CO, O3 and SO2. The dataset uploaded here contains 270 hourly time series which have been categorized using city, station name and air quality measurement. The original dataset contains missing values. The leading missing values of a given series were replaced by zeros and the remaining missing

values were replaced by carrying forward the corresponding last observations (LOCF method).

- **Temperature.** This dataset contains 32072 daily time series showing the temperature observations and rain forecasts, gathered by the Australian Bureau of Meteorology for 422 weather stations across Australia, between 02/05/2015 and 26/04/2017. The original dataset contains missing values and they have been simply replaced by zeros. We extracted the mean temperature column here.
- **Rain.** This data set comes from same source as Temperature dataset, extracted the rain column.
- **NN5.** This dataset was used in the NN5 forecasting competition. It contains 111 time series from the banking domain. The goal is predicting the daily cash withdrawals from ATMs in UK. The original dataset contains missing values. A missing value on a particular day is replaced by the median across all the same days of the week along the whole series.
- **Fred-MD.** This dataset contains 107 monthly time series showing a set of macro-economic indicators from the Federal Reserve Bank. It was extracted from the FRED-MD database. The series are differentiated and log-transformed as suggested in the literature.
- **Exchange.** This dataset contains daily exchange rate between 8 currencies.

More over, we also describe the dataset used in unseen data generation experiment:

- **Stock.** This dataset contains daily stock price of symbol GOOG, listed in NASDAQ.
- **Web.** This dataset was used in the Kaggle Wikipedia Web Traffic forecasting competition. It contains 145063 daily time series representing the number of hits or web traffic for a set of Wikipedia pages from 2015-07-01 to 2017-09-10. The original dataset contains missing values. They have been simply replaced by zeros.

A.2 Marginal Distribution Distance Score for Generation Quality Experiment

The additional marginal distribution distance score results for the generation quality experiment are shown in Table 4. Table 4 shows that TimeDP gets overall best result on the MDD metrics.

A.3 Marginal Distribution Distance Score for Ablation Study

The additional marginal distribution distance score results for the ablation study experiment are shown in Table 8. From Table 8 we can observe that the marginal distribution distance score performance of TimeDP is generally stable.

A.4 Visualization Analysis on Domain Prompts

In this section, we analyze the correlation among domain prompts through visualization.

Figure 2 provides a heatmap for domain prompts, each column showing the weights for each prototype.

		TimeDP	TimeGAN	GT-GAN	TimeVAE	TimeVQVAE	TimeVQVAE-C
Marginal Distribution Distance	Electricity	0.005 ± 0.002	0.075 ± 0.035	0.047 ± 0.008	0.098 ± 0.003	0.067 ± 0.004	<u>0.005</u> ± 0.001
	Solar	<u>56.414</u> ± 21.890	70.334 ± 11.928	83.855 ± 3.100	16.721 ± 0.041	57.401 ± 0.041	59.043 ± 0.440
	Wind	0.084 ± 0.009	0.226 ± 0.061	0.138 ± 0.015	0.201 ± 0.004	0.159 ± 0.011	0.085 ± 0.008
	Traffic	0.049 ± 0.004	0.149 ± 0.018	0.153 ± 0.001	0.110 ± 0.001	0.119 ± 0.005	<u>0.053</u> ± 0.002
	Taxi	0.081 ± 0.008	0.104 ± 0.010	0.109 ± 0.004	0.094 ± 0.001	0.096 ± 0.004	<u>0.084</u> ± 0.004
	Pedestrian	0.071 ± 0.012	0.096 ± 0.024	0.097 ± 0.006	0.086 ± 0.002	0.143 ± 0.007	<u>0.079</u> ± 0.002
	Air	0.042 ± 0.002	0.139 ± 0.025	0.171 ± 0.011	0.085 ± 0.001	0.092 ± 0.008	<u>0.058</u> ± 0.006
	Temperature	0.142 ± 0.010	0.189 ± 0.009	0.208 ± 0.012	0.259 ± 0.004	0.191 ± 0.004	<u>0.156</u> ± 0.012
	Rain	0.067 ± 0.015	0.228 ± 0.090	0.142 ± 0.020	0.228 ± 0.008	0.177 ± 0.016	<u>0.104</u> ± 0.006
	NN5	0.140 ± 0.005	0.295 ± 0.021	0.240 ± 0.014	0.304 ± 0.010	0.220 ± 0.012	<u>0.175</u> ± 0.011
	Fred-MD	0.021 ± 0.002	0.079 ± 0.016	0.098 ± 0.024	0.104 ± 0.009	0.126 ± 0.009	<u>0.076</u> ± 0.025
	Exchange	0.358 ± 0.010	<u>0.351</u> ± 0.098	0.345 ± 0.018	0.442 ± 0.020	0.515 ± 0.021	0.486 ± 0.040

Table 4: Marginal Distribution Distance (MDD) results of in-domain generation for sequence length 168. Best results are highlighted in bold face and second best results are underlined.

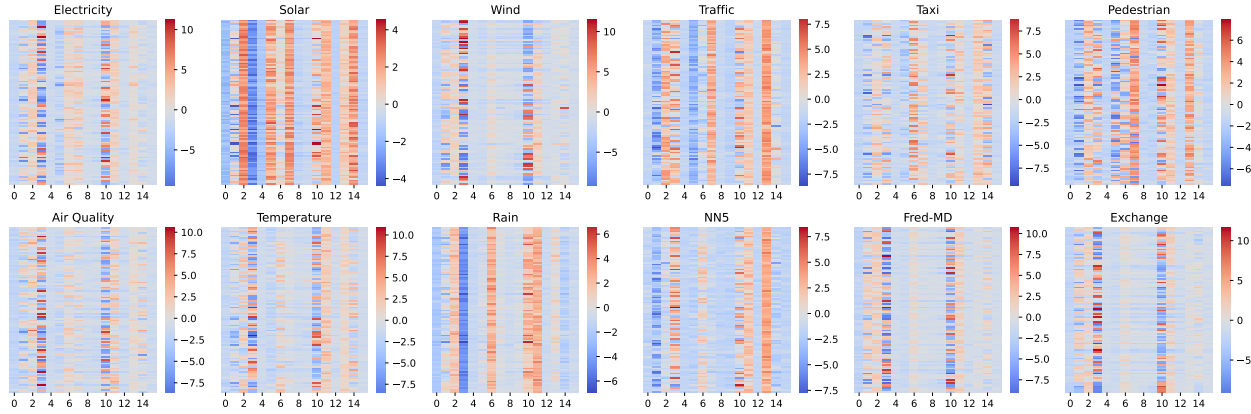


Figure 2: Heatmap of Domain Prompts.

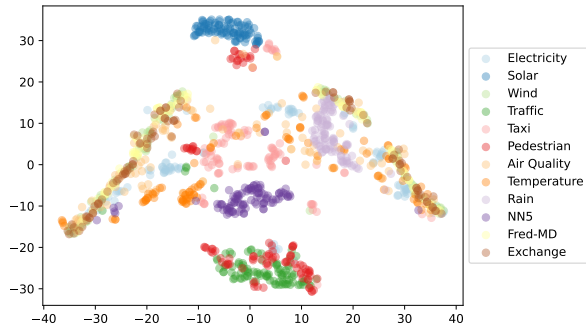


Figure 3: T-SNE visualization of domain prompts. Domain prompt generated for different dataset are marked with the different color.

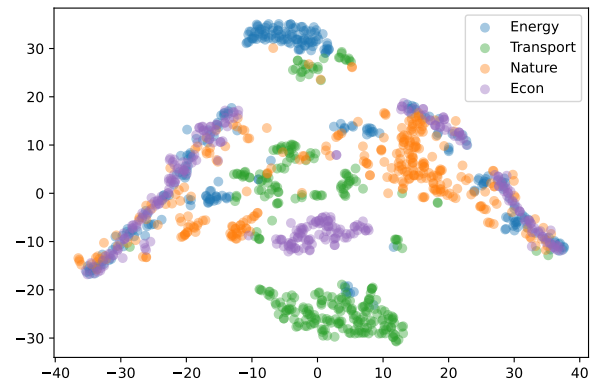


Figure 4: T-SNE visualization of domain prompts. Datasets in the same domain are marked with the same color.

Figure 3 and Figure 4 provide t-SNE analysis on the similarity of prompts among domain. We can observe that the model has learned to differentiate among domain with domain prompts.

		TimeDP	TimeGAN	GT-GAN	TimeVAE	TimeVQVAE	TimeVQVAE-C
Maximum Mean Discrepancy	Electricity	0.001 ± 0.001	0.270 ± 0.095	0.189 ± 0.085	0.980 ± 0.016	0.142 ± 0.017	<u>0.002</u> ± 0.001
	Solar	0.023 ± 0.016	0.513 ± 0.031	0.517 ± 0.027	0.321 ± 0.005	0.411 ± 0.004	<u>0.036</u> ± 0.011
	Wind	0.042 ± 0.019	0.348 ± 0.019	0.308 ± 0.015	0.419 ± 0.006	0.129 ± 0.008	0.034 ± 0.011
	Traffic	0.009 ± 0.005	0.466 ± 0.071	0.522 ± 0.072	0.146 ± 0.005	0.189 ± 0.009	<u>0.014</u> ± 0.004
	Taxi	0.032 ± 0.005	0.156 ± 0.046	0.243 ± 0.030	0.049 ± 0.004	<u>0.036</u> ± 0.003	0.036 ± 0.004
	Pedestrian	0.056 ± 0.011	0.085 ± 0.022	0.121 ± 0.011	0.111 ± 0.003	0.087 ± 0.005	<u>0.063</u> ± 0.004
	Air	0.008 ± 0.004	0.146 ± 0.022	0.206 ± 0.007	0.192 ± 0.010	<u>0.010</u> ± 0.002	0.017 ± 0.005
	Temperature	0.086 ± 0.015	0.801 ± 0.048	0.801 ± 0.015	1.003 ± 0.012	0.215 ± 0.018	<u>0.111</u> ± 0.019
	Rain	0.014 ± 0.010	0.182 ± 0.096	0.085 ± 0.052	0.737 ± 0.048	<u>0.036</u> ± 0.006	0.037 ± 0.004
	NN5	0.072 ± 0.015	0.704 ± 0.028	0.507 ± 0.078	0.708 ± 0.037	0.162 ± 0.011	0.086 ± 0.004
K-L	Fred-MD	0.002 ± 0.001	1.113 ± 0.048	0.822 ± 0.139	0.187 ± 0.014	0.003 ± 0.001	<u>0.003</u> ± 0.000
	Exchange	0.127 ± 0.066	0.504 ± 0.024	0.484 ± 0.074	0.662 ± 0.018	0.276 ± 0.026	<u>0.155</u> ± 0.039
	Electricity	0.016 ± 0.015	0.333 ± 0.134	0.423 ± 0.178	1.245 ± 0.018	0.302 ± 0.022	<u>0.023</u> ± 0.007
	Solar	0.017 ± 0.009	0.410 ± 0.042	<u>0.102</u> ± 0.039	0.458 ± 0.016	0.878 ± 0.031	0.120 ± 0.036
	Wind	0.316 ± 0.057	2.378 ± 0.989	<u>0.317</u> ± 0.175	1.105 ± 0.019	0.566 ± 0.030	0.374 ± 0.053
	Traffic	<u>0.014</u> ± 0.005	1.061 ± 0.177	1.211 ± 0.286	0.446 ± 0.116	0.219 ± 0.016	0.013 ± 0.005
	Taxi	0.006 ± 0.002	0.547 ± 0.150	0.872 ± 0.148	0.263 ± 0.022	0.121 ± 0.015	<u>0.015</u> ± 0.007
	Pedestrian	0.016 ± 0.010	0.369 ± 0.092	0.408 ± 0.105	0.211 ± 0.015	0.431 ± 0.017	<u>0.023</u> ± 0.008
	Air	0.029 ± 0.007	0.477 ± 0.154	0.923 ± 0.119	0.683 ± 0.067	0.108 ± 0.015	<u>0.057</u> ± 0.006
	Temperature	0.240 ± 0.029	7.672 ± 2.287	1.419 ± 0.461	1.288 ± 0.145	0.764 ± 0.076	<u>0.357</u> ± 0.055
Marginal Distribution Distance	Rain	0.037 ± 0.019	0.356 ± 0.089	0.468 ± 0.158	0.987 ± 0.073	<u>0.077</u> ± 0.020	0.089 ± 0.037
	NN5	<u>0.074</u> ± 0.028	2.273 ± 0.231	1.154 ± 0.288	1.434 ± 0.117	0.904 ± 0.175	0.063 ± 0.010
	Fred-MD	0.515 ± 0.174	3.356 ± 0.639	2.161 ± 0.483	0.701 ± 0.091	<u>0.533</u> ± 0.067	0.574 ± 0.168
	Exchange	7.270 ± 2.080	5.482 ± 2.108	3.897 ± 2.172	8.682 ± 1.334	<u>4.201</u> ± 0.945	4.561 ± 0.212
	Electricity	<u>0.004</u> ± 0.002	0.059 ± 0.011	0.052 ± 0.014	0.114 ± 0.001	0.063 ± 0.002	0.004 ± 0.001
	Solar	70.117 ± 16.651	64.855 ± 3.985	76.076 ± 4.681	46.304 ± 0.010	<u>50.554</u> ± 0.146	53.143 ± 0.574
	Wind	0.115 ± 0.017	0.195 ± 0.016	0.138 ± 0.013	0.291 ± 0.003	0.188 ± 0.007	<u>0.133</u> ± 0.010
	Traffic	<u>0.017</u> ± 0.005	0.112 ± 0.008	0.124 ± 0.003	0.079 ± 0.002	0.080 ± 0.002	0.016 ± 0.002
	Taxi	<u>0.016</u> ± 0.002	0.036 ± 0.005	0.049 ± 0.002	0.028 ± 0.001	0.029 ± 0.002	0.014 ± 0.001
	Pedestrian	<u>0.049</u> ± 0.006	0.060 ± 0.004	0.075 ± 0.002	0.069 ± 0.002	0.092 ± 0.002	0.046 ± 0.001
	Air	0.028 ± 0.005	0.076 ± 0.016	0.112 ± 0.003	0.100 ± 0.003	0.053 ± 0.003	<u>0.035</u> ± 0.003
	Temperature	0.102 ± 0.007	0.145 ± 0.004	0.175 ± 0.009	0.180 ± 0.002	0.157 ± 0.006	<u>0.109</u> ± 0.007
	Rain	0.007 ± 0.001	0.063 ± 0.017	0.046 ± 0.011	0.122 ± 0.002	0.052 ± 0.001	<u>0.012</u> ± 0.003
	NN5	<u>0.073</u> ± 0.005	0.248 ± 0.005	0.185 ± 0.019	0.243 ± 0.008	0.111 ± 0.004	0.068 ± 0.003
	Fred-MD	0.010 ± 0.002	0.107 ± 0.002	0.098 ± 0.004	0.053 ± 0.007	0.038 ± 0.003	<u>0.011</u> ± 0.001
	Exchange	0.348 ± 0.011	<u>0.321</u> ± 0.040	0.313 ± 0.017	0.417 ± 0.009	0.466 ± 0.009	0.441 ± 0.021

Table 5: Maximum mean discrepancy (MMD) and K-L divergence (K-L) and Marginal Distribution Distance (MDD) results of in-domain generation for sequence length 24. Best results are highlighted in bold face and second best results are underlined.

A.5 Visualization on Unseen Domain Generated Time Series

Figure 6 and Figure 5 show the few-shot prompting results.

		TimeDP	TimeGAN	GT-GAN	TimeVAE	TimeVQVAE	TimeVQVAE-C
Maximum Mean Discrepancy	Electricity	0.002 ± 0.001	0.401 ± 0.219	0.317 ± 0.311	0.648 ± 0.012	0.153 ± 0.023	<u>0.002</u> ± 0.000
	Solar	0.019 ± 0.004	0.584 ± 0.056	0.634 ± 0.119	0.369 ± 0.012	0.439 ± 0.021	<u>0.035</u> ± 0.008
	Wind	<u>0.029</u> ± 0.015	0.195 ± 0.030	0.217 ± 0.134	0.179 ± 0.002	0.129 ± 0.016	0.013 ± 0.007
	Traffic	0.018 ± 0.005	0.456 ± 0.100	0.513 ± 0.073	0.162 ± 0.005	0.189 ± 0.014	<u>0.027</u> ± 0.002
	Taxi	0.055 ± 0.007	0.197 ± 0.057	0.328 ± 0.071	<u>0.060</u> ± 0.002	0.089 ± 0.005	0.061 ± 0.002
	Pedestrian	0.040 ± 0.010	0.068 ± 0.025	0.159 ± 0.149	0.064 ± 0.001	0.066 ± 0.011	<u>0.049</u> ± 0.005
	Air	0.010 ± 0.006	0.103 ± 0.046	0.244 ± 0.160	0.085 ± 0.011	<u>0.022</u> ± 0.003	0.036 ± 0.007
	Temperature	0.178 ± 0.047	0.827 ± 0.092	0.859 ± 0.109	0.958 ± 0.012	0.413 ± 0.034	<u>0.205</u> ± 0.033
	Rain	0.044 ± 0.027	0.285 ± 0.199	0.166 ± 0.203	0.327 ± 0.020	<u>0.068</u> ± 0.008	0.072 ± 0.001
	NN5	0.234 ± 0.047	0.888 ± 0.091	0.706 ± 0.116	0.875 ± 0.053	0.321 ± 0.008	0.241 ± 0.011
K-L	Fred-MD	<u>0.005</u> ± 0.002	0.055 ± 0.022	0.150 ± 0.184	0.046 ± 0.013	0.004 ± 0.001	0.006 ± 0.001
	Exchange	0.343 ± 0.210	0.449 ± 0.018	0.567 ± 0.141	0.592 ± 0.124	<u>0.337</u> ± 0.056	0.224 ± 0.028
	Electricity	0.015 ± 0.017	0.552 ± 0.241	0.447 ± 0.147	0.790 ± 0.018	0.245 ± 0.039	<u>0.031</u> ± 0.005
	Solar	0.010 ± 0.002	0.922 ± 0.953	<u>0.117</u> ± 0.047	0.288 ± 0.015	0.812 ± 0.077	0.151 ± 0.114
	Wind	<u>0.182</u> ± 0.029	2.288 ± 1.135	0.164 ± 0.071	0.527 ± 0.009	0.510 ± 0.112	0.284 ± 0.115
	Traffic	0.010 ± 0.006	1.241 ± 0.757	1.198 ± 0.340	0.205 ± 0.008	0.192 ± 0.016	<u>0.014</u> ± 0.004
	Taxi	0.010 ± 0.007	0.550 ± 0.333	0.801 ± 0.270	0.121 ± 0.003	0.102 ± 0.017	<u>0.027</u> ± 0.007
	Pedestrian	0.012 ± 0.005	0.458 ± 0.368	0.327 ± 0.148	0.074 ± 0.004	0.373 ± 0.027	<u>0.034</u> ± 0.006
	Air	0.025 ± 0.011	0.317 ± 0.102	0.510 ± 0.147	0.168 ± 0.012	<u>0.061</u> ± 0.012	0.096 ± 0.019
	Temperature	0.392 ± 0.097	11.298 ± 1.212	5.830 ± 4.888	1.737 ± 0.068	1.135 ± 0.110	<u>0.567</u> ± 0.050
Marginal Distribution Distance	Rain	0.022 ± 0.009	0.549 ± 0.257	0.481 ± 0.146	0.237 ± 0.025	<u>0.055</u> ± 0.021	0.091 ± 0.010
	NN5	0.069 ± 0.033	5.829 ± 6.120	2.709 ± 2.080	1.537 ± 0.224	1.038 ± 0.168	<u>0.103</u> ± 0.072
	Fred-MD	1.183 ± 0.280	0.628 ± 0.335	<u>0.486</u> ± 0.108	0.441 ± 0.063	0.813 ± 0.093	1.240 ± 0.232
	Exchange	14.336 ± 2.652	6.445 ± 2.210	10.170 ± 6.381	9.937 ± 4.322	3.740 ± 1.261	<u>4.408</u> ± 0.700
	Electricity	<u>0.005</u> ± 0.002	0.070 ± 0.020	0.046 ± 0.011	0.082 ± 0.002	0.058 ± 0.002	0.005 ± 0.001
	Solar	77.545 ± 18.384	60.819 ± 9.001	77.198 ± 5.522	47.123 ± 0.114	<u>50.619</u> ± 0.099	53.220 ± 0.832
	Wind	<u>0.089</u> ± 0.012	0.207 ± 0.054	0.135 ± 0.035	0.208 ± 0.002	0.159 ± 0.009	0.069 ± 0.014
	Traffic	0.024 ± 0.003	0.109 ± 0.018	0.122 ± 0.009	0.083 ± 0.001	0.089 ± 0.002	<u>0.028</u> ± 0.002
	Taxi	<u>0.043</u> ± 0.004	0.072 ± 0.008	0.092 ± 0.010	0.064 ± 0.001	0.070 ± 0.003	0.040 ± 0.002
	Pedestrian	0.058 ± 0.007	0.074 ± 0.018	0.088 ± 0.025	0.073 ± 0.001	0.116 ± 0.003	<u>0.058</u> ± 0.003
	Air	0.031 ± 0.007	0.091 ± 0.021	0.145 ± 0.022	0.070 ± 0.002	0.066 ± 0.004	<u>0.056</u> ± 0.007
	Temperature	0.135 ± 0.016	0.154 ± 0.004	0.185 ± 0.007	0.196 ± 0.001	0.198 ± 0.006	<u>0.143</u> ± 0.010
	Rain	0.039 ± 0.006	0.192 ± 0.059	0.113 ± 0.011	0.196 ± 0.003	0.134 ± 0.009	<u>0.047</u> ± 0.012
	NN5	<u>0.175</u> ± 0.016	0.270 ± 0.014	0.225 ± 0.017	0.268 ± 0.006	0.196 ± 0.007	0.158 ± 0.004
	Fred-MD	0.023 ± 0.005	0.063 ± 0.009	0.076 ± 0.023	0.055 ± 0.012	0.083 ± 0.004	<u>0.045</u> ± 0.017
	Exchange	0.402 ± 0.083	<u>0.342</u> ± 0.044	0.334 ± 0.024	0.425 ± 0.027	0.503 ± 0.016	0.426 ± 0.033

Table 6: Maximum mean discrepancy (MMD) and K-L divergence (K-L) and Marginal Distribution Distance (MDD) results of in-domain generation for sequence length 96. Best results are highlighted in bold face and second best results are underlined.

		TimeDP	TimeGAN	GT-GAN	TimeVAE	TimeVQVAE	TimeVQVAE-C
Maximum Mean Discrepancy	Electricity	0.001 ± 0.002	0.320 ± 0.197	0.347 ± 0.524	0.504 ± 0.012	0.135 ± 0.018	<u>0.003± 0.000</u>
	Solar	0.074 ± 0.012	0.646 ± 0.032	0.735 ± 0.045	0.400 ± 0.015	0.493 ± 0.024	<u>0.081± 0.006</u>
	Wind	<u>0.024± 0.003</u>	0.208 ± 0.095	0.221 ± 0.010	0.145 ± 0.005	0.139 ± 0.015	0.014 ± 0.007
	Traffic	0.077 ± 0.009	0.518 ± 0.028	0.704 ± 0.101	0.235 ± 0.011	0.233 ± 0.017	<u>0.091± 0.005</u>
	Taxi	0.189 ± 0.034	0.329 ± 0.011	0.390 ± 0.078	0.166 ± 0.002	0.202 ± 0.004	<u>0.189± 0.007</u>
	Pedestrian	0.035 ± 0.008	0.140 ± 0.067	0.158 ± 0.019	0.061 ± 0.003	0.074 ± 0.007	<u>0.067± 0.008</u>
	Air	0.018 ± 0.005	0.213 ± 0.037	0.261 ± 0.062	0.075 ± 0.005	<u>0.038± 0.003</u>	0.053 ± 0.007
	Temperature	0.185 ± 0.021	0.914 ± 0.031	0.934 ± 0.051	0.976 ± 0.022	0.318 ± 0.008	<u>0.197± 0.014</u>
	Rain	0.096 ± 0.086	0.151 ± 0.108	0.198 ± 0.295	0.135 ± 0.018	0.066 ± 0.002	<u>0.081± 0.005</u>
	NN5	0.191 ± 0.016	0.663 ± 0.027	0.951 ± 0.069	0.806 ± 0.086	0.333 ± 0.009	<u>0.260± 0.049</u>
K-L	Fred-MD	0.006 ± 0.004	0.156 ± 0.097	0.063 ± 0.006	0.052 ± 0.020	0.010 ± 0.006	<u>0.007± 0.001</u>
	Exchange	0.683 ± 0.217	0.724 ± 0.130	0.553 ± 0.238	0.598 ± 0.166	<u>0.318± 0.077</u>	0.275 ± 0.198
	Electricity	0.016 ± 0.016	0.396 ± 0.108	1.169 ± 2.029	0.719 ± 0.020	0.257 ± 0.023	<u>0.042± 0.033</u>
	Solar	0.017 ± 0.006	<u>0.179± 0.062</u>	2.150 ± 2.179	0.249 ± 0.012	0.870 ± 0.110	<u>0.201± 0.017</u>
	Wind	<u>0.158± 0.032</u>	0.169 ± 0.064	4.264 ± 2.790	0.373 ± 0.019	0.495 ± 0.043	0.121 ± 0.037
	Traffic	0.008 ± 0.004	1.443 ± 0.354	2.992 ± 0.592	0.241 ± 0.012	0.210 ± 0.034	<u>0.023± 0.006</u>
	Taxi	0.076 ± 0.034	0.937 ± 0.151	1.461 ± 0.367	0.160 ± 0.015	0.230 ± 0.026	<u>0.094± 0.007</u>
	Pedestrian	0.009 ± 0.003	0.452 ± 0.131	1.544 ± 0.908	<u>0.044± 0.007</u>	0.407 ± 0.069	0.053 ± 0.017
	Air	0.017 ± 0.010	1.031 ± 0.145	1.481 ± 0.553	0.180 ± 0.015	0.121 ± 0.028	<u>0.079± 0.016</u>
	Temperature	0.037 ± 0.016	2.479 ± 1.867	9.139 ± 2.247	1.726 ± 0.122	1.119 ± 0.179	<u>0.224± 0.123</u>
Marginal Distribution Distance	Rain	0.011 ± 0.002	0.388 ± 0.112	0.271 ± 0.046	0.028 ± 0.005	<u>0.025± 0.013</u>	0.028 ± 0.005
	NN5	0.075 ± 0.034	1.418 ± 0.122	8.449 ± 5.473	1.233 ± 0.256	0.836 ± 0.250	<u>0.180± 0.181</u>
	Fred-MD	0.815 ± 0.630	<u>0.309± 0.145</u>	0.568 ± 0.321	0.212 ± 0.047	0.586 ± 0.161	1.099 ± 0.020
	Exchange	18.426 ± 4.493	14.116 ± 3.798	13.875 ± 2.965	10.670 ± 5.335	<u>7.755± 2.566</u>	7.651 ± 3.841
	Electricity	0.006 ± 0.004	0.048 ± 0.009	0.061 ± 0.058	0.099 ± 0.002	0.069 ± 0.003	<u>0.008± 0.003</u>
	Solar	<u>49.170± 21.121</u>	82.860 ± 4.163	74.403 ± 18.119	13.755 ± 0.077	54.065 ± 0.183	55.878 ± 0.249
	Wind	<u>0.091± 0.007</u>	0.146 ± 0.028	0.251 ± 0.060	0.172 ± 0.004	0.168 ± 0.009	0.070 ± 0.009
	Traffic	0.066 ± 0.002	0.191 ± 0.005	0.206 ± 0.031	0.134 ± 0.001	0.158 ± 0.005	<u>0.072± 0.005</u>
	Taxi	0.680 ± 0.021	1.073 ± 0.162	0.986 ± 0.126	0.764 ± 0.003	0.750 ± 0.003	<u>0.707± 0.001</u>
	Pedestrian	0.074 ± 0.006	0.121 ± 0.027	0.153 ± 0.070	<u>0.086± 0.002</u>	0.161 ± 0.007	0.098 ± 0.006
	Air	0.049 ± 0.004	0.157 ± 0.002	0.166 ± 0.039	0.078 ± 0.002	0.099 ± 0.005	<u>0.056± 0.005</u>
	Temperature	0.153 ± 0.010	0.230 ± 0.007	0.218 ± 0.013	0.341 ± 0.005	0.204 ± 0.009	<u>0.155± 0.006</u>
	Rain	0.100 ± 0.007	0.144 ± 0.011	0.238 ± 0.184	0.210 ± 0.014	0.222 ± 0.009	<u>0.138± 0.014</u>
	NN5	<u>1.079± 0.016</u>	1.128 ± 0.017	1.199 ± 0.024	1.209 ± 0.021	1.109 ± 0.006	1.074 ± 0.012
	Fred-MD	0.031 ± 0.007	0.106 ± 0.038	0.085 ± 0.006	0.093 ± 0.035	0.128 ± 0.016	<u>0.060± 0.025</u>
	Exchange	0.510 ± 0.131	0.337 ± 0.023	<u>0.383± 0.091</u>	0.443 ± 0.027	0.537 ± 0.036	<u>0.452± 0.048</u>

Table 7: Maximum mean discrepancy (MMD) and K-L divergence (K-L) and Marginal Distribution Distance (MDD) results of in-domain generation for sequence length 336. Best results are highlighted in bold face and second best results are underlined.

	# Prototypes	TimeDP					- PAM	- Prompt
		4	8	16	32	64		
Marginal Distribution Distance	Electricity	0.006 \pm 0.001	0.005 \pm 0.001	0.005 \pm 0.002	0.005 \pm 0.002	<u>0.005</u> \pm 0.002	0.008 \pm 0.004	0.014 \pm 0.008
	Solar	46.294 \pm 19.573	45.134 \pm 21.804	56.414 \pm 21.890	<u>38.374</u> \pm 22.157	64.258 \pm 29.679	47.291 \pm 16.147	32.032 \pm 6.336
	Wind	0.089 \pm 0.014	0.084 \pm 0.011	<u>0.084</u> \pm 0.009	0.084 \pm 0.012	0.084 \pm 0.009	0.087 \pm 0.006	0.090 \pm 0.023
	Traffic	0.072 \pm 0.021	0.051 \pm 0.003	0.049 \pm 0.004	<u>0.050</u> \pm 0.002	0.050 \pm 0.001	0.131 \pm 0.001	0.133 \pm 0.004
	Taxi	0.087 \pm 0.004	0.082 \pm 0.005	0.081 \pm 0.008	<u>0.082</u> \pm 0.008	<u>0.081</u> \pm 0.007	0.089 \pm 0.002	0.089 \pm 0.004
	Pedestrian	0.080 \pm 0.010	0.072 \pm 0.006	0.071 \pm 0.012	0.073 \pm 0.009	<u>0.072</u> \pm 0.010	0.076 \pm 0.003	0.076 \pm 0.011
	Air	0.050 \pm 0.005	0.043 \pm 0.003	0.042 \pm 0.002	0.042 \pm 0.003	<u>0.042</u> \pm 0.003	0.046 \pm 0.002	0.048 \pm 0.007
	Temperature	0.153 \pm 0.013	<u>0.138</u> \pm 0.007	0.142 \pm 0.010	0.143 \pm 0.011	0.144 \pm 0.011	0.163 \pm 0.002	0.128 \pm 0.021
	Fred-MD	0.024 \pm 0.004	<u>0.020</u> \pm 0.003	0.021 \pm 0.002	<u>0.020</u> \pm 0.002	0.020 \pm 0.001	0.070 \pm 0.002	0.029 \pm 0.010
	Exchange	0.358 \pm 0.032	0.358 \pm 0.008	0.358 \pm 0.010	<u>0.357</u> \pm 0.006	0.360 \pm 0.008	0.493 \pm 0.003	0.327 \pm 0.006
Average		3.955	3.849	4.790	<u>3.286</u>	5.443	4.060	2.769

Table 8: Marginal Distribution Distance (MDD) results for ablation study on generation sequence length 168. Best results are highlighted in bold face. Second best results are underlined.

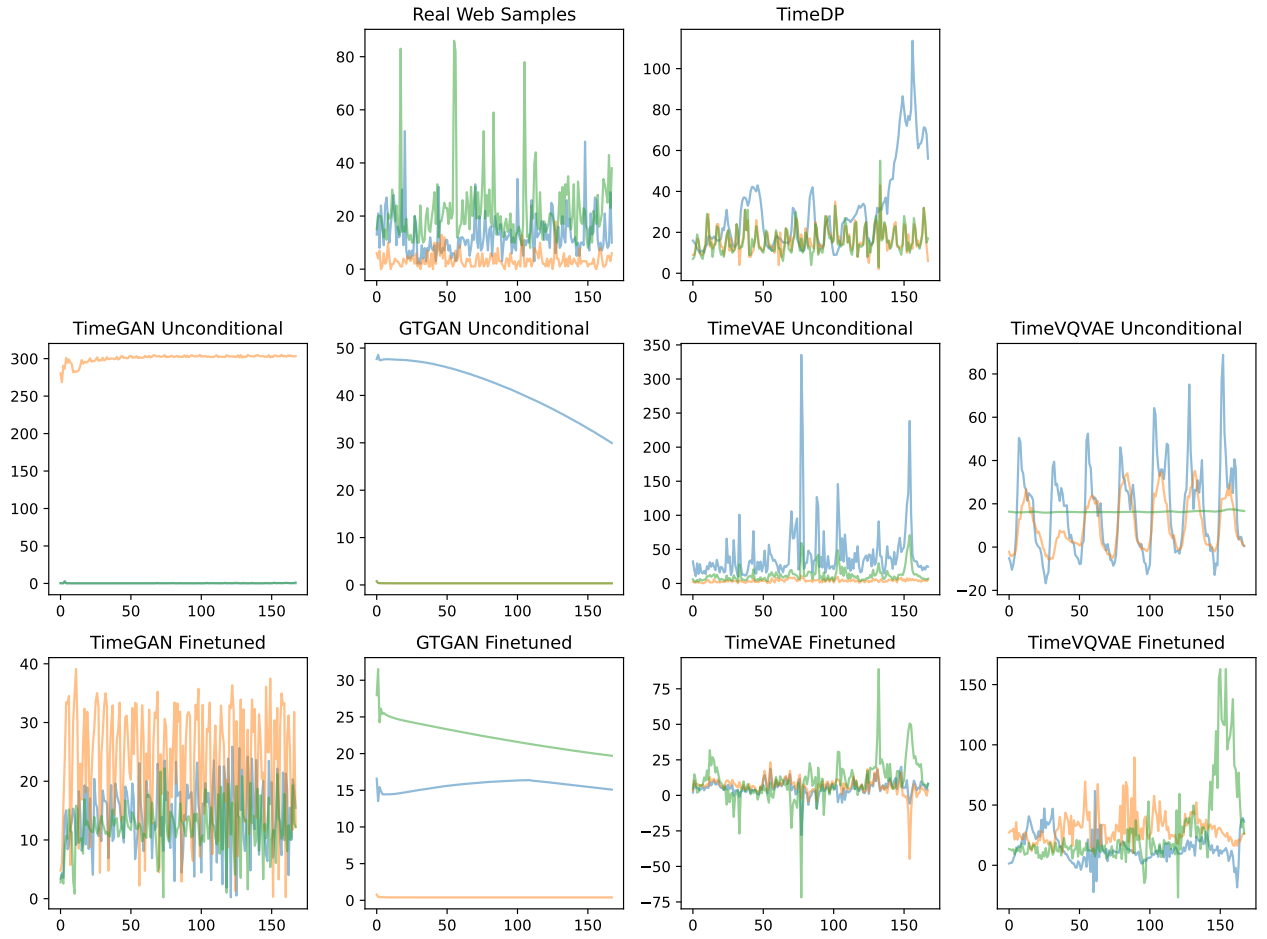


Figure 5: 10-shot generation results of Web dataset, compared with real Web dataset samples.

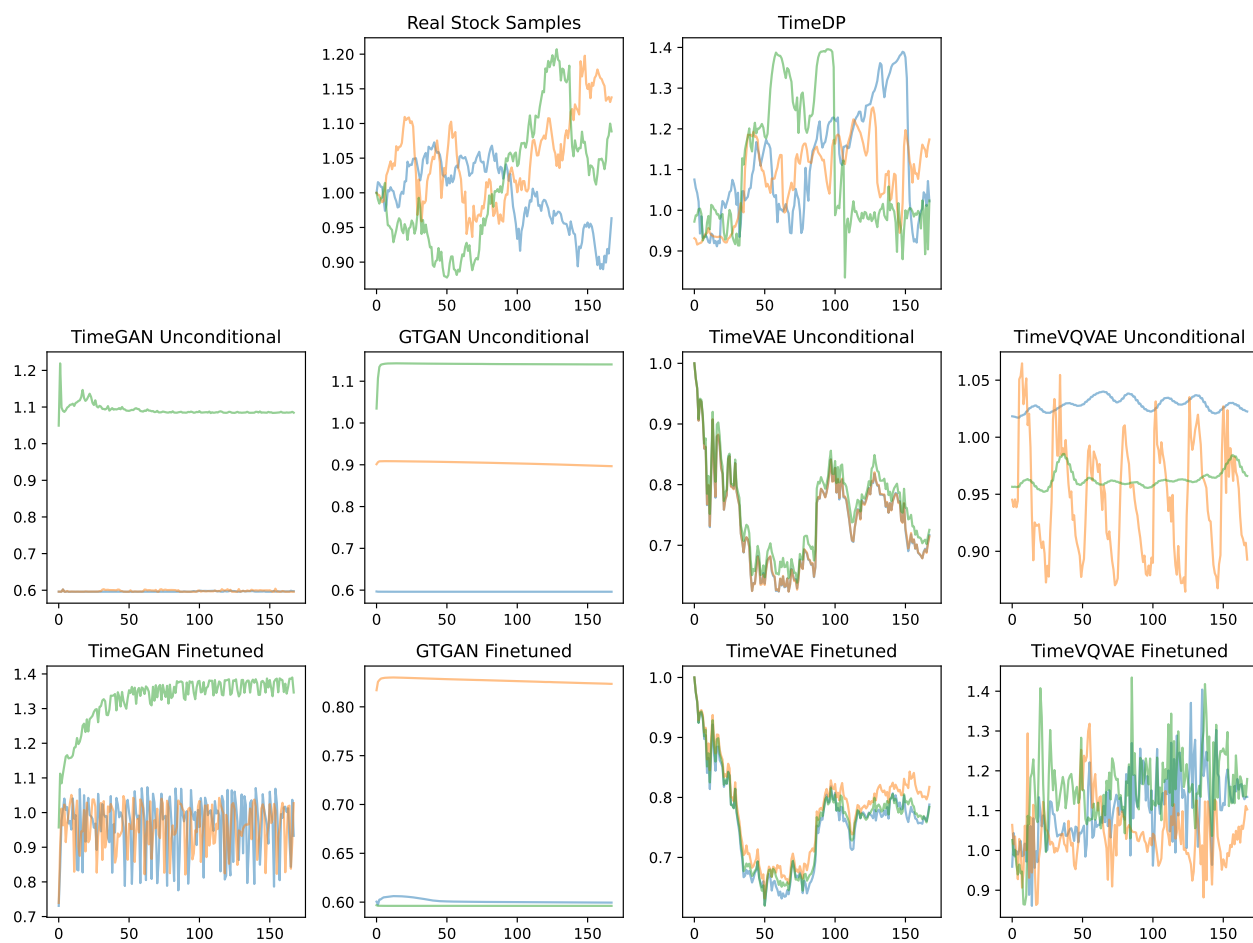


Figure 6: 10-shot generation results of Stock dataset, compared with real Stock dataset samples.