Cumulative Difference Learning VAE for Time-Series with Temporally Correlated Inflow-Outflow

Tianchun Li[1] **Chengxiang Wu**[2], **Pengyi Shi**[3], **Xiaoqian Wang**[1]
[1]School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA
[2]Department of Computer Science, Purdue University, West Lafayette, Indiana, USA
[3]School of Business, Purdue University, West Lafayette, Indiana, USA
{li2657, wu1522, joywang, shi178}@purdue.edu

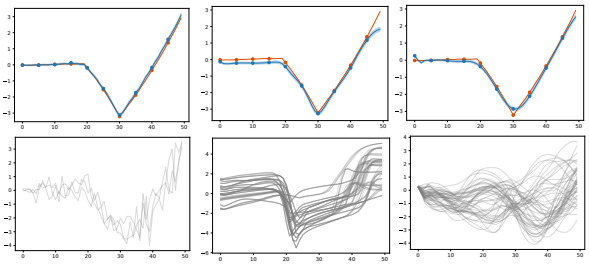
Abstract

Time-series generation has crucial practical significance for decision-making under uncertainty. Existing methods have various limitations like accumulating errors over time, significantly impacting downstream tasks. We develop a novel generation method, DT-VAE, that incorporates generalizable domain knowledge, is mathematically justified, and significantly outperforms existing methods by mitigating error accumulation through a cumulative difference learning mechanism. We evaluate the performance of DT-VAE on several downstream tasks using both semi-synthetic and real time-series datasets, including benchmark datasets and our newly curated COVID-19 hospitalization datasets. The COVID-19 datasets enrich existing resources for time-series analysis. Additionally, we introduce Diverse Trend Preserving (DTP), a time-series clustering-based evaluation for direct and interpretable assessments of generated samples, serving as a valuable tool for evaluating time-series generative models.

Introduction

Time-series generation provides augmented datasets that can be used for testing and validating models, improving the accuracy of predictive models, avoiding model overfitting based on limited observations, as well as for generating scenarios for decision-making (e.g., demand scenarios for evaluating staffing level) (van Dyk and Meng 2001; Forestier et al. 2017; Wen et al. 2021). These benefits are highly valuable within various domains. For example, generated synthetic data has been shown to benefit significantly healthcare and medicine fields (Abbasimehr, Paki, and Bahrini 2022; Chellasamy and Nagarathinam 2022; Er, Yang, and Zhao 2023; Kiyasseh et al. 2020), by addressing the concerns about data privacy and restricted data access limit, and thus boosting model comparisons and the development of new models.

This work is motivated by the burgeoning need to synthesize hospital census (number of hospitalized patients in different units/hospitals/regions) since the COVID-19 pandemic outbreak. Learning the census *distribution* and generating future census sample paths, rather than just point estimates, is crucial for downstream prediction tasks and decision-making, e.g., bed capacity planning, nurse staffing, and medical equipment allocation via stochastic optimization.



(a) DT-VAE (Ours) (b) PSA-GAN (c) TimeGAN

Figure 1: The mean of the generated series (dark blue) and 0.95 confidence interval (light blue range) align with the true mean (orange). However, significantly more samples (grey curves in (b) and (c)) from the PSA-GAN and TimeGAN deviate from the original trend with chaotic shapes. For better visualization, up to 50 of the 500 generated samples that notably deviated are highlighted.

Existing methods fall short in addressing the challenges that arise in meeting this pressing demand, prompting the initiation of this research. Yet, these challenges are not unique to the healthcare domain; they are pervasive in many other applications such as crime prediction, financial trading, and traffic flow (Calatayud, Jornet, and Mateu 2023; Gontis, Ruseckas, and Kononovičius 2010; Tahmasbi and Hashemi 2014). The two primary challenges are (i) the errors tend to accumulate over time due to generation in an autoregressive/recursive manner (Yoon, Jarrett, and Van der Schaar 2019; Jeha et al. 2021) and (ii) general-purpose (non-domain specific) generative methods often ignore problem structure and lack theoretical justification or interpretability.

Figure 1 illustrates the prevailing issue (i) via comparison using synthetic data. At each time step, the synthetic data is drawn from a Gaussian distribution with time-varying means, which produces a distinct "V-shaped" trend. While the means of generated data from all models align closely with the ground truth mean, a substantial portion of samples generated by general-purpose methods deviates from the trend in synthetic data, as shown in plots (b) and (c), primarily due to early-stage errors accumulating over time, resulting in disruptions to the trends. Notably, even though samples generated by PSA-GAN shown in (b) exhibit relatively more

resilience to errors compared to samples from TimeGAN in (c), issue (i) still significantly harms its performance: the early-stage generation errors cause the large misalignment with the crucial trend where the minimum occurs and lead to chaotic patterns in later stages. More details are in the "Gaussian Synthetic Data" section in the appendix. Issue (ii) imposes another layer of challenge to general-purpose generative methods. The time series observations are often driven by unobserved variables with complex interactions in various domains. For instance, the daily hospital census is driven by daily patient arrivals (i.e., *inflow*) and discharges (i.e., *outflow*). The arrivals and discharges are correlated, resulting in highly complex spatial-temporal correlations in the census, but their records are often absent. General-purpose methods either rely on step-wise learning (causing the issue (i)) or otherwise have to impose assumptions on temporal correlation structures in the time series observations, which might inadequately capture the complexities. Disregarding the interactions between the unobserved and observed variables may also risk misinterpreting the true drivers behind the time-series patterns and thereby affect decision-making.

In this work, our method, named *DT-VAE*, addresses the above issues by leveraging a mathematically justified, domain-aware time-series framework that adeptly integrates generalizable domain knowledge. Our analytical framework focuses on the type of time series driven by temporally (and potentially spatially) correlated and unobservable inflow and outflow variables. The common problem structure follows a relationship in which the observation of a new time step arises from the preceding time step, with inflow added and outflow subtracted.

To overcome the issue (i), we utilize the inflow/outflow structure to mathematically transform the original learning task to learn the cumulative difference between the observation on any time step t and the initial time step (the *DT* part), achieving success as demonstrated in Figure 1(a). Meanwhile, our framework captures the interplay between unobserved variables and time series observations through its generative structure, overcoming issue (ii). This framework allows us to mathematically derive the learning objective, which guides the design of the encoder and decoder networks of the *VAE* part. This not only provides theoretical justification and interpretability that contrast purely black-box models or heuristic designs, but also offers a versatile foundation for capturing complex correlation structures. We demonstrate this latter point by empirically comparing with direct methods such as ARIMA and Temporal-VAE (which applies VAE to learn the census directly with assumed correlation structure); see the "Experiment" section. Our framework is generally applicable across various domains, including healthcare (Littig and Isken 2007), traffic management (Lebacque 2005), and energy collection (Mugi and Chandramohan 2021). For other domains, our framework may lose some interpretability but remain applicable. We establish the broad applicability of our framework conceptually in the "Method" section and empirically validate it using five public benchmark datasets from various domains.

We introduce additional key contributions and summarize our contributions as follows:

- We propose a new **theoretical modeling framework** that incorporates generalizable domain knowledge by assuming the observed time series are driven by temporally linked inflow and outflow variables.
- We propose a novel **a VAE-based method** called *cumulative Difference Temporal VAE (DT-VAE)*, which merges VAE with cumulative learning, addressing prevalent issues with existing methods. DT-VAE also maintains high flexibility, demonstrated by connecting DT-VAE with a GAN (DT-VAE-GAN) to enhance time series generation.
- Our methods are evaluated on popular benchmark datasets and **two of our newly curated COVID-19 hospitalization datasets**. To the best of our knowledge, these two datasets are the first publicly accessible complied datasets focused on COVID-19 hospitalization.
- Beyond standard evaluations, we propose a **time-series clustering based evaluation metric** named *Diverse Trend Preserving (DTP) evaluation*. This provides not just numerical results but also a visually interpretable assessment of generated sample quality.

Related Work

Variational Autoencoder

This work is related to VAE (Kingma and Welling 2013). VAE learns a probabilistic mapping between input and designated latent spaces. Combining VAE with Generative Adversarial Networks (GAN) (Goodfellow et al. 2020) facilitates the VAE-GAN (Larsen et al. 2016), in which the decoder serves the role of the GAN's generator. Refer to the "VAE" section and "VAE-GAN Preliminary" in the appendix for more information.

Time-series Generation and Issues

Existing methods of time-series generation fall into three major categories: (i) Monte Carlo simulation that samples from prior defined distributions; (ii) deep-learning-based methods to learn the distribution from real-world data; and (iii) discrete-event simulation. Desai et al. (2021) summarized the pros and cons of approaches in (i)-(ii)

Simulations in (iii) need precise data calibration. For example, in the healthcare domain, while calibrating simulations requires patient flow attributes such as arrivals and departures, many datasets, such as COVID-19 hospitalization (Regenstrief Institute COVID-19 Dashboard 2023; WHO COVID-19 Dashboard 2023; CDC COVID-19 Tracking 2023), offer just the daily census. This unobservable issue is prevalent across different domains such as in social justice (Calatayud, Jornet, and Mateu 2023), traffic management (Dheeru and Karra Taniskidou 2017; Cuturi 2011), and energy conservation (Lai et al. 2018a), making (iii) impractical to realize. Contrasting (iii), direct methods like ARIMA or RNN overlook the dependence between the observed data and the in-/outflow variables, missing the spatial-temporal correlations.

The majority of methods in (ii) can be divided into two categories, GANs- or VAEs-based generative models. These frameworks improved solutions and enhanced flexibility for the aforementioned challenge. Early models, C-RNN-GAN

(Mogren 2016) and RCGAN (Esteban, Hyland, and Rätsch 2017), integrated RNNs and GANs for music and eICU data generation, respectively. TimeGAN (Yoon, Jarrett, and Van der Schaar 2019) aligns latent representations of real and generated data, considered as state-of-art. COT-GAN (Xu et al. 2020) employs optimal transport for adversarial training. PSA-GAN (Jeha et al. 2021) uses self-attention for long time-series generation and presents Context-FID, inspired by the FID score (Heusel et al. 2017). TsT-GAN (Srinivasan and Knottenbelt 2022) integrates transformers with GANs. VAE-based methods include Variational Recurrent AutoEncoder (Fabius and Van Amersfoort 2014) for classical songs, Stochastic WaveNet (Lai et al. 2018b) for adaptive prior distribution learning, and Stochastic TCN (Aksan and Hilliges 2019) merging ELBO with TCN (Bai, Kolter, and Koltun 2018). TimeVAE (Desai et al. 2021) uses interpretable components for time series generation, while CR-VAE (Li, Yu, and Principe 2023) used VAE to combine the causality and RNN. Other time-series generation methods involve generative models with Fourier Flows (Alaa, Chan, and van der Schaar 2020) or Contrastive Imitation (Jarrett, Bica, and van der Schaar 2021).

Method

Problem Definition

Consider a collection of random variables that form a time-series sequence $\{X_t, t = 0, 1, \ldots, T\}$ with the length of $T + 1$. We denote this sequence as $X_{0:T}$. The training dataset $\mathcal{D}$ consists of N observed sequences, denoted as $\mathcal{D} = \{x_{0:T}^{(1)}, \ldots, x_{0:T}^{(N)}\}$. For a given sequence, the observation at time t, x_t, is a vector in $\mathbb{R}^k$, where k is the feature space dimension. Going forward, N and k will be omitted unless explicitly mentioned. Our goal is to learn the joint distribution $p(X_{0:T})$ from the training data.

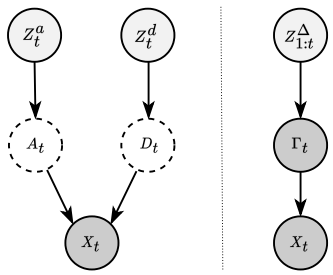
Problem Formulation

We adopt the generative modeling framework, where the dependency structure between the latent variables and observations is summarized in Figure 2 and are specified as follows. We assume that each random variable X_t is driven by the previous random variable X_{t-1}, an "inflow" variable A_t, and an "outflow" variable D_t. The relationship of X_t, X_{t-1}, A_t, D_t is described as

$$X_t | X_{t-1}, A_t, D_t = X_{t-1} + A_t - D_t + \epsilon, \quad t = 0, \ldots, T, \quad \epsilon \sim N(0, \tau), \tag{1}$$

where τ denotes the covariance matrix in the Gaussian distribution. Note that the noise may change over time as $\epsilon_t \sim N(0, \tau_t)$ with τ_t being time-varying. We continue with the fixed covariance matrix τ for now. Since the random variable X_t is dependent on other unobserved variables, we will refer to X_t as the "dependent variable".

The inflow and outflow variables A_t and D_t are further driven by some underlying "environmental factors" $\{Z_t\}$. We use the following dependence formulation, which is commonly used in deep latent Gaussian models (Rezende, Mohamed, and Wierstra 2014) and is a discretized version of stochastic differential equations (SDE) for variational inference in neural-SDE (Tzen and Raginsky 2019). The auto-regressive updates capture how the inflow and outflow variables change over time, similar to the evolution of a diffusion process described by SDE (Song et al. 2020).



(a) Original Generation View (b) Cum. Diff. Learning View

Figure 2: Geneative model structure.

$$A_0 = a_0; \quad D_0 = d_0; \tag{2}$$
$$A_t = A_{t-1} + b_a(A_{t-1}) + \sigma_a Z_t^a, t = 1, \ldots, T; \tag{3}$$
$$D_t = D_{t-1} + b_d(D_{t-1}) + \sigma_d Z_t^d, t = 1, \ldots, T.$$

Here, the sequences of latent variables $Z_1^a, \ldots, Z_T^a \sim^{iid} \mathcal{N}(0, I_k)$ and $Z_1^d, \ldots, Z_T^d \sim^{iid} \mathcal{N}(0, I_k)$ are all i.i.d. standard Gaussian vectors in $\mathbb{R}^k$. The drift function $b_a(\cdot)$ and $b_d(\cdot)$ and diffusion terms $\sigma_a Z_t^a$ and $\sigma_d Z_t^d$ drive the inflow and outflow processes and the temporal correlations. Additionally, it can integrate spatial correlations via the diffusion terms $\sigma_a Z_t^a$ and $\sigma_d Z_t^d$ depending on the application domains.

The formulation in (1)-(3) is generally applicable in various domains. We illustrate the applicability with a few examples. In healthcare, X_t corresponds to the daily patient census, i.e., the count of hospitalized patients in a specific hospital unit, county, or region on day t. The census changes with daily patient arrivals A_t and discharges D_t. In traffic management, X_t corresponds to the volume of traffic flow at time t, which is determined by incoming vehicles A_t and outgoing vehicles D_t. In the energy sector, X_t signifies the observed energy volume at time t. This amount is driven by the newly collected energy A_t and energy loss or consumption D_t. X_t for crime incidents can be interpreted similarly (Calatayud, Jornet, and Mateu 2023).

The parameters to be learned include those parametrizing the drift and the diffusion terms. This learning task poses challenges since the inflow and outflow variables are often absent in observed datasets and only values of X_t's are recorded. In other words, we cannot directly learn the parameters from inflow variable $\{A_t\}$ and outflow variable $\{D_t\}$ but need to learn from $\{X_t\}$ via the dependence structures given in Equations (1) and (2)-(3).

Cumulative Difference Learning

Beyond the unobserved inflow and outflow variables, a key challenge arises from the accumulation of errors. A common approach to learning the joint distribution of X_t within the generative modeling framework is through step-wise learning,

specifically, learning the conditional distribution $X_t|X_{0:t-1}$. However, this approach encounters a prominent challenge. For each time step $l < T$, a highly inaccurate estimation of the dependent variable X_l results in considerable deviations in the estimation of all subsequent values from $l+1$ to time T. This is due to the recursive nature of step-wise learning, where the calculation of the dependent variable X_t depends on X_{t-1}. In simpler terms, errors accumulate over time, and their effect on the dependent variable amplifies significantly for more distant future time points.

To overcome this issue, we develop a novel cumulative difference learning scheme, specified as follows. First, we define a new variable that captures the *cumulative* difference:

$$\Gamma_t = \sum_{i=1}^{t} \Delta_i = \sum_{i=1}^{t} A_i - \sum_{i=1}^{t} D_i. \tag{4}$$

Here, $\Delta_t = A_t - D_t$ represents the difference between inflow and outflow variables A_t and D_t (i.e., the changes in X_t's from X_{t-1}'s), and Γ_t represents the cumulative difference (i.e., the net changes in X_t's). In the training dataset, this cumulative difference can be observed by $\gamma_t = x_t - x_0$ (with measurement errors). The new relationship between X_0, X_t, and Γ_t can be described as follows:

$$X_t|(X_0, \Gamma_t) = X_0 + \Gamma_t + \epsilon'_t, \qquad \epsilon'_t \sim N(0, \tau'_t). \tag{5}$$

It is important to note that, different from (1), the noise ϵ'_t must change over time as the covariance matrix τ'_t is time-varying when learning the cumulative difference.

From (5), it appears that the currently observed variable X_t only depends on the current information of Γ_t for any given starting census X_0. However, this is not the case as the correlations are encapsulated in all latent variables up to time t, as they drive the cumulative difference Γ_t. This becomes evident from the following recursive equations for Γ_t:

$$\begin{aligned}
&\Gamma_0 = \Delta_0 = a_0 - d_0 \\
&\Gamma_t = \sum_{i=1}^{t} \Delta_i = \sum_{i=1}^{t-1} \Delta_i + \Delta_t = \Gamma_{t-1} + \Delta_t \\
&\quad = \Gamma_{t-1} + b_\Delta(\Delta_{t-1}) + \sigma_\Delta Z_t^\Delta, \quad t = 1, \ldots, T.
\end{aligned} \tag{6}$$

Note that (5)-(6) present an equivalent transformation of the original learning task, with the unknown parameters to be learned as the drift function $b_\Delta(\cdot)$ and diffusion matrix σ_Δ.

The relationship presented in (5)-(6) reveals the complex temporal correlations. Due to interactions between inflow and outflow variables described in (1)-(3), such complexity hinders any heuristic assumptions on correlation structures between X_t and X_{t-1}. However, by expanding (6), the correlation structure becomes clear. The difference variables $\Delta_{1:t-1}$ with the drift function $b_\Delta(\cdot)$ contributes to Γ_t, demonstrating that the temporal correlations X_t and X_{t-1} are also influenced by the sequence of difference variables $\Delta_{1:t-1}$.

Furthermore, using (5)-(6), reconstructing X_t from Γ_t prevents error accumulation, as Γ_t is a cumulative difference and only needs the initial value X_0 to reconstruct X_t, bypassing the recursive reconstruction in step-wise learning. By expanding (6), it becomes clear that Γ_t is driven by the latent variable sequence $Z_{1:t}^\Delta$. Therefore, we are essentially learning a mapping from the sequence of latent variables $Z_{1:t}^\Delta$ to the variable Γ_t. This means potential biases in the learned $\widehat{\Gamma}_{t-1}$ will not affect Γ_t learning since it is solely determined by latent variables. Meanwhile, as discussed, temporal correlations within sequences are not lost. This is because the dependence of Γ_t on $Z_{1:t}^\Delta$ naturally encodes temporal correlations between Γ_{t-1} and Γ_t.

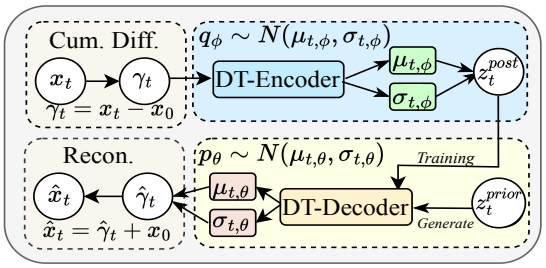


Figure 3: The architecture of DT-VAE with training and generation procedure. Encoder q_ϕ encodes input data to latent space, while decoder p_θ generates data from encoder samples during training and a prior distribution during generation.

Cumulative Difference Temporal VAE (DT-VAE)

We develop a VAE-based framework, *DT-VAE*, to learn (5) and (6).

ELBO. For any given $X_0 = x_0$, denote the observed sequence of cumulative difference variables $\Gamma_{1:T}$ as $\gamma_{1:T} = \{\gamma_1, \ldots, \gamma_T\}$. As from (6), Γ_t is driven by latent variables $Z_{1:t}^\Delta$ at each time step $t = \{1, \cdots, T\}$. Let $z_{1:T} = (z_1, \ldots, z_T)$ denote the sequence of prior variables from prior distribution $p(z_{1:T})$.

To overcome the intractability of the likelihood function $p_\theta(\gamma_{1:T})$, VAE optimizes the evidence lower bound (ELBO) as the surrogate objective:

$$\begin{aligned}
\log p_\theta(\gamma_{1:T}) &= \log \int p_\theta(\gamma_{1:T}, z_{1:T}) dz_{1:T} \\
&= \log \int p_\theta(\gamma_{1:T}, z_{1:T}) \frac{q_\phi(z_{1:T} \mid \gamma_{1:T})}{q_\phi(z_{1:T} \mid \gamma_{1:T})} dz_{1:T} \\
&\geq \mathbb{E}_{z_{1:T} \sim q_\phi} \left[\log \left(\frac{p_\theta(\gamma_{1:T}, z_{1:T})}{q_\phi(z_{1:T}|\gamma_{1:T})} \right) \right] \\
&= \mathbb{E}_{z_{1:T} \sim q_\phi} \left[\log \left(\frac{\prod_{t=1}^{T} p_\theta(\gamma_t|z_{1:t})) p(z_t|z_{1:t-1})}{\prod_{t=1}^{T} q_\phi(z_t|z_{1:t-1}, \gamma_{1:t})} \right) \right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{z_{1:t}} \log p_\theta(\gamma_t|z_{1:t}) \\
&\quad - \mathbb{E}_{z_{1:t-1}} D_{KL} \Big(q_\phi(z_t|z_{1:t-1}, \gamma_{1:t}) || N(0, I) \Big) \\
&= \mathcal{L}(\gamma_{1:T}).
\end{aligned} \tag{7}$$

The detailed derivation is given in the section A in the appendix. We describe the joint distribution $p_\theta(\gamma_{1:T}, z_{1:T})$ as the generative process for generating $\gamma_{1:T}$, and $q_\phi(z_{1:T}|\gamma_{1:T})$ denotes the variational distribution with parameter ϕ that

approximates the true posterior distribution. Next, we will describe the generative process $p_\theta(\gamma_{1:T}, z_{1:T})$ and variational distribution $q_\phi(z_{1:T}|\gamma_{1:T})$ separately.

Decoder design. For the generative process, the decoder of DT-VAE $f_\theta(\cdot)$, parameterized by θ, decodes latent variables $z_{1:t}$ to generate γ_t. In other words, the decoder $f_\theta(\cdot)$ learns the conditional distribution $p_\theta(\gamma_t|z_{1:t})$. Going forward, we will use $f_\theta(z_{1:t})$ and $p_\theta(\gamma_t|z_{1:t})$ interchangeably and use z^{prior} to denote samples from the prior distribution

A key step in deriving the ELBO in (7), particularly from line 3 to line 4, is via the decomposition for $p_\theta(\gamma_{1:T}, z_{1:T})$:

$$\begin{aligned} p_\theta(\gamma_{1:T}, z_{1:T}) &= p_\theta(\gamma_{1:T}|z_{1:T})p(z_{1:T}) \\ &= \Big(\prod_{t=1}^{T} p_\theta(\gamma_t|z_{1:t})\Big)p(z_{1:T}) \\ &= \prod_{t=1}^{T} p_\theta(\gamma_t|z_{1:t}) \prod_{t=1}^{T} p(z_t|z_{1:t-1}), \end{aligned} \tag{8}$$

where $p_\theta(\gamma_t|z_{1:t})$ denotes the approximation of the true conditional distribution $p(\gamma_t|z_{1:t})$ and $p(z_t|z_{1:t-1})$ denotes the conditional prior distribution for latent variables z_t.

From (8), we make an important assumption on the conditional distribution $p_\theta(\gamma_{1:T}|z_{1:T})$ and prior distribution $p(z_{1:T})$. As previously mentioned, for each γ_t, it solely depends on latent variables $z_{1:t}$ to avoid error accumulation. This essentially makes γ_t *conditionally independent* across different time steps given the latent variables $z_{1:t}^{prior}$. That is, for any two time steps $w \neq v \leq T$, the cumulative difference variable $(\gamma_w|z_{1:w}) \perp (\gamma_v|z_{1:v})$ are independent conditional on corresponding latent variables. This assumption is crucial, allowing the transformation from $p_\theta(\gamma_{1:T}|z_{1:T})$ to the product form $\prod_{t=1}^{T} p_\theta(\gamma_t|z_{1:t})$.

Following the VAE literature, we assume the conditional distribution $p_\theta(\gamma_t|z_{1:t}) \sim N(\mu_{t,\theta}, \sigma_{t,\theta})$, i.e., a Gaussian distribution with a diagonal covariance matrix. Also note that $\sigma_{t,\theta}$ is time-varying as from (5). For the prior distribution, we assume it is an independent Gaussian, namely, $p(z_t|z_{1:t-1}) \sim N(0, I)$. Though z_t^{prior}'s are independent, the decoder f_θ still allows us to capture the underlying correlation via the relationship between γ_t and $z_{1:t}^{prior}$. We design the decoder via a recurrent network f_{θ_1}, enclosing all time steps information $z_{1:t}^{prior}$ recursively, with a feedforward network f_{θ_2}, further transforming the input to $\mu_{t,\theta}$ and $\sigma_{t,\theta}$, i.e.,

$$h_{t,\theta_1} = f_{\theta_1}(h_{t-1,\theta_1}, z_t) \quad (\mu_{t,\theta}, \sigma_{t,\theta}) = f_{\theta_2}(h_{t,\theta_1}) \tag{9}$$

where h_{t,θ_1} is the hidden state in the RNN structure f_{θ_1}.

We stress that the decoder f_θ can be parameterized by any architecture, as long as the output γ_t depends on the latent variables $z_{1:t}^{prior}$, such as self-attention structures (Vaswani et al. 2017) or temporal convolutions (Oord et al. 2016).

Encoder design. Next, we will describe the posterior distribution, also known as the encoder. DT-VAE learns an encoder $f_\phi(\cdot)$ with parameter ϕ to encode observed $\gamma_{1:t}$ into the variational (posterior) distribution $q_\phi(z_{1:T}|\gamma_{1:T})$. We will use $f_\phi(\gamma_{1:t})$ and $p_\phi(z_{1:t}|\gamma_{1:t})$ interchangeably. To distinguish, we denote z^{post} for samples from the posterior distribution.

We factor the posterior distribution $q_\phi(z_{1:T}|\Gamma_{1:T})$ as

$$q_\phi(z_{1:T}|\gamma_{1:T}) = \prod_{t=1}^{T} q_\phi(z_t|z_{1:t-1}, \gamma_{1:t}) \tag{10}$$

During the training stage, we will sample z_t^{post} from the posterior distribution $q_\phi(z_t|z_{1:t-1}, \gamma_{1:t})$ and let the decoder reconstruct the observed γ_t's. For sampling from the posterior distribution, at each time step t, we sample z_t^{post} from the distribution conditioned on the historical posterior variables $z_{1:t-1}^{post}$ and all observed $\gamma_{1:t}$.

Following the VAE literature, we assume that variational distribution $q_\phi(z_t|z_{1:t-1}, \gamma_{1:t}) \sim N(\mu_{t,\phi}, \sigma_{t,\phi})$, i.e. a Gaussian distribution with a diagonal covariance matrix, where the $\mu_{t,\phi}$ and $\sigma_{t,\phi}$ are learned using the encoder f_ϕ. To capture that the historical information relies on both z^{post}'s and γ's, we decompose f_ϕ into three functions with parameters ϕ_1, ϕ_2 and ϕ_3:

$$\begin{aligned} h_{t,\phi_1} &= f_{\phi_1}(h_{t-1,\phi_1}, \gamma_t) \\ \mu_{t,\phi} &= f_{\phi_2}(h_{t,\phi_1}, \mu_{t-1,\phi}) \\ \sigma_{t,\phi} &= f_{\phi_3}(h_{t,\phi_1}, \sigma_{t-1,\phi}) \end{aligned} \tag{11}$$

where h_{t,ϕ_1} is the hidden state in RNN structure f_{ϕ_1}. For each time step, h_{t,ϕ_1} encodes all observed $\gamma_{1:t}$. The RNN structure f_{ϕ_2} will output the mean of posterior distribution $\mu_{t,\phi}$ by utilizing the h_{t,ϕ_1} and previous $\mu_{t-1,\phi}$. Therefore, for each time step, the current mean $\mu_{t,\phi}$ contains information of previous means $\mu_{1:t-1,\phi}$, which resembles the conditional structure in $q_\phi(z_t|z_{1:t-1}, \gamma_{1:t})$ from (10). Similarly, the RNN structure f_{ϕ_3} outputs $\sigma_{t,\phi}$ by utilizing h_{t,ϕ_1} and $\sigma_{t-1,\phi}$ that contain prior information of $\gamma_{1:t}$ and $z_{1:t-1}^{post}$.

DT-VAE-GAN. Our DT-VAE framework is general, flexible, and can be integrated with various generative models. We demonstrate its effectiveness when paired with VAE-GAN for improved generative accuracy. We use an RNN-based discriminator network. Different from VAE-GAN, we still maintain the original reconstruction loss as in DT-VAE (see section L in the appendix).

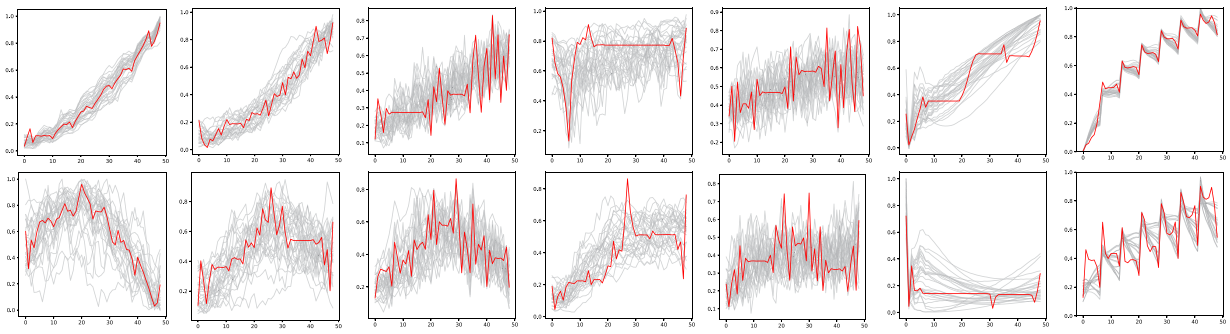
Experiment

In this section, we evaluate our methods on five datasets with various metrics and compare them with traditional and state-of-the-art methods. Our code is available at https://github.com/boilerchun/DT-VAE-codebase

Datasets and Baselines

Semi-Synthetic Data. We demonstrate the superiority of our methods over traditional statistical approaches like AR model by simulating the hospital census data. The daily arrivals $a(t)$ follow the discretized Cox-Ingersoll-Ross (CIR) process (Cox, Ingersoll Jr, and Ross 2005) with the drift function based on the day of the week, while daily discharges $d(t)$ are generated by simulating patient flow. See appendix section F.

COVID-19 County Daily Hospitalization (County-DH). We collect 2-year COVID-19 county hospitalization data from CA (CA Covid-19 Data Log 2023), NY (NY COVID-19 Data Log 2023), and PA (PA COVID-19 Data Log 2023), covering Medical/Surgical units and ICUs with our Week-of-day Quantile-Min normalization. See section F in the appendix.



(a) Real (I+E) (b) DT-VAE-GAN (c) DT-VAE (d) T-VAE-GAN (e) T-VAE (f) TimeGAN (g) PSA-GAN

Figure 4: First column provides the visualization for the groups of clusters (test set) in the State-DH dataset, including Increasing (I) and Early Peak (E) trends. Each of the next columns represents one of the 6 benchmarks. The centroids of each cluster (red) can represent the general trend for the corresponding time-series data (grey).

		DTP Score *(the lower the better)*				
Dataset	**Trends**	**TimeGAN**	**PSA-GAN**	**T-VAE-GAN**	**DT-VAE**	**DT-VAE-GAN**
County-DH	INC	0.4942±0.00001	0.4959±0.00009	0.4164±0.00054	0.8115±0.00055	**0.3717±0.00017**
	FLV	1.085±0.00000	1.3065±0.00033	0.8186±0.000362	0.8501±0.00048	**0.7636±0.00008**
	DEC	0.8640±0.00019	0.9469±0.00012	1.2546±0.00075	0.6259±0.00012	**0.3271±0.00003**
	LTP	**0.4954±0.00001**	0.6506±0.00120	1.0099±0.000031	0.8759±0.00022	0.5225±0.00002
States-DH	INC	0.3252±0.00124	0.3420±0.00009	0.6570±0.00282	0.8852±0.00217	**0.3240±0.00029**
	ERP	0.8574±0.00011	1.3928±0.00001	0.8067±0.11438	0.7603±0.00173	**0.7564±0.00191**
	DEC	0.4392±0.00066	0.6419±0.00051	0.3359±0.00055	0.3178±0.00002	**0.3135±0.00003**
	LTP	0.7095±0.02287	1.1213±0.00328	0.9031±0.00304	0.9356±0.00015	**0.4566±0.00232**
Energy	INC	**0.2286±0.00001**	0.5515±0.00001	1.3726±0.00063	0.8860±0.00043	0.3897±0.00059
	DEC	0.3780±0.00002	0.6840±0.00002	1.2716±0.01282	0.6679±0.00034	**0.2366±0.00068**
Traffic	INC	0.2153±0.00000	0.4534±0.00017	1.5622±0.00165	0.6945±0.00002	**0.2119±0.00000**
	DEC	0.7286±0.00006	0.9257±0.00023	1.9073±0.00596	1.0044±0.00005	**0.3707±0.00003**
Stock	INC	0.6723 ± 0.01629	0.3246±0.00039	0.3066±0.00396	0.3135±0.00012	**0.2601±0.00022**
	DEC	0.4796 ± 0.01595	0.3361±0.00022	0.3847±0.00213	0.3885±0.00302	**0.2860±0.00023**

Table 1: Results for MS units in County/States-DH and other benchmarks, with their trends: Increasing (INC), Flat Valley (FLV), Decreasing (DEC), and Late Peak (LTP). Bold indicates the best performance, underlined shows the next-best performance

COVID-19 States Daily Hospitalization (States-DH). Similarly, we obtain 2-year state-level hospitalization data from US Health Department (Covid-19 Data by states 2023).

Solar Energy (Energy). Hourly solar energy collection data in Alabama State from 137 stations (Lai et al. 2018a).

Traffic Management (Traffic). Hourly occupancy rates within 963 lanes in San Francisco (Cuturi 2011).

Stock Market (Stock). Daily Google stock data from 2004-2023 (Yahoo Finance stock history 2023). Notably, the lack of typical inflow/outflow assumptions in this domain demonstrates the wide applicability of our methods.

Baselines. We compare with Autoregressive (AR) models and state-of-the-art models (TimeGAN and PSA-GAN (Yoon, Jarrett, and Van der Schaar 2019; Jeha et al. 2021)). Temporal VAE (T-VAE) and Temporal VAE-GAN (T-VAE-GAN) denote the same structure as DT-methods but directly learn from the data. See sections M and K in the appendix.

Evaluation Metrics

Accurate temporal patterns are crucial for downstream tasks (Shi et al. 2022; Park, Yun, and Ahn 2009; Dalamagkidis et al. 2007). Our evaluation metrics follow two principles: (i) assessing the extent to which the generative model preserves patterns in real data, (ii) evaluating the performance of using generated data in challenging downstream tasks.

Evaluation for Semi-Synthetic Data. For purposes (i) and (ii), we employ *Direct Comparison* to assess since we know the "ground-truth" temporal trend from the model parameters. We mainly compare our methods with traditional statistical methods like auto-regressive models.

Evaluation for Real Dataset. Without ground truth, we design new evaluation frameworks for purposes (i) and (ii).

Diverse Trend Preservation (DTP) evaluation. We use the K-means Dynamic Time Warping Barycenter Averaging (DBA) algorithm to identify distinct trend patterns in time-series datasets (Tavenard et al. 2020; Petitjean, Ketterlin, and

			Long-term Prediction Score *(the lower the better)*				
Dataset	**Trends**	**Original**	**TimeGAN**	**PSA-GAN**	**T-VAE-GAN**	**DT-VAE**	**DT-VAE-GAN**
County-DH	INC	0.2401±0.0951	0.9848±0.1055	1.4362±0.0822	6.2268±0.1393	3.1878±1.1980	**0.9396±0.5801**
	DEC	0.0959±0.0144	1.8062±0.1364	0.7098±0.0665	2.5431±0.1062	**0.5983±0.0546**	0.6804±0.2742
State-DH	INC	0.1252±0.0544	0.5258±0.0175	0.4873±0.0746	2.2192±0.2093	0.3223±0.0412	**0.1666±0.0196**
	DEC	0.0186±0.0129	0.2020±0.0293	0.1639±0.0100	0.7234±0.1095	0.1642±0.0453	**0.1091±0.0904**
Energy	INC	0.0159±0.0017	0.0196±0.0015	0.0612±0.0032	0.9028±0.4065	0.0646±0.0089	**0.0152±0.0005**
	DEC	0.0034±0.0008	0.0059±0.0016	0.0357±0.0174	0.2949±0.1229	0.0311±0.0062	**0.0056±0.0015**
Traffic	INC	0.0055±0.0002	0.0097±0.0022	0.0146±0.0023	0.0391±0.0056	0.0134±0.0010	**0.0080±0.0003**
	DEC	0.0009±0.0000	0.0056±0.0013	0.0058±0.0015	0.0276±0.0073	0.0027±0.0007	**0.0010±0.0001**
Stock	INC	0.2650±0.0179	0.3263±0.0042	0.3050±0.0111	1.1663±0.0168	0.2706±0.0084	**0.2665±0.0081**
	DEC	0.2161±0.0114	0.3446±0.0154	0.4333±0.0486	1.9552±0.0956	0.3116±0.0160	**0.2594±0.0307**

Table 2: Results on long-term prediction task using first 40 time steps to predict next 9 time steps. MSE and corresponding standard deviations are reported. Bold indicates the best performance, underlined indicates the next-best performance.

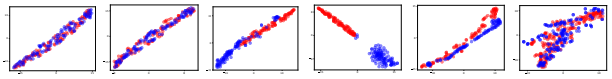


Figure 5: t-SNE for original data (red) and generated data (blue) in State DH's decreasing trends with DT-VAE-GAN, DT-VAE, T-VAE-GAN, T-VAE, TimeGAN, PSA-GAN.

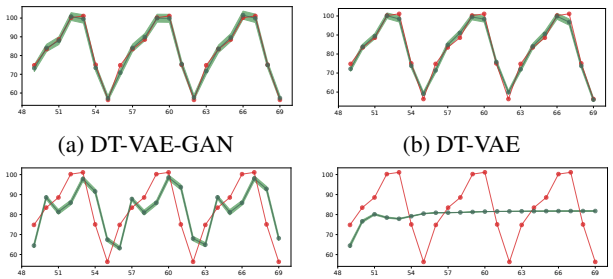
Gançarski 2011). Using the Dynamic Time Warping (DTW) distance, the DBA clusters similar patterns (Müller 2007). Distinct trends are identified in each dataset; see section H in the appendix. From the identified clusters, we construct training and test data by sampling sequences from combinations of these distinct clusters. This introduces an input containing multiple patterns. We then reapply the K-means DBA on the generated data, examining its ability to represent the original input patterns. Our Diverse Trend Preservation (DTP) visualization portrays the trends detected in the hold-out test dataset and the generated dataset. For a quantitative assessment, we use the DTW distance to determine the DTP score, measuring the resemblance between cluster centroids from the test dataset and those in the generated dataset. Our DTP evaluation, unlike PSA-GAN's FID score and TimeGAN's discriminative score, provides visual insights into patterns in generated data and operates independently from pre-trained models. See section I in the appendix.

Evaluations on downstream tasks. For the "Long-term Prediction" task, we use historical data to forecast future values, which is a prevalent task in practice. We employ the "Train on Synthetic, Test on Real" (TSTR) strategy (Esteban, Hyland, and Rätsch 2017), training LSTM on generated data and testing on a hold-out set. The post-hoc model predicts future segments based on initial chunks, with performance measured by Mean Square Error (MSE) on the hold-out set.

Results Comparison

Direct comparison. Figure 6 shows the mean and confidence interval of generated samples. We can observe the superior performance of our methods in capturing the temporal trend.

DTP Visualization and Score. Figure 4 shows our methods outperform other benchmarks in preserving the original trends. The generated data closely matches the increasing trend (first row) observed in the real data. This observation is confirmed by the DTP scores in Table 1. Our methods demonstrate up to a 30% ↓ improvement, averaging at 12% ↓ decrease. Additionally, our methods tend to perform equally well for different clusters. See appendix table 3 and section I.



(a) DT-VAE-GAN (b) DT-VAE

(c) Short Term AR(3) (d) Long Term AR(3)

Figure 6: Synthetic data's mean from days 49-70 (test period) is in red. Generated time series mean is in dark green, with the 0.95 confidence interval in light green.

Prediction Scores. We use different periods of time-series data to predict subsequent periods: the first 40 time steps predict the next 9 time steps in table 2, the first 49 time steps predict the next 21 time steps in table 4 in the appendix, and the first 56 predict the next 35 time steps in table 5 in the appendix. As shown in table 2, our methods consistently outperform benchmarks, achieving on average a 30% ↓ improvement and up to 65% ↓. More insights are provided in the table 4, table 5, and the section J in the appendix. Note that we reduce the number of clusters mainly to increase the number of training/testing samples for the prediction task.

t-SNE visualization. Figure 5 demonstrates our methods align better with the original data compared to benchmarks. TimeGAN's narrow t-SNE pattern indicates limited trend diversity, aligning with DTP results. See section N in appendix.

Acknowledgments

This work was partially supported by NSF IIS 1955890, IIS 2146091. T. Li is supported by Gayle and Frank Johnston student scholarship

References

Abbasimehr, H.; Paki, R.; and Bahrini, A. 2022. A novel approach based on combining deep learning models with statistical methods for COVID-19 time series forecasting. *Neural Computing and Applications*, 1–15.

Aksan, E.; and Hilliges, O. 2019. STCN: Stochastic temporal convolutional networks. *arXiv preprint arXiv:1902.06568*.

Alaa, A.; Chan, A. J.; and van der Schaar, M. 2020. Generative time-series modeling with fourier flows. In *International Conference on Learning Representations*.

Bai, S.; Kolter, J. Z.; and Koltun, V. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

CA Covid-19 Data Log. 2023. Covid-19 hospital data. https://data.ca.gov/dataset/covid-19-hospital-data1. Accessed: 2023-03-26.

Calatayud, J.; Jornet, M.; and Mateu, J. 2023. Spatio-temporal stochastic differential equations for crime incidence modeling. *Stochastic Environmental Research and Risk Assessment*, 37(5): 1839–1854.

CDC COVID-19 Tracking . 2023. Covid Data Tracker Weekly Review. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html. Accessed: 2023-03-18.

Chellasamy, A.; and Nagarathinam, A. 2022. An Overview of Augmenting AI Application in Healthcare. In Pandian, A. P.; Fernando, X.; and Haoxiang, W., eds., *Computer Networks, Big Data and IoT*, 397–407. Singapore: Springer Nature Singapore. ISBN 978-981-19-0898-9.

Covid-19 Data by states. 2023. Covid-19 reported patient impact and hospital capacity by state timeseries. https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh. Accessed: 2023-03-25.

Cox, J. C.; Ingersoll Jr, J. E.; and Ross, S. A. 2005. A theory of the term structure of interest rates. In *Theory of valuation*, 129–164. World Scientific.

Cuturi, M. 2011. PEMS-SF. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C52G70.

Dalamagkidis, K.; Kolokotsa, D.; Kalaitzakis, K.; and Stavrakakis, G. 2007. Reinforcement learning for energy conservation and comfort in buildings. *Building and Environment*, 42(7): 2686–2698.

Desai, A.; Freeman, C.; Wang, Z.; and Beaver, I. 2021. TimeVAE: A Variational Auto-Encoder for Multivariate Time Series Generation. *arXiv preprint arXiv:2111.08095*.

Dheeru, D.; and Karra Taniskidou, E. 2017. UCI Machine Learning Repository.

Er, S.; Yang, S.; and Zhao, T. 2023. County augmented transformer for COVID-19 state hospitalizations prediction. *Scientific Reports*, 13(1): 9955.

Esteban, C.; Hyland, S. L.; and Rätsch, G. 2017. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint arXiv:1706.02633*.

Fabius, O.; and Van Amersfoort, J. R. 2014. Variational recurrent auto-encoders. *arXiv preprint arXiv:1412.6581*.

Forestier, G.; Petitjean, F.; Dau, H. A.; Webb, G. I.; and Keogh, E. 2017. Generating synthetic time series to augment sparse datasets. In *2017 IEEE international conference on data mining (ICDM)*, 865–870. IEEE.

Gontis, V.; Ruseckas, J.; and Kononovičius, A. 2010. A long-range memory stochastic model of the return in financial markets. *Physica A: Statistical Mechanics and its Applications*, 389(1): 100–106.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Jarrett, D.; Bica, I.; and van der Schaar, M. 2021. Time-series generation by contrastive imitation. *Advances in Neural Information Processing Systems*, 34: 28968–28982.

Jeha, P.; Bohlke-Schneider, M.; Mercado, P.; Kapoor, S.; Nirwan, R. S.; Flunkert, V.; Gasthaus, J.; and Januschowski, T. 2021. PSA-GAN: Progressive Self Attention GANs for Synthetic Time Series. In *International Conference on Learning Representations*.

Kingma, D. P.; and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kiyasseh, D.; Tadesse, G. A.; Thwaites, L.; Zhu, T.; Clifton, D.; et al. 2020. PlethAugment: GAN-based PPG augmentation for medical diagnosis in low-resource settings. *IEEE journal of biomedical and health informatics*, 24(11): 3226–3235.

Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. 2018a. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. arXiv:1703.07015.

Lai, G.; Li, B.; Zheng, G.; and Yang, Y. 2018b. Stochastic wavenet: A generative latent variable model for sequential data. *arXiv preprint arXiv:1806.06116*.

Larsen, A. B. L.; Sønderby, S. K.; Larochelle, H.; and Winther, O. 2016. Autoencoding beyond pixels using a learned similarity metric. In *International conference on machine learning*, 1558–1566. PMLR.

Lebacque, J.-P. 2005. Intersection modeling, application to macroscopic network traffic flow models and traffic management. In *Traffic and Granular Flow'03*, 261–278. Springer.

Li, H.; Yu, S.; and Principe, J. 2023. Causal Recurrent Variational Autoencoder for Medical Time Series Generation. *arXiv preprint arXiv:2301.06574*.

Littig, S. J.; and Isken, M. W. 2007. Short term hospital occupancy prediction. *Health care management science*, 10: 47–66.

Mogren, O. 2016. C-RNN-GAN: Continuous recurrent neural networks with adversarial training. *arXiv preprint arXiv:1611.09904*.

Mugi, V. R.; and Chandramohan, V. 2021. Energy and exergy analysis of forced and natural convection indirect solar dryers: Estimation of exergy inflow, outflow, losses, exergy efficiencies and sustainability indicators from drying experiments. *Journal of Cleaner Production*, 282: 124421.

Müller, M. 2007. Dynamic time warping. *Information retrieval for music and motion*, 69–84.

NY COVID-19 Data Log. 2023. Covid-19 Hospitalizations Weekly data. https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Hospitalizations/jw46-jpb7/data. Accessed: 2023-03-16.

Oord, A. v. d.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; and Kavukcuoglu, K. 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

PA COVID-19 Data Log. 2023. Covid-19 Hospitalizations Weekly data. https://data.pa.gov/Covid-19/COVID-19-Aggregate-Hospitalizations-Current-Weekly/kayn-sjhx. Accessed: 2023-03-26.

Park, B. Yun, I.; and Ahn, K. 2009. Stochastic Optimization for Sustainable Traffic Signal Control. *International Journal of Sustainable Transportation*, 3(4): 263–284.

Petitjean, F.; Ketterlin, A.; and Gançarski, P. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition*, 44(3): 678–693.

Regenstrief Institute COVID-19 Dashboard. 2023. COVID-19 Dashboard. https://www.regenstrief.org/covid-dashboard/. Accessed: 2023-1-24.

Rezende, D. J.; Mohamed, S.; and Wierstra, D. 2014. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, 1278–1286. PMLR.

Shi, P.; Helm, J. E.; Chen, C.; Lim, J.; Parker, R. P.; Tinsley, T.; and Cecil, J. 2022. Operations (management) warp speed: Rapid deployment of hospital-focused predictive/prescriptive analytics for the COVID-19 pandemic. *Production and Operations Management*.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

Srinivasan, P.; and Knottenbelt, W. J. 2022. Time-series Transformer Generative Adversarial Networks. *arXiv preprint arXiv:2205.11164*.

Tahmasbi, R.; and Hashemi, S. M. 2014. Modeling and Forecasting the Urban Volume Using Stochastic Differential Equations. *IEEE Transactions on Intelligent Transportation Systems*, 15(1): 250–259.

Tavenard, R.; Faouzi, J.; Vandewiele, G.; Divo, F.; Androz, G.; Holtz, C.; Payne, M.; Yurchak, R.; Rußwurm, M.; Kolar, K.; and Woods, E. 2020. Tslearn, A Machine Learning Toolkit for Time Series Data. *Journal of Machine Learning Research*, 21(118): 1–6.

Tzen, B.; and Raginsky, M. 2019. Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. *arXiv preprint arXiv:1905.09883*.

van Dyk, D. A.; and Meng, X.-L. 2001. The Art of Data Augmentation. *Journal of Computational and Graphical Statistics*, 10(1): 1–50.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wen, Q.; Sun, L.; Yang, F.; Song, X.; Gao, J.; Wang, X.; and Xu, H. 2021. Time Series Data Augmentation for Deep Learning: A Survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization.

WHO COVID-19 Dashboard. 2023. WHO coronavirus (COVID-19) dashboard. https://covid19.who.int/? Accessed: 2023-03-24.

Xu, T.; Wenliang, L. K.; Munn, M.; and Acciaio, B. 2020. Cot-gan: Generating sequential data via causal optimal transport. *Advances in Neural Information Processing Systems*, 33: 8798–8809.

Yahoo Finance stock history. 2023. Finance 2023.

Yoon, J.; Jarrett, D.; and Van der Schaar, M. 2019. Time-series generative adversarial networks. *Advances in neural information processing systems*, 32.