RDPI: A Refine Diffusion Probability Generation Method for Spatiotemporal Data Imputation

Zijin Liu[1], **Xiang Zhao**[2], **You Song**[2*]

[1]School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2]School of Software, Beihang University, Beijing 100191, China
{liuzijin, by2321118, songyou}@buaa.edu.cn

Abstract

Spatiotemporal data imputation plays a crucial role in various fields such as traffic flow monitoring, air quality assessment, and climate prediction. However, spatiotemporal data collected by sensors often suffer from temporal incompleteness, and the sparse and uneven distribution of sensors leads to missing data in the spatial dimension. Among existing methods, autoregressive approaches are prone to error accumulation, while simple conditional diffusion models fail to adequately capture the spatiotemporal relationships between observed and missing data. To address these issues, we propose a novel two-stage **R**efined **D**iffusion **P**robability **I**mpuation (RDPI) framework based on an initial network and a conditional diffusion model. In the initial stage, deterministic imputation methods are used to generate preliminary estimates of the missing data. In the refinement stage, residuals are treated as the diffusion target, and observed values are innovatively incorporated into the forward process. This results in a conditional diffusion model better suited for spatiotemporal data imputation, bridging the gap between the preliminary estimates and the true values. Experiments on multiple datasets demonstrate that RDPI not only achieves state-of-the-art imputation performance but also significantly reduces sampling computational costs.

Introduction

Spatiotemporal data encompass both temporal and spatial dimensions, primarily including traffic flow data, meteorological data, environmental monitoring data, etc. playing a crucial role in scientific research and social development. However, due to equipment aging, sensor malfunctions, or communication issues, spatiotemporal data often experiences partial or even complete data loss from sensors, which severely affects the reliability and effectiveness of subsequent analyses. Therefore, effectively imputing spatiotemporal data has become a crucial research issue.

Currently, methods for addressing spatiotemporal data imputation issues are mainly categorized into deterministic methods, probabilistic methods, and diffusion methods. Deterministic methods attempt to impute values using rules or deterministic algorithms, such as Recurrent Neural Networks (RNNs) (Cao et al. 2018; Yoon, Zame, and van der Schaar 2018; Che et al. 2018), Transformer-based Networks (Vaswani et al. 2017; Du, Côté, and Liu 2023; Shan, Li, and Oliva 2023), and Graph Neural Networks (GNNs) (Cini, Marisca, and Alippi 2021; Marisca, Cini, and Alippi 2022) have been widely employed. These methods leverage the powerful representation learning capabilities of neural networks to learn complex spatiotemporal patterns from data and use learned models to predict missing data. However, deterministic methods often lack modeling of data uncertainty, which can pose challenges when dealing with complex spatiotemporal dependencies.

Probabilistic methods use statistical models or machine learning techniques to model the probability distribution of data and perform imputation based on these distributions. In addition to Variational Autoencoders (VAE) (Kim et al. 2023; Mulyadi, Jun, and Suk 2021; Fortuin et al. 2020) and Generative Adversarial Networks (GAN) (Luo et al. 2018; Liu et al. 2019; Miao et al. 2021), other methods such as generative models and Bayesian networks are also utilized. These methods not only learned distributions from existing data, but also leveraged the latent structures and patterns within the data to handle data uncertainty. For example, VAE learns a latent-space representation of the data, effectively imputing missing data.

To better simulate the diffusion process of data over time and space and capture the dynamic changes and complex dependencies of spatiotemporal data, researchers have proposed methods based on diffusion models (Ho, Jain, and Abbeel 2020; Han, Zheng, and Zhou 2022). The Conditional Diffusion model for Imputation (CDI) (Tashiro et al. 2021; Liu et al. 2023; Wang et al. 2023) is widely adopted as a typical approach. However, traditional CDI models often only consider conditional factors during denoising network training, neglecting them in the forward and imputation process. This limitation may result in conditional models that fail to fully capture the complex dependencies in spatiotemporal traffic data, thereby affecting their performance in practical data imputation tasks.

To address the key challenges in spatiotemporal data imputation, we propose a novel two-stage refined diffusion probability imputation (RDPI) framework. In the first stage, a preliminary estimate is generated using an initial network. In the second stage, a novel conditional diffusion model is proposed to refine the results. By integrating conditional in-

*Corresponding author.

formation throughout the diffusion process, the framework significantly improves the performance and effectiveness of imputing missing data.

Key contributions of this paper include:

- Introduction of the RDPI framework, which employs a two-stage imputation strategy to achieve higher performance in data imputation.
- Incorporation of observed values into the forward process, leading to a conditional diffusion model that is better suited for spatiotemporal data imputation and utilizing residuals as the diffusion target to effectively enhance imputation efficiency.
- Extensive empirical validation demonstrates that the proposed RDPI excels in traffic data imputation and air quality data imputation tasks, achieving state-of-the-art performance.

These achievements provide new methodologies and theoretical support for addressing urban traffic data imputation challenges, potentially offering more reliable data foundations for urban traffic management and planning.

Related Work

Deterministic Methods Deterministic methods rely on rules or algorithms to impute missing spatiotemporal data, learning patterns to predict missing values. Early methods like KNN (Hastie et al. 2009) and Kriging (Wu et al. 2021) used simple historical or global information, while autoregressive approaches such as ARIMA (Box et al. 2015) and VAR (Zivot and Wang 2006) incorporated temporal trends. More recent advancements focus on spatiotemporal dependencies, with methods like GRAPE (You et al. 2020), GRIN (Cini, Marisca, and Alippi 2021), and BRITS (Cao et al. 2018) leveraging graph-based and recurrent neural networks to enhance imputation.

Probabilistic Methods Probabilistic methods use generative models and Bayesian frameworks to model data distributions and handle uncertainty. Examples include GP-VAE (Fortuin et al. 2020), which captures feature correlations, and MIWVAE (Mattei and Frellsen 2019), which avoids missing data costs using importance weighting. Other approaches, such as GAIN (Yoon, Jordon, and Schaar 2018), SSGAN (Miao et al. 2021), and STGNP (Hu et al. 2023), apply adversarial networks and graph neural processes to improve spatiotemporal imputation while accounting for uncertainty.

Diffusion Methods Diffusion models have emerged as a leading approach for spatiotemporal imputation. CSDI (Tashiro et al. 2021) pioneered conditional diffusion for time-series data, while SSSD (Alcaraz and Strodthoff 2022) integrated state-space models with Diffwave (Kong et al. 2021). PriSTI (Liu et al. 2023) introduced noise prediction models and global context priors, and MIDM (Wang et al. 2023) leveraged Gaussian processes for handling latent distributions. Despite their success, many diffusion models overlook conditional factors during imputation, limiting their ability to capture complex spatiotemporal dependencies and reducing effectiveness in practical scenarios.

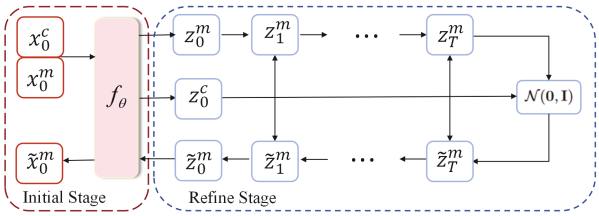

Figure 1: The directed graphical model of RDPI framework. In the initial stage, the rough imputation result z_0^m is computed from the observed values x_0^c and the missing data x_0^m using a deterministic model f_θ. In the refine stage, a novel conditional diffusion model is introduced to generate the residual z_0^m between the rough imputation result $f_\theta(x_0^c)$ and the true values x_0^m, ultimately yielding a refined imputation result for the missing data. In this figure, z_t^m correspond to representation of residual at t-th step of diffusion process, z_0^c is the representation of x_0^c obtained from f_θ, $\widetilde{x}_0^m$ is the ultimate imputation result, $\mathcal{N}(\mathbf{0},\mathbf{I})$ correspond the Standard Gaussian.

Preliminaries

We formalize a spatiotemporal data set $\mathbf{x} \in \mathbb{R}^{N\times D}$ as sequential data consisting of N time series of length L. Any data in the dataset can be missing, meaning their values are unknown. We divide the dataset into two parts: observed data $\mathbf{x}^c$ and missing data $\mathbf{x}^m$. Specifically, both observed and missing data are obtained from a binary mask matrix $\mathbf{m} \in \{0,1\}^{N\times D}$, where observed data are denoted as $\mathbf{x}^c = \{\mathbf{x}^{i,j}|\mathbf{m}^{i,j} = 1\}$ and missing data as $\mathbf{x}^m = \{\mathbf{x}^{i,j}|\mathbf{m}^{i,j} = 0\}$.

The problem of spatiotemporal imputation can be defined as estimating $p(\mathbf{x}^m|\mathbf{x}^c)$.

Method

The proposed RDPI framework, shown in Figure 1, consists of two stages: initialization and refinement. In the initialization stage, missing data is first imputed using a deterministic model. In the refinement stage, a novel conditional diffusion probability model is proposed to enhance the imputation results, aiming to narrow the gap between the initial values and the true values.

Existing conditional diffusion imputation methods focus primarily on the reverse process, ignoring the relationship between missing and observed data in the forward process. We address this by defining conditional probability distributions for both processes and deriving a novel the Evidence Lower Bound (ELBO) that includes forward conditions. Based on this, we introduce a new conditional diffusion imputation model (see Appendix of (Liu, Zhao, and Song 2024) for details), followed by the reverse process, forward process, ELBO, and training and imputation algorithms.

Initial Stage

In our experiments, the deterministic model f_θ is implemented using GRIN (Cini, Marisca, and Alippi 2021), with

the loss function for the initial stage defined as:

$$\mathcal{L}_{\text{init}} = \|f_\theta(\mathbf{x}_0^c) - \mathbf{x}_0^m\| \tag{1}$$

where $\mathbf{x}_0^m$ denotes the ground truth for the missing data, and $f_\theta(\mathbf{x}_0^c)$ represents the initial imputation.

The noise-disturbed target for our novel conditional diffusion model g_θ is the residual $\mathbf{z}_0^m$, which is defined as the difference between $\mathbf{x}_0^m$ and $f_\theta(\mathbf{x}_0^c)$. While pretraining the initial model is not strictly required, as gradients from joint training propagate into g_θ through f_θ (Whang et al. 2022), we recommend pretraining to enhance the stability of the diffusion model.

Unlike (Whang et al. 2022), we define the loss function as $\|f_\theta(\mathbf{x}_0^c) - \mathbf{x}_0^m\|$ rather than $\|\mathbf{x}_0^m - f_\theta(\mathbf{x}_0^c)\|$. Experimental results show that this formulation gradually reduces the loss of the initial model during training, ultimately converging at the imputation capacity limit of initial model f_θ. This ensures that the diffusion model g_θ can effectively and stably model the data distribution, leading to improved residual generation capacity.

Reverse Process

Following the steps of Denoising Diffusion Probabilistic Models (DDPM) (Ho, Jain, and Abbeel 2020), the proposed conditional diffusion model is a latent variable model of the form $p(\mathbf{z}_0^m|\mathbf{z}_0^c) = \int p(\mathbf{z}_{0:T}^m|\mathbf{z}_0^c)d\mathbf{z}_{1:T}^m$, where $\mathbf{z}_{1:T}^m$ is a latent variable with the same dimension as the residual $\mathbf{z}_0^m$ conditioned on observed data $\mathbf{z}_0^c$. The joint probability $p(\mathbf{z}_{0:T}^m|\mathbf{z}_0^c)$ is referred to as the reverse process, defined on a Markov chain:

$$\begin{aligned} p(\mathbf{z}_{0:T}^m|\mathbf{z}_0^c) &:= p(\mathbf{z}_T^m|\mathbf{z}_0^c)\prod_{t=1}^{T} p(\mathbf{z}_{t-1}^m|\mathbf{z}_t^m, \mathbf{z}_0^c) \\ p(\mathbf{z}_{t-1}^m|\mathbf{z}_t^m, \mathbf{z}_0^c) &:= \mathcal{N}(\mathbf{z}_{t-1}^m; \mu_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, t), \mathbf{\Sigma}_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, t)) \end{aligned} \tag{2}$$

where $p(\mathbf{z}_T^m|\mathbf{z}_0^c)$ represents the endpoint of the conditional diffusion process, which is assumed to follow a standard Gaussian distribution. In this context, μ_θ and $\mathbf{\Sigma}_\theta$ denote the mean and variance, respectively, and can be estimated using neural networks. We fix the $\mathbf{\Sigma}_\theta$ to be constant.

Forward Process

Given a data point $\mathbf{z}_0^m$ sampled from the true spatiotemporal data distribution $p(\mathbf{z}_0^m)$, we define the forward process or diffusion process as a Markov chain with learned Gaussian transitions that starts from the sampled data distribution:

$$\begin{aligned} q(\mathbf{z}_{1:T}^m|\mathbf{z}_0^m, \mathbf{z}_0^c) &:= \prod_{t=1}^{T} q(\mathbf{z}_t^m|\mathbf{z}_{t-1}^m, \mathbf{z}_0^c) \\ q(\mathbf{z}_t^m|\mathbf{z}_{t-1}^m, \mathbf{z}_0^c) &:= \mathcal{N}(\mathbf{z}_t^m; \sqrt{1-\beta_t}\mathbf{z}_{t-1}^m + \sqrt{1-\beta_t}\mathbf{z}_0^c, \beta_t\mathbf{I}) \end{aligned} \tag{3}$$

where β_t is variance from a given variance schedule $\beta_1, \cdots, \beta_T$.

At each transition step, a small amount of Gaussian noise is added to the data through Gaussian transitions. The step size is controlled by a variable schedule $\{\beta_t \in (0,1)\}_{t=1}^T$. Let $\overline{\alpha}_t := 1 - \beta_t$ and $\alpha_t = \prod_{s=1}^{t} \overline{\alpha}_t$, we sample $\mathbf{z}_t^m$ directly from $\mathbf{z}_0^m$ with an arbitrary timestep t:

$$q(\mathbf{z}_t^m|\mathbf{z}_0^m, \mathbf{z}_0^c) = \mathcal{N}(\mathbf{z}_t^m; \sqrt{\alpha_t}(\mathbf{z}_0^m + \mathbf{z}_0^c), (1-\alpha_t)\mathbf{I}) \tag{4}$$

Training

With the light of reverse process and forward process, a novel ELBO of the conditional diffusion model can be derived as follows :

$$\begin{aligned} &\log p(\mathbf{z}_0^m|\mathbf{z}_0^c) \\ &= \mathbb{E}_{q(\mathbf{z}_{1:T}^m|\mathbf{z}_0^m, \mathbf{z}_0^c)}[D_{KL}(q(\mathbf{z}_T^m|\mathbf{z}_0^m, \mathbf{z}_0^c)||p(\mathbf{z}_T^m|\mathbf{z}_0^c)) \\ &+ \sum_{i=2}^{T} D_{KL}(q(\mathbf{z}_{t-1}^m|\mathbf{z}_t^m, \mathbf{z}_0^m, \mathbf{z}_0^c)||p(\mathbf{z}_{t-1}^m|\mathbf{z}_t^m, \mathbf{z}_0^c)) \\ &+ \log p(\mathbf{z}_0^m|\mathbf{z}_1^m, \mathbf{z}_0^c)] \end{aligned} \tag{5}$$

Using Bayes' rule and following the standard Gaussian density function, the mean and variance can be parameterized as follows:

$$\widetilde{\beta}_t = \frac{1-\alpha_{t-1}}{1-\alpha_t}\beta_t \tag{6}$$

$$\begin{aligned} &\widetilde{\mu}_t(\mathbf{z}_t^m, \mathbf{z}_0^c, t) \\ &= \frac{1}{\sqrt{\overline{\alpha}_t}}(\mathbf{z}_t^m - \frac{\overline{\alpha}_t\sqrt{\overline{\alpha}_t}(1-\alpha_{t-1})}{1-\alpha_t}\mathbf{z}_0^c - \frac{1-\overline{\alpha}_t}{\sqrt{(1-\alpha_t)}}\epsilon_t) \end{aligned} \tag{7}$$

We need to learn a neural network to estimate the conditioned probability distributions in the reverse diffusion process $p(\mathbf{z}_{t-1}^m|\mathbf{z}_t^m, \mathbf{z}_0^c) = \mathcal{N}(\mathbf{z}_{t-1}^m; \mu_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, t), \mathbf{\Sigma}_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, t))$. We would like to train μ_θ to predict $\widetilde{\mu}_t$. We can reparameterize the Gaussian noise term instead to make it predict ϵ from the input $\mathbf{z}_t^m$ at time step t:

$$\begin{aligned} \mu_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, t) = \frac{1}{\sqrt{\overline{\alpha}_t}}(\mathbf{z}_t^m - \frac{\overline{\alpha}_t\sqrt{\overline{\alpha}_t}(1-\alpha_{t-1})}{1-\alpha_t}\mathbf{z}_0^c \\ - \frac{1-\overline{\alpha}_t}{\sqrt{(1-\alpha_t)}}\epsilon_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, \mathbf{z}_0^c, t)) \end{aligned} \tag{8}$$

The simplified training objective can be written:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{\mathbf{z}_0^m, \epsilon}[\|\epsilon_t - \epsilon_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, \mathbf{z}_0^c, t)\|] \tag{9}$$

We follow the convention to assume it will be close to zero by carefully diffusing the observed response variable $\mathbf{z}_0^c$ towards a pre-assumed distribution $p(\mathbf{z}_T^m|\mathbf{z}_0^c)$.

At last, joint loss of RDPI becomes:

$$\mathcal{L}_{\text{joint}} = \mathcal{L}_{\text{simple}} + \lambda\mathcal{L}_{\text{init}} \tag{10}$$

where λ is the hyperparameter that trades off between the initial loss and diffusion loss. The complete training procedure has been displayed in Algorithm 1.

Imputation

After training denoising networks at each diffusion step, the likelihood of residual distributions is derived using Eq. (8). Missing data is generated by sampling from standard Gaussian noise, as shown in Algorithm 2. Similar to DDPM, diffusion models face high computational demands during iterative sampling. To accelerate this process, RDPI can adopt

Algorithm 1: Training of RDPI

Pre-train the deterministic imputation model f_θ
while not converged **do**
  Draw $\mathbf{x}_0^m \sim q(\mathbf{x}_0^m|\mathbf{x}_0^c)$
  $\mathbf{z}_0^m = f_\theta(\mathbf{x}_0^c) - \mathbf{x}_0^m$
  Compute initial loss $\mathcal{L}_{\text{init}} = \|\mathbf{z}_0^m\|$
  Draw $t \sim \text{Uniform}(\{1, \cdots, T\})$
  Draw $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Compute diffusion loss $\mathcal{L}_{\text{simple}} = \|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, \mathbf{z}_0^c, t)\|^2$
  Compute joint loss $\mathcal{L}_{\text{joint}} = \mathcal{L}_{\text{simple}} + \lambda\mathcal{L}_{\text{init}}$
  Take gradient descent step on $\nabla_\theta\mathcal{L}_{\text{joint}}$
end while

Algorithm 2: Imputatation (Sampling) with RDPI

$\mathbf{x}_{init}^m = f_\theta(\mathbf{x}_0^c)$
$\mathbf{z}_T^m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T$ to 1 **do**
  $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$ else $\boldsymbol{\epsilon}_t = \mathbf{0}$
  $\mathbf{z}_{t-1}^m = \frac{1}{\sqrt{\overline{\alpha}_t}}\left(\mathbf{z}_t^m - \frac{\overline{\alpha}_t\sqrt{\overline{\alpha}_t}(1-\alpha_{t-1})}{1-\alpha_t}\mathbf{z}_0^c\right.$
  $\left. - \frac{1-\overline{\alpha}_t}{\sqrt{(1-\alpha_t)}}\epsilon_\theta(\mathbf{z}_t^m, \mathbf{z}_0^c, \mathbf{z}_0^c, t)\right) + \sigma_t\boldsymbol{\epsilon}_t$
end for
return $\mathbf{x}_{init}^m - \mathbf{z}_0^m$

any unconditional DDPM sampling scheme, such as DDIM (Song, Meng, and Ermon 2021). However, unlike the original DDIM, removing randomness may degrade imputation quality. Therefore, RDPI retains accelerated schemes while preserving random components to ensure quality.

Denoising Model

In this section, we introduce the proposed denoising model. The specific process is illustrated in Figure 2. Our denoising model consists of four components: embedding module, temporal self-attention module, graph neural network module and spatial self-attention module.

To keep the native information in the raw data, we first concatenates observed and missing data, then utilize a 1D convolutional layer to obtain a feature embedding $E^f \in \mathbb{R}^{T\times N\times d}$:

$$E^f = \text{CNN}(\mathbf{z}_0^c||\mathbf{z}_0^m) \tag{11}$$

In spatiotemporal sequences, data from the same node exhibit temporal heterogeneity, while data from different nodes show spatial heterogeneity. To capture this, the embedding module converts time, spatial, and diffusion step information into trainable embeddings. We encode time windows into a learnable embedding table $E_d^{tem} \in \mathbb{R}^{N_W\times d}$, where N_W and d denote the time window length and feature embedding dimension. A spatial embedding $E_d^{spa} \in \mathbb{R}^{N_S\times d}$ captures spatial heterogeneity, where N_S is the number of nodes. The final spatiotemporal representation is obtained

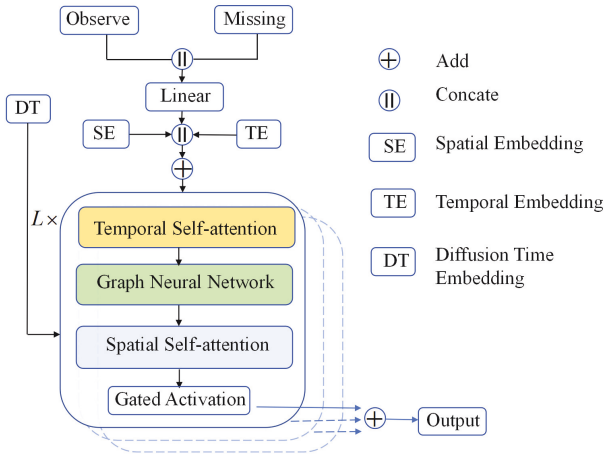

Figure 2: Architecture of denoising model.

by summing these embeddings:

$$Z^{emb} = E^f + E_d^{tem} + E_d^{spa} \tag{12}$$

Since the attention mechanism performs exceptionally well in modeling sequential data, we use self-attention to model the spatial and temporal dependencies of Z_t^m, which can be formulated as follows:

$$\begin{aligned} \text{Attn}(Z^{emb}) &= \text{softmax}\left(QK^T/\sqrt{d}\right)V, \\ Q = W_Q Z^{emb}, \quad K &= W_K Z^{emb}, \quad V = W_V Z^{emb} \end{aligned} \tag{13}$$

where W_Q, W_K and $W_V \in \mathbb{R}^{d\times d}$ are learnable projection parameters. To capture the relationships between nodes in spatiotemporal data, we employ graph neural networks to further model spatial dependencies which can be formalized as follows:

$$Z^{emb} = \text{GNN}(Z^{emb}, A) \tag{14}$$

where A denotes adjacency matrix.

Experiments

This section introduces the experimental datasets, comparison methods, evaluation metrics, and experimental settings. We evaluate the RDPI framework through extensive spatiotemporal imputation experiments to address the following questions:

RQ 1: Does RDPI outperform existing methods?

RQ 2: Does imputation benefit from conditional observational data and the initial model?

RQ 3: Can RDPI maintain effective imputation performance even when all nodes are missing?

Datasets and Experimental Settings

Datasets. We conducted comparative experiments on four real datasets: PEMS-BAY (Li et al. 2017), METR-LA (Li et al. 2017), AQI (Yi et al. 2016)and AQI36 (Yi et al. 2016). PEMS-BAY contains traffic speed data from 207 nodes on California highways over four months. METR-LA collected traffic speed data from 325 nodes on Los Angeles highways

over six months. Both PEMS-BAY and METR-LA are sampled every 5 minutes. AQI contains hourly sampled air quality data from 437 observation nodes in 43 Chinese cities over twelve months. AQI36 contains 36 observation nodes from AQI in Beijing over twelve months.

Training Strategies. In air quality datasets, missing data and their corresponding true data are already annotated in the dataset. As in previous works (Cini, Marisca, and Alippi 2021; Liu et al. 2023), we employed two different training scenarios: (1) **In-sample**, where model was trained according to the specified missing positions in the dataset; (2) **Out-of-sample**, where the training set combined the missing data specified with the observed data, and new missing positions were randomly generated. In traffic datasets, missing data is sparse and lacks ground truth values. Therefore, in addition to the actual observed data, we employed a method to manually design target data imputation using random masking. Similar to prior studies, the imputation targets were divided into two scenarios: (1) **Point missing**, where 25% of observed data was randomly masked; (2) **Block missing**, where initially 5% of observed data was masked, and then for each node, data within 1 to 4 hours was randomly masked at a probability of 0.15%.

Baseline

In our study, we conducted a comparative analysis using various baseline methods, including statistic methods (MEAN, KNN, MICE (White, Royston, and Wood 2011), and VAR), deep autoregressive methods (GRIN (Cini, Marisca, and Alippi 2021), BRITS (Cao et al. 2018), rGAIN (Yoon, Jordon, and Schaar 2018), and GP-VAE (Fortuin et al. 2020)), and diffusion-based models (CSDI (Tashiro et al. 2021) and MIDM (Wang et al. 2023)).

Metrics

We apply three evaluation metrics to measure the performance of spatiotemporal imputation: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean relative error (MRE).

Main Results (RQ1)

We evaluated the spatiotemporal imputation performance of RDPI and compared it with baselines. Table 1 and 2 show the results on the AQI36, AQI, PEMS-BAY, and METR-LA datasets. RDPI consistently outperforms all baselines across all metrics, particularly in In-sample mode, where MSE is reduced by 34% for AQI36 and 50% for AQI. It also performs exceptionally well on traffic datasets, demonstrating robustness in handling varying spatial dependencies. This improvement is mainly due to the proposed imputation framework, which quickly generates rough imputations through the initial model and refines them by efficiently estimating residual distributions, reducing computational cost and enhancing performance.

Although RDPI achieves the best MAE and MSE performance, the improvement in MRE is less significant. This is likely due to the probabilistic sampling in the diffusion model, which smooths the residuals and reduces their magnitude, benefiting MSE but not fully capturing the relative error emphasized by MRE.

Ablation Study (RQ2)

We conducted ablation experiments on the AQI36 and METR-LA datasets to evaluate the contributions of the conditional diffusion model, the two-stage framework, and training strategies. The following variants were compared:

- w/o cond-forw: Excluding observational conditions significantly reduced performance, highlighting the importance of observed-missing data relationships.
- w/o residual: The conditional diffusion model still performed reasonably well without the two-stage framework.
- w/o joint / w/o pre-train: Without joint training, residual overfitting led to poor results; pretraining stabilized the residuals and improved the diffusion model's performance.
- Predicting $\mathbf{x}_\theta$: Directly predicting data was less effective, as noise prediction proved more efficient.
- $-f_\theta(\mathbf{x}_0^c)$: Using $\|\mathbf{x}_0^m - f_\theta(\mathbf{x}_0^c)\|$ as the loss function degraded performance, destabilizing the initial model and impairing residual modeling.

Table 3 summarizes the results, leading to these observations:

- w/o cond-forw: Excluding observational conditions significantly worsens performance, highlighting the importance of leveraging observed-missing data relationships.
- w/o residual: The conditional diffusion model shows strong generative capability, achieving reasonable results even without a two-stage framework.
- w/o joint / w/o pre-train: Both highlight the influence of the initial model on final imputation. Without joint training, overfitting residuals lead to the worst results. Pretraining stabilizes residuals, improving diffusion model performance. Predicting $\mathbf{x}_\theta$: Predicting data directly is suboptimal due to low signal-to-noise ratios in residual inputs, making noise prediction more effective.
- $-f_\theta(\mathbf{x}_0^c)$: Performance degrades when the loss depends on $\|\mathbf{x}_0^m - f_\theta(\mathbf{x}_0^c)\|$, as optimizing the denoising model destabilizes the initial model's performance, impairing residual modeling.

Probabilistic Imputation (RQ3)

Figure 3 presents imputation examples from the AQI36 dataset, with red crosses representing observed values and blue circles for missing ground truths. After 50 imputations, the median line closely follows the original time series, and almost all missing data falls within the 5th and 95th percentiles, demonstrating the effectiveness of the proposed model.

We also addressed the Kriging problem, where data is missing at specific locations, such as entry points in traffic networks. To evaluate imputation performance, experiments

Dataset	Model	In-sample			Out-of-sample		
		MAE	MSE	MRE(%)	MAE	MSE	MRE(%)
AQI-36	MEAN	53.48±0.00	4578.08±00.00	76.77±0.00	53.48±0.00	4578.08±00.00	76.77±0.00
	KNN	30.21±0.00	2892.31±00.00	43.36±0.00	30.21±0.00	2892.21±00.00	43.36±0.00
	MICE	29.89±0.11	2575.53±07.67	42.90±0.15	30.37±0.09	2594.06 ±07.17	43.59±0.13
	VAR	13.16±0.21	513.90±12.39	18.89±0.31	15.64±0.08	833.46±13.85	22.02±0.11
	BRITS	12.24±0.26	495.94±43.56	17.57±0.38	14.50±0.35	662.36±65.16	20.41±0.50
	GRIN	10.51±0.28	371.47±17.38	15.09±0.40	12.08±0.47	523.14±57.17	17.00±0.67
	rGAIN	12.23±0.17	393.76±12.66	17.55±0.25	15.37±0.26	641.92±33.89	21.63±0.36
	GP-VAE	14.11±0.24	483.91±24.36	18.43±0.45	25.71±0.30	2589.53± 59.14	-
	CSDI	9.60±0.14	372.49±16.90	15.49±0.37	9.51±0.10	352.46±7.50	-
	MIDM	9.41±0.20	361.28±21.33	14.87±0.41	-	-	-
	RDPI	**7.98±0.24**	**238.25±13.22**	**11.67±0.35**	**9.14±0.03**	**306.40±21.33**	**13.45 ±0.02**
AQI	MEAN	39.60±0.00	3231.04±00.00	59.25±0.00	39.60±0.00	3231.04±00.00	59.25±0.00
	KNN	34.10±0.00	3471.14±00.00	51.02±0.00	34.10±0.00	3471.14±00.00	51.02±0.00
	MICE	26.39±0.13	1872.53±15.97	39.49±0.19	26.98±0.10	1930.92±10.08	40.37±0.15
	VAR	18.13±0.84	918.68±56.55	27.13±1.26	22.95±0.30	1402.84±52.63	33.99±0.44
	BRITS	17.24±0.13	924.34±18.26	25.79±0.20	20.21±0.22	1157.89±25.66	29.94±0.33
	GRIN	13.10±0.08	615.80±10.09	19.60±0.11	14.73±0.15	775.91±28.49	21.82±0.23
	rGAIN	17.69±0.17	861.66±17.49	26.48±0.25	21.78±0.50	1274.93±60.28	32.26±0.75
	GP-VAE	17.84±0.16	893.27±20.39	27.46±0.19	-	-	-
	CSDI	11.37±0.12	589.31±11.20	18.26±0.24	-	-	-
	MIDM	10.06±0.11	562.84±12.01	**16.87±0.19**	-	-	-
	RDPI	**9.10±0.33**	**266.81±13.65**	17.17±0.15	**11.22±0.21**	**388.63±0.22**	**21.16±0.02**

Table 1: Imputation results on AQI-36 and AQI. Performance averaged over 5 runs.

Dataset	Model	Block-missing			Point-missing		
		MAE	MSE	MRE(%)	MAE	MSE	MRE(%)
PEMS-BAY	MEAN	5.46±0.00	87.56± 0.00	8.75±0.00	5.42±0.00	86.59±0.00	8.67±0.00
	KNN	4.30±0.00	49.90±0.00	6.90±0.00	4.30±0.00	49.80±0.00	6.88±0.00
	MICE	2.94±0.02	28.28±0.37	4.71±0.03	3.09±0.02	31.43 ±0.41	4.95±0.02
	VAR	2.09±0.10	16.06±0.73	3.35±0.02	1.30±0.00	6.52±0.01	2.07±0.01
	BRITS	1.70±0.01	10.50±0.07	2.72±0.01	1.47±0.00	7.94±0.03	2.36±0.00
	GRIN	1.14±0.01	6.60±0.10	1.83±0.02	0.67±0.00	1.55±0.01	1.08±0.00
	rGAIN	2.18±0.01	13.96±0.20	3.50±0.02	1.88±0.02	10.37±0.20	3.01±0.04
	GP-VAE	2.39±0.03	14.81±0.15	4.32±0.02	1.92±0.01	12.43±0.08	3.67±0.02
	CSDI	1.16±0.01	7.02±0.09	1.96±0.01	0.83±0.00	1.79±0.00	1.42±0.00
	MIDM	1.03±0.01	5.83±0.11	1.77±0.02	0.60±0.00	1.54±0.02	0.93±0.00
	RDPI	**0.90±0.01**	**4.76±0.02**	**1.45 ±0.01**	**0.59±0.02**	**1.42±0.01**	**0.90±0.02**
METR-LA	MEAN	7.48±0.00	139.54±0.00	12.96±0.00	7.56±0.00	142.22±0.00	13.10±0.00
	KNN	7.79±0.00	124.61±0.00	13.49±0.00	7.88±0.00	129.29±0.00	13.65±0.00
	MICE	4.22±0.05	51.07±1.25	7.31±0.09	4.42±0.07	55.07±1.46	7.65±0.12
	VAR	3.11±0.08	28.00±0.76	5.38±0.13	2.69±0.00	21.10±0.02	4.66±0.00
	BRITS	2.34±0.01	17.00±0.14	4.05±0.01	2.34±0.00	16.46±0.05	4.05±0.00
	GRIN	2.03±0.00	13.26±0.05	3.52±0.01	1.91±0.00	10.41±0.003	3.30±0.00
	rGAIN	2.90±0.01	21.67±0.15	5.02±0.02	2.83±0.01	20.03±0.09	4.91±0.01
	GP-VAE	6.55±0.09	122.33±2.05	-	6.57±0.10	127.26±3.97	-
	CSDI	1.98±0.00	12.62±0.60	-	1.79±0.00	8.96±0.08	
	RDPI	**1.96±0.01**	**12.55±0.34**	**3.40±0.02**	**1.73±0.01**	**8.45±0.03**	**3.02±0.02**

Table 2: Imputation results on PEMS-BAY and METR-LA. Performance averaged over 5 runs.

Method	AQI36			METR-LA		
	In-sample			Block-missing		
	MAE	MSE	MRE(%)	MAE	MSE	MRE(%)
w/o cond-forw	9.25±0.32	310.44±10.22	12.45±4.35	2.05±0.01	14.73±0.03	4.74±0.03
w/o residual	8.31±4.35	280.57±10.03	12.23±0.03	1.97±0.01	13.53±0.03	4.02±0.03
w/o joint	9.20±0.01	263.25±0.03	12.83±0.03	2.04±0.01	15.01±0.02	4.97±0.01
w/o pre-train	8.87±0.44	295.03±26.88	12.52±0.32	2.12±0.01	16.97±0.02	4.21±0.01
predicting $\mathbf{x}_\theta$	8.29±0.32	**153.66±15.90**	11.69±0.24	1.97±0.04	13.21±0.01	3.97±0.01
$-f_\theta(\mathbf{x}_0^c)$	9.13±0.12	274.35±0.32	12.64±0.32	2.63±0.02	18.97±0.01	5.24±0.02
RDPI	**7.98±0.24**	238.25±13.22	**11.67±0.35**	**1.96±0.01**	**12.55±0.34**	**3.40±0.02**

Table 3: Ablation studies on AQI36 and METR-LA.

Model	Node-14			Node-31		
	MAE	MSE	MRE(%)	MAE	MSE	MRE(%)
GRIN	13.75	658.33	19.57	20.55	1151.35	39.72
RDPI	**9.50**	**316.70**	**13.52**	**15.28**	**705.41**	**29.53**

Table 4: Imputation result for node 14 and node 31 in AQI36.

on the AQI36 dataset were conducted using nodes with different levels of connectivity: node 14 (high connectivity) and node 31 (low connectivity). As illustrated in Figure 4, the imputed results closely align with the ground truth values. Quantitative results are provided in Table 4, showing that RDPI outperforms GRIN by reducing MAE, MSE, and MRE by 31%, 52%, and 31% for node 14, and by 26%, 39%, and 26% for node 31. These results highlight RDPI's effectiveness in leveraging spatial relationships for imputation. Assuming that the spatiotemporal data exhibit sufficient dependence on spatial geographic location, RDPI can potentially reconstruct time series data for locations within the study area where no sensors are deployed.

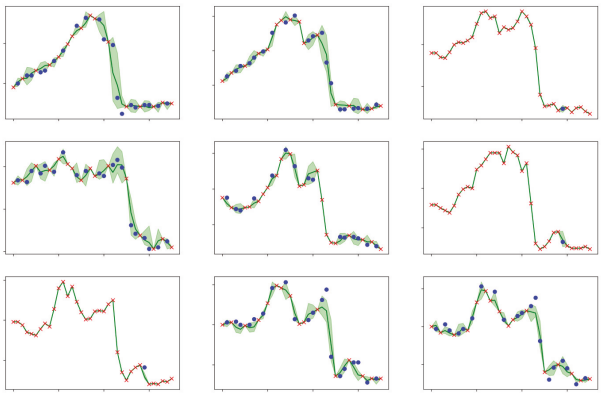

Figure 3: Imputation examples of RDPI on the AQI36 dataset. The horizontal axis represents time, and the vertical axis represents value.

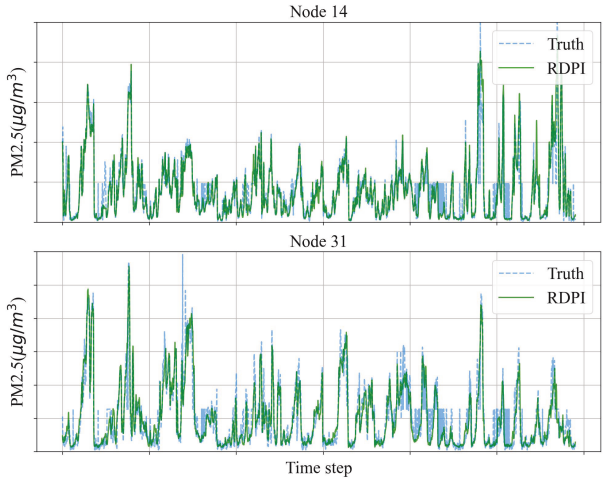

Figure 4: The imputation for unobserved sensors in AQI36. The blue dotted line represents the truth of the ground, and the green solid line represents the deterministic imputation result.

Conclusion

This paper presents a novel two-stage refined diffusion probabilistic imputation (RDPI) framework, for spatiotemporal data imputation. In the initial stage, a deterministic method is used to generate rough estimates of the missing data. In the refinement stage, we incorporate observed conditions into the forward process, with the residuals as the diffusion target, to derive a new conditional diffusion model that optimizes the initial imputation results. By increasing the utilization of observed data, RDPI generates more targeted imputation results. Extensive experimental results demonstrate that the RDPI method significantly improves the performance of spatiotemporal data imputation, while effectively reducing sampling computation costs and improving imputation efficiency. Future work will focus on decoupling spatiotemporal data and exploring conditional diffusion models for graph structures and multivariate time series, further enhancing the imputation and prediction performance of the existing framework in capturing spatiotemporal dependencies.

References

Alcaraz, J. M. L.; and Strodthoff, N. 2022. Diffusion-based time series imputation and forecasting with structured state space models. *arXiv preprint arXiv:2208.09399*.

Box, G. E.; Jenkins, G. M.; Reinsel, G. C.; and Ljung, G. M. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Cao, W.; Wang, D.; Li, J.; Zhou, H.; Li, L.; and Li, Y. 2018. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31.

Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; and Liu, Y. 2018. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1): 6085.

Cini, A.; Marisca, I.; and Alippi, C. 2021. Filling the G_ap_s: Multivariate Time Series Imputation by Graph Neural Networks. In *International Conference on Learning Representations*.

Du, W.; Côté, D.; and Liu, Y. 2023. Saits: Self-attention-based imputation for time series. *Expert Systems with Applications*, 219: 119619.

Fortuin, V.; Baranchuk, D.; Rätsch, G.; and Mandt, S. 2020. Gp-vae: Deep probabilistic time series imputation. In *International conference on artificial intelligence and statistics*, 1651–1661. PMLR.

Han, X.; Zheng, H.; and Zhou, M. 2022. Card: Classification and regression diffusion models. *Advances in Neural Information Processing Systems*, 35: 18100–18115.

Hastie, T.; Tibshirani, R.; Friedman, J. H.; and Friedman, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Hu, J.; Liang, Y.; Fan, Z.; Chen, H.; Zheng, Y.; and Zimmermann, R. 2023. Graph Neural Processes for Spatio-Temporal Extrapolation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 752–763.

Kim, S.; Kim, H.; Yun, E.; Lee, H.; Lee, J.; and Lee, J. 2023. Probabilistic imputation for time-series classification with missing data. In *International Conference on Machine Learning*, 16654–16667. PMLR.

Kong, Z.; Ping, W.; Huang, J.; Zhao, K.; and Catanzaro, B. 2021. DiffWave: A Versatile Diffusion Model for Audio Synthesis. In *International Conference on Learning Representations*.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.

Liu, M.; Huang, H.; Feng, H.; Sun, L.; Du, B.; and Fu, Y. 2023. Pristi: A conditional diffusion framework for spatiotemporal imputation. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 1927–1939. IEEE.

Liu, Y.; Yu, R.; Zheng, S.; Zhan, E.; and Yue, Y. 2019. Naomi: Non-autoregressive multiresolution sequence imputation. *Advances in neural information processing systems*, 32.

Liu, Z.; Zhao, X.; and Song, Y. 2024. RDPI: A Refine Diffusion Probability Generation Method for Spatiotemporal Data Imputation. *arXiv preprint arXiv:2412.12642*.

Luo, Y.; Cai, X.; Zhang, Y.; Xu, J.; et al. 2018. Multivariate time series imputation with generative adversarial networks. *Advances in neural information processing systems*, 31.

Marisca, I.; Cini, A.; and Alippi, C. 2022. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Advances in Neural Information Processing Systems*, 35: 32069–32082.

Mattei, P.-A.; and Frellsen, J. 2019. MIWAE: Deep generative modelling and imputation of incomplete data sets. In *International conference on machine learning*, 4413–4423. PMLR.

Miao, X.; Wu, Y.; Wang, J.; Gao, Y.; Mao, X.; and Yin, J. 2021. Generative semi-supervised learning for multivariate time series imputation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 8983–8991.

Mulyadi, A. W.; Jun, E.; and Suk, H.-I. 2021. Uncertainty-aware variational-recurrent imputation network for clinical time series. *IEEE Transactions on Cybernetics*, 52(9): 9684–9694.

Shan, S.; Li, Y.; and Oliva, J. B. 2023. Nrtsi: Non-recurrent time series imputation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Song, J.; Meng, C.; and Ermon, S. 2021. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*.

Tashiro, Y.; Song, J.; Song, Y.; and Ermon, S. 2021. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems*, 34: 24804–24816.

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.

Wang, X.; Zhang, H.; Wang, P.; Zhang, Y.; Wang, B.; Zhou, Z.; and Wang, Y. 2023. An observed value consistent diffusion model for imputing missing values in multivariate time series. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2409–2418.

Whang, J.; Delbracio, M.; Talebi, H.; Saharia, C.; Dimakis, A. G.; and Milanfar, P. 2022. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16293–16303.

White, I. R.; Royston, P.; and Wood, A. M. 2011. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, 30(4): 377–399.

Wu, Y.; Zhuang, D.; Labbe, A.; and Sun, L. 2021. Inductive graph neural networks for spatiotemporal kriging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4478–4485.

Yi, X.; Zheng, Y.; Zhang, J.; and Li, T. 2016. ST-MVL: Filling missing values in geo-sensory time series data. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*.

Yoon, J.; Jordon, J.; and Schaar, M. 2018. Gain: Missing data imputation using generative adversarial nets. In *International conference on machine learning*, 5689–5698. PMLR.

Yoon, J.; Zame, W. R.; and van der Schaar, M. 2018. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Transactions on Biomedical Engineering*, 66(5): 1477–1490.

You, J.; Ma, X.; Ding, Y.; Kochenderfer, M. J.; and Leskovec, J. 2020. Handling missing data with graph representation learning. *Advances in Neural Information Processing Systems*, 33: 19075–19087.

Zivot, E.; and Wang, J. 2006. Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®*, 385–429.