FMT-I: Flexible Mass Transport for Noise-Resilient Missing Data Imputation

Hao Wang, Licheng Pan, Yuan Lu, Zhengnan Li, Shuting He, Xiaoyu Jiang, *Member, IEEE*, Zhichao Chen, Xinggao Liu

***Abstract*—The vulnerability of existing missing data imputation methods to noisy observations often limits their utility in real-world industrial applications. To address this, we introduce a noise-resilient approach to missing data imputation based on optimal transport theory. Specifically, we first present the Flexible Mass Transport (FMT) discrepancy, which quantifies the discrepancy between two distributions as the minimal cost to match samples between them. Unlike standard transport problems, FMT allows for matching a flexible mass of each sample, thereby aligning reliable samples while excluding noisy ones. Subsequently, we propose the FMT for Imputation (FMT-I) framework, which iteratively minimizes the FMT discrepancy between arbitrary mini-batch samples and utilizes the resulting gradients to update imputed values. This process ensures the consistency of statistical properties between the imputed data and the entire dataset, thereby producing feasible imputations. Experiments on real-world industrial datasets confirm that FMT-I exhibits strong noise resilience and outperforms state-of-the-art imputation methods, offering a reliable solution for missing data imputation problems in complex industrial contexts.**

***Index Terms*—missing value imputation, industrial data analysis, noise interference.**

I. Introduction

MISSING data is pervasive in industrial data analytics, frequently encountered in domains such as process monitoring [1, 2], inferential sensing [3, 4], and quality control [5, 6]. For example, temperature sensors in steel plants may fail due to overheating, leading to missing temperature records that obscure hazard warnings [7]; vibration sensors on machinery may disconnect due to electromagnetic interference, leading to missing records that obscure fault detection [8]. *Consequently, missing data constitutes a significant obstacle to industrial data analytics, compromising the security and reliability of industrial process automation.*

This work was supported by the China Postdoctoral Science Foundation under Grant Number 2025M781449 (Corresponding Author: Zhichao Chen).

Hao Wang and Zhichao Chen are with the School of Intelligence Science and Technology, Peking University, Beijing 100871, China. They are also with the College of Control Sceince and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: haohaow@zju.edu.cn, 12032042@zju.edu.cn).

Licheng Pan and Xinggao Liu are with the College of Control Sceince and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: 22132045@zju.edu.cn, lxg@zju.edu.cn).

Yuan Lu is with the Xiaohongshu Inc, Shanghai 200003, China (e-mail: luyuan2@xiaohongshu.com).

Zhengnan Li is with the School of Data Science, The Chinese University of Hong Kong, Shenzhen 518172, China (e-mail: zhengnanli@link.cuhk.edu.cn).

Shuting He is with the MoE Key Laboratory of Interdisciplinary Research of Computation and Economics, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: shuting.he@sufe.edu.cn).

Xiaoyu Jiang is with Hangzhou International Innovation Institute, Beihang University, Hangzhou 311115, China (e-mail: jiangxiaoyu@buaa.edu.cn).

To address this issue, diverse missing data imputation (MDI) methods have been developed, generally categorized into discriminative and generative approaches [9]. Early simple discriminative methods employ statistical measures (e.g., mean, median, or mode) to impute missing values; however, they struggle to capture intricate feature dependencies and exhibit limited performance [10]. Iterative discriminative methods improve accuracy by leveraging machine learning models to predict missing features, yet they introduce challenges in model selection and suffer from limited model capacity for modeling intricate feature dependencies [11]. Generative methods conceptualize MDI as a data generation process, which allows for modeling intricate feature dependencies with advanced deep generative models. Nevertheless, they often necessitate meticulous hyperparameter tuning and ample high-quality training data [12–14]. Consequently, each class of imputation methods exhibits distinct strengths and limitations, rendering them suitable for different application scenarios.

A unique and critical challenge for applying MDI to industrial data is the ubiquitous presence of noise. In industry, sensor measurements are routinely affected by hardware malfunctions, human errors, or abrupt operational changes [15, 16]. For instance, pressure sensors may produce inaccurate readings under extreme stress, while manual recordings can be false due to misreadings [17]. These factors lead to noisy observations that invalidate most MDI methods. For example, the iterative discriminative methods [11, 18] and generative methods [12, 14] require training on observed data, where noisy samples poison the learning process and reduce imputation performance. *Therefore, there is a critical need for developing noise-resilient MDI methods to perform reliable imputation in noisy industrial data.*

To address this gap, this study frames MDI as a distributional discrepancy minimization problem and employs a noise-resilient optimal transport approach. First, we formulate the Flexible Mass Transport (FMT) problem to quantify the discrepancy between distributions. Unlike standard optimal transport, FMT enables the selective matching of reliable samples while excluding noisy ones, thereby yielding a noise-resilient discrepancy measure. The FMT problem can be efficiently solved using iterative matrix-vector projections, enabling GPU acceleration and scalability to large datasets. Subsequently, we propose the FMT for Imputation (FMT-I) framework, which iteratively minimizes the FMT discrepancy between randomly sampled mini-batches. This optimization ensures that the statistical properties of the imputed values align with the entire dataset, generating statistically feasible imputations.

The selective matching mechanism of FMT makes FMT-I notably resilient to noise, positioning it as an ideal solution for MDI in complex industrial contexts.

Contributions. The key contributions of this study are summarized as follows. ❶ **We investigate the problem of MDI with noise interference**, which is a challenging yet widespread scenario in industrial MDI applications. ❷ **We develop a novel imputation method with noise resilience**, which enhances noise-resilience through a selective matching mechanism within the optimal transport framework. ❸ **We validate the efficacy of the proposed method on publicly available industrial datasets.** Moreover, our empirical studies under noisy conditions serve as a valuable benchmark for both researchers and industry professionals, shedding light on the utility of existing imputation strategies in industrial contexts with noise interference.

Organization. Section II introduces existing MDI models, classifying them into two paradigms and analyzing their respective strengths and weaknesses. Section III encapsulates the technical background and preliminaries to facilitate understanding our proposed approach. Section IV details the computational workflow of the proposed FMT-I framework. Section V applies FMT-I to a specific industrial case study—debutanizer column—and empirically assesses its imputation performance. Finally, we summarize our conclusions, limitations, and outline potential avenues for future research.

II. Related works

The ubiquity of missing data drives the development of MDI techniques in various fields such as traffic [19, 20], signal processing [21], and bioinformatics [22, 23]. Existing approaches are generally categorized into two paradigms: discriminative and generative[9]. These paradigms are distinguished by their reliance on generative models: generative methods perform imputation based on generative models (e.g., diffusion models [12]), whereas discriminative methods do not rely on generative models [9].

The discriminative paradigm formulates MDI as a statistical inference problem. Early methods in this category employed simple statistical measures, such as the mean or median for imputation. While straightforward, these approaches often struggle to capture intricate feature dependencies, resulting in limited imputation accuracy [24, 25]. To address this, iterative methods were developed to model each missing feature as a function of observed features, a strategy pioneered by Imputation by Chained Equations (ICE) [11]. Subsequent works extend the capacity of ICE by integrating modern statistical models—such as neural networks [26, 27], Bayesian models [11] and decision trees [18]—as well as advanced training strategies such as multiple imputation [11], ensemble learning [18], and variable selection [28]. Meanwhile, the factorization-based methods treat the dataset as a matrix or tensor to be completed, aiming to recover latent factors that reconstruct the data via constraint optimization [29]. Subsequent extensions modify the optimization problem formulation to improve resilience against noise [30–32]. In general, discriminative methods are straightforward to implement and effective on small-scale datasets. However, their reliance on statistical models often limits their capacity to model highly nonlinear relationships and scale to large-scale datasets.

The generative paradigm formulates MDI as a conditional data generation problem, and employs advanced deep generative models to solve it, such as autoencoders [33–36], generative adversarial networks [14, 37, 38], and diffusion models [13, 19, 39]. Recently, these models have seen widespread adaptation in industrial scenarios. For example, autoencoder models have been augmented with position-encoding mechanisms to capture autocorrelation in industrial logs [40], Transformer encoders to reconstruct blackout data [41], and multiscale attention modules to extract multifaceted representation for imputation [42]. GANs have been enhanced with multidimensional attention generators for multimode batch processes [43], soft sensor modules to improve imputation accuracy on quality-related variables [44], and modified loss functions to stabilize adversarial training [45]. Diffusion models have been integrated with state-space models to learn the representation of industrial logs [46], accelerated sampling strategies to reduce imputation overhead [47], and double-weighted Transformer encoders to depict evolving working conditions [48]. By utilizing capable generative models rather than statistical models, generative methods excel at capturing complex, nonlinear relationships and scaling to large datasets. However, these advantages are counterbalanced by practical challenges: they often necessitate meticulous hyperparameter tuning and ample high-quality training data, which are often unavailable in industrial settings.

More recently, discrepancy-based methods have emerged as a promising subset of discriminative approaches. These methods iteratively minimize the distributional discrepancy between randomly sampled data batches [7, 49, 50], exploiting the statistical property that the empirical distributions of any two random batches from i.i.d. data should be similar. Unlike existing methods, they are neither constrained by parametric statistical models nor reliant on extensive hyperparameter tuning. However, their performance is notably susceptible to sample noise. Our work directly addresses this limitation, thereby extending the applicability of discrepancy-based imputation to industrial scenarios characterized by noisy observations.

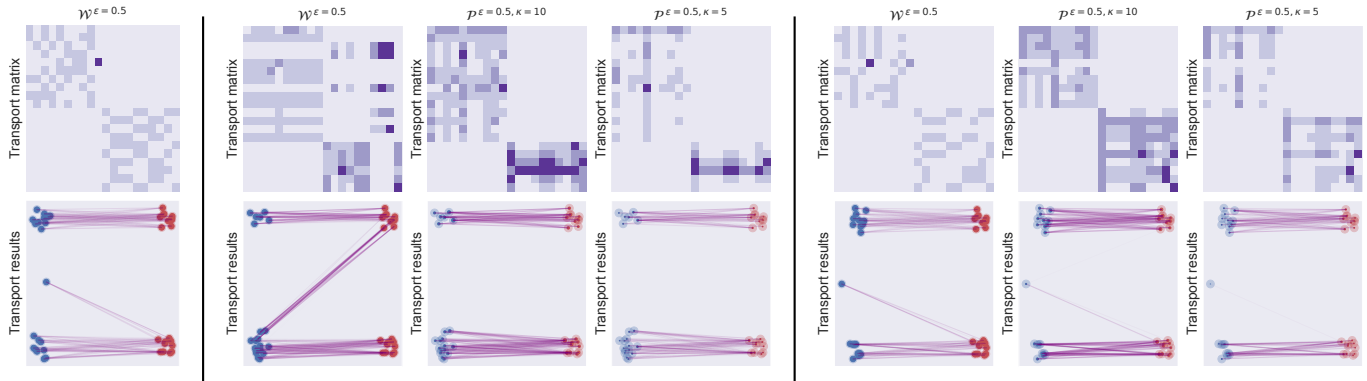
III. Preliminaries

A. Problem formulation

Suppose $\mathbf{X}^{(\mathrm{id})} \in \mathbb{R}^{\mathrm{N}\times\mathrm{D}}$ is the ideally complete industrial data matrix with N samples and D features. The presence of missing entries in $\mathbf{X}^{(\mathrm{id})}$ is indicated by a binary matrix $\mathbf{M} \in \{0,1\}^{\mathrm{N}\times\mathrm{D}}$, where each entry $\mathbf{M}_{n,d}$ is set to 1 if the corresponding entry $\mathbf{X}^{\mathrm{id}}_{n,d}$ is missing, and 0 otherwise. Consequently, the observed dataset $\mathbf{X}^{(\mathrm{obs})}$ can be derived using the Hadamard product:

$$\mathbf{X}^{(\mathrm{obs})} := \mathbf{X}^{(\mathrm{id})} \odot (1 - \mathbf{M}) + \mathrm{nan} \odot \mathbf{M}. \tag{1}$$

The goal of MDI is to recover the missing entries by constructing an imputation matrix $\mathbf{X}^{(\mathrm{imp})} \in \mathbb{R}^{\mathrm{N}\times\mathrm{D}}$.



(a) Ideal case (b) Mode imbalance case (c) Outlier case

Fig. 1. Visualization of the case study. The toy dataset includes two batches of samples, each distinguished by the color of the samples. The dataset contains two modes, differentiated by the vertical positioning of the samples.

B. Optimal transport

Optimal Transport (OT) is a mathematical framework designed to quantify the discrepancy between two distributions by identifying the minimum cost required to transform one distribution into the other. Originally proposed by Monge [51], the formulation involved finding an optimal mapping between two continuous distributions, which posed challenges related to the existence and uniqueness of solutions. Addressing these issues, Kantorovich [52] proposed a more computationally feasible approach, defined as follows:

Definition III.1. *Consider empirical distributions $\alpha = \alpha_{1:\mathrm{n}}$ and $\beta = \beta_{1:\mathrm{m}}$, each with n and m samples, respectively. The Kantorovich problem seeks a feasible plan $\mathbf{P} \in \mathbb{R}_{+}^{\mathrm{n}\times\mathrm{m}}$ to transport α to β at the minimum possible cost:*

$$\mathcal{W}^{\varepsilon}(\alpha,\beta) := \min_{\mathbf{P}\in\Pi(\alpha,\beta)} \langle \mathbf{D},\mathbf{P}\rangle + \varepsilon \mathrm{H}(\mathbf{P}),$$
$$\Pi(\alpha,\beta) := \begin{cases} \mathbf{P}_{i,1} + ... + \mathbf{P}_{i,\mathrm{m}} = \mathbf{a}_i, i = 1, ..., \mathrm{n}, \\ \mathbf{P}_{1,j} + ... + \mathbf{P}_{\mathrm{n},j} = \mathbf{b}_j, j = 1, ..., \mathrm{m}, \\ \mathbf{P}_{i,j} \geq 0, i = 1, ..., \mathrm{n}, j = 1, ..., \mathrm{m}, \end{cases} \tag{2}$$

where $\mathcal{W}^{\varepsilon}(\alpha,\beta)$ denotes the minimum transport cost; $\mathbf{D} \in \mathbb{R}_{+}^{\mathrm{n}\times\mathrm{m}}$ represents the pairwise distances calculated as $\mathbf{D}_{i,j} = \|\alpha_i - \beta_j\|_2$; $\mathbf{a} = [\mathbf{a}_1, \ldots, \mathbf{a}_{\mathrm{n}}]$ and $\mathbf{b} = [\mathbf{b}_1, \ldots, \mathbf{b}_{\mathrm{m}}]$ are the masses of samples in α and β, respectively; Π defines the set of constraints; H is the entropy regularization term, with ε indicating its strength.

The formulation above is a quadratic programming problem solvable via convex optimization techniques [53].

IV. Methodology

A. Motivation

Industrial sensor data is frequently contaminated by noise arising from hardware degradation (e.g., sensor degradation [15]) or human error (e.g., misreadings), causing observations to deviate from ground truths. However, **existing MDI methods struggle to handle such noisy observations.** (i) Methods based on global statistics (e.g., mean, median) are vulnerable to noisy observations; a single outlier can substantially skew the statistics and bias imputations across all missing entries. (ii) Iterative discriminative methods [11, 18] and generative methods [12, 14] are vulnerable to noisy observations; they require training on observed data, where noisy samples inevitably poison the learning process. Factorization-based methods [30, 31] represent an attempt to improve robustness but are constrained by rigid assumptions (e.g., low-rank structure) and limited model capacity, which exhibits limited performance in complex, large-scale datasets.

Therefore, developing an effective MDI strategy resilient to noisy observations remains a critical yet unresolved challenge. To address this, we propose FMT-I, a novel framework leveraging generalized OT to robustly accommodate noisy observations. Specifically, we first formulate the FMT discrepancy by relaxing the marginal constraints of the classical OT problem (2) with soft penalties, demonstrating its noise resilience on illustrative datasets (Section IV-B). Subsequently, we present the solution algorithm to compute FMT discrepancy (Section IV-C). Finally, we delineate the workflow of the FMT-I framework (Section IV-D).

B. Flexible mass transport

The standard OT problem in (2) is vulnerable to noisy samples due to its constraint set $\Pi(\alpha,\beta)$. Specifically, the constraints require matching all masses of each sample, mandating that the sum of the rows and columns of $\mathbf{P}$ equals the mass vectors $\mathbf{a}$ and $\mathbf{b}$, respectively. Consequently, when a noisy sample δ_z is present in α, the constraint compels a match between δ_z and normal samples in β, distorting the intended matching strategy and yielding an imprecise discrepancy measurement. This vulnerability is justified in Lemma IV.1, where the OT discrepancy increases as the outlier's deviation from the typical elements of β increases.

Lemma IV.1. *For two probability distributions α and β, suppose that $\tilde{\alpha} = \zeta\delta_z + (1-\zeta)\alpha$ is a distribution perturbed by a Dirac outlier at z with relative mass ζ. Abbreviate $\mathcal{W}^{\varepsilon=0}$ as $\mathcal{W}$; for an arbitrary sample y_* in the support of β, Fatras et al. [54] demonstrate that:*

$$\mathcal{W}(\tilde{\alpha},\beta) \geq (1-\zeta)\mathcal{W}(\alpha,\beta) + \zeta\left(D(z,y^*) - g(y^*) + \int g d\beta\right)$$

where $D(z, y^*)$ is the deviation of δ_z, (f, g) are the optimal dual potentials of $\mathcal{W}(\alpha, \beta)$.

To handle the noise vulnerability defect with OT, it is plausible to relax the marginal constraint to enable the creation and destruction of mass of each sample. To this end, inspired by the weak transport principles [55, 56], we suggest a Flexible Mass Transport (FMT) formulation, which replaces the marginal constraints with soft Kullback-Leibler (KL) penalties:

$$\mathrm{D}_{\mathrm{KL}}(\mathbf{P}_i||\mathbf{a}_i) = \left(\sum_{j=1}^{\mathrm{m}} \mathbf{P}_{i,j}\right) \log\left[\frac{\left(\sum_{j=1}^{\mathrm{m}} \mathbf{P}_{i,j}\right)}{\mathbf{a}_i}\right], \quad i = 1, ..., \mathrm{n},$$

$$\mathrm{D}_{\mathrm{KL}}(\mathbf{P}_j||\mathbf{b}_j) = \left(\sum_{i=1}^{\mathrm{n}} \mathbf{P}_{i,j}\right) \log\left[\frac{\left(\sum_{i=1}^{\mathrm{n}} \mathbf{P}_{i,j}\right)}{\mathbf{b}_j}\right], \quad j = 1, ..., \mathrm{m},$$

This replacement obviates the necessity for matching all masses across samples. Subsequently, without loss of generality [57, 58], assuming that the total mass of α and β is normalized to 1 and that each sample within these distributions has equal mass, we have $\mathbf{a} = \mathbf{1}_{\mathrm{n}}/\mathrm{n}$ and $\mathbf{b} = \mathbf{1}_{\mathrm{m}}/\mathrm{m}$, where $\mathbf{1}_{\mathrm{n}}$ and $\mathbf{1}_{\mathrm{m}}$ are vectors of ones with n and m samples, respectively. Defining simplex vectors $\Delta_{\mathrm{n}} = \mathbf{1}_{\mathrm{n}}/\mathrm{n}$ and $\Delta_{\mathrm{m}} = \mathbf{1}_{\mathrm{m}}/\mathrm{m}$, the FMT problem can be encapsulated in Definition IV.1.

Definition IV.1. *The FMT problem seeks a transport plan $\mathbf{P} \in \mathbb{R}_+^{\mathrm{n}\times\mathrm{m}}$ that transports α to β at the minimum cost:*

$$\begin{aligned} \mathcal{P}^{\varepsilon,\kappa}(\alpha, \beta) &:= \langle \mathbf{D}, \mathbf{P}^\star \rangle \\ \mathbf{P}^\star &:= \arg\min_{\mathbf{P}\geq 0} \langle \mathbf{D}, \mathbf{P} \rangle + \varepsilon \mathrm{H}(\mathbf{P}) \\ &\quad + \kappa(\mathrm{D}_{\mathrm{KL}}(\mathbf{P}\mathbf{1}_{\mathrm{m}}||\Delta_{\mathrm{m}}) + \mathrm{D}_{\mathrm{KL}}(\mathbf{P}^{\mathrm{T}}\mathbf{1}_{\mathrm{n}}||\Delta_{\mathrm{n}})) \end{aligned} \tag{3}$$

where $\mathcal{P}^{\varepsilon,\kappa}$ is the FMT discrepancy; κ is the mass-preservation strength, ε is the entropic regularization strength.

The relaxation of constraints in the FMT problem effectively enhances its resilience to noise. Considering two empirical distributions α and β sampled from the same industrial dataset, normal samples predominantly occupy overlapping regions between α and β, whereas noisy samples tend to appear in non-overlapping areas. The FMT problem minimizes transport costs by preferentially matching the mass of samples within these overlapping zones and ignoring those that are distantly located. It enables FMT to pair normal samples while disregarding outliers and noise, making it exceptionally suitable for handling industrial data characterized by noisy observations.

Case Study. To validate FMT's noise resilience, we conduct a comparative case study using a synthetic dataset, illustrated in Fig. 1. The dataset consists of two sample batches (in different colors), each initially containing two vertical modes. In this ideal, noise-free scenario, the basic OT correctly matches typical samples. We then introduce two common forms of noise interference to simulate industrial conditions. ❶ **Mode Imbalance:** We alter the proportion of samples in one mode relative to the other. The basic OT, bound by its constraint to match the entire mass, incorrectly pairs samples across different modes. In contrast, FMT's flexible matching mechanism prioritizes the matching of several samples, effectively avoiding erroneous matches (for $\kappa \leq 10$). ❷ **Outliers:** We introduce an outlier that does not overlap with the primary modes. The basic OT is forced to match this outlier, which distorts the entire transport plan and corrupts the matching of other samples. FMT mitigates this effect: as κ decreases, the influence of the outlier is progressively reduced until it is entirely excluded from the matching process (at $\kappa = 5$).

This case study empirically demonstrates that the flexible matching mechanism of the FMT problem confers resilience to noise interference. Such resilience is particularly advantageous for industrial applications, where noisy data are commonplace.

Algorithm 1 The solution algorithm to the FMT problem.

Input: $\mathbf{D}$: the pair-wise distance matrix.
Parameter: ε: the entropic regularization strength; κ: the mass-preservation strength; $\ell_{\max}$: the max iterations.
Output: $\mathcal{P}^{\varepsilon,\kappa}$: the FMT discrepancy.

$\mathbf{f} \leftarrow \mathbf{0}_n$, $\mathbf{g} \leftarrow \mathbf{0}_m$, $\ell \leftarrow 1$.
$\mathbf{a} \leftarrow \Delta_{\mathrm{n}}$, $\mathbf{b} \leftarrow \Delta_{\mathrm{m}}$
while $\ell < \ell_{\max}$ **do**
  $\mathbf{u} \leftarrow \exp(\mathbf{f}/\varepsilon)$, $\mathbf{v} \leftarrow \exp(\mathbf{g}/\varepsilon)$
  $\mathbf{P}^\ell \leftarrow \mathrm{diag}(\mathbf{u}) \exp(-\mathbf{D}/\varepsilon) \mathrm{diag}(\mathbf{v})$.
  $\mathbf{a}' \leftarrow \mathbf{P}\mathbf{1}_n$, $\mathbf{b}' \leftarrow \mathbf{P}^{\mathrm{T}}\mathbf{1}_m$.
  if $\ell // 2 = 0$ **then**
    $\mathbf{f} \leftarrow \left[\frac{\mathbf{f}}{\varepsilon} + \log(\mathbf{a}) - \log(\mathbf{a}')\right] \frac{\varepsilon\kappa}{\varepsilon+\kappa}$
  else
    $\mathbf{g} \leftarrow \left[\frac{\mathbf{g}}{\varepsilon} + \log(\mathbf{b}) - \log(\mathbf{b}')\right] \frac{\varepsilon\kappa}{\varepsilon+\kappa}$
  $\ell \leftarrow \ell + 1$.
$\mathbf{P}^* \leftarrow \mathrm{diag}(\mathbf{u}) \exp(-\mathbf{D}/\varepsilon) \mathrm{diag}(\mathbf{v})$.
$\mathcal{P}^{\varepsilon,\kappa} \leftarrow \langle \mathbf{D}, \mathbf{P}^* \rangle$.

Algorithm 2 The imputation workflow of FMT-I.

Input: $\mathbf{X}^{(\mathrm{obs})}$: the incomplete data; $\mathbf{M}$: the mask matrix.
Parameter: κ: the mass-preservation strength; ε the entropic regularization strength; $\ell_{\max}$: the max number of iterations to solve FMT; $\mathrm{T}_{\max}$: the max number of epochs; B: the batch size; η: the update rate.
Output: $\mathbf{X}^{(\mathrm{imp})}$: the imputed dataset.

$t \leftarrow 0$.
$\mathbf{X}^t \leftarrow \mathrm{Pre\text{-}Impute}(\mathbf{X}^{(\mathrm{obs})})$
while $t < \mathrm{T}_{\max}$ **do**
  $\alpha, \beta, \mathbf{M}^\alpha, \mathbf{M}^\beta \leftarrow \mathrm{Sample}(\mathbf{X}^t, \mathbf{M}; \mathrm{B})$.
  $\mathbf{D}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2, \quad 1 \leq i, j \leq \mathrm{B}$
  $\mathcal{P} \leftarrow \mathrm{Algorithm1}(\mathbf{D}; \varepsilon, \kappa, \ell_{\max})$.
  $\alpha' \leftarrow \alpha - \eta \nabla_\alpha \mathcal{P} \odot \mathbf{M}$.
  $\beta' \leftarrow \beta - \eta \nabla_\beta \mathcal{P} \odot \mathbf{M}$.
  $\mathbf{X}^{t+1} \leftarrow \mathrm{Update}(\alpha', \beta', \mathbf{X}^t)$.
  $t \leftarrow t + 1$.
  if $\left\|\mathbf{X}^t - \mathbf{X}^{t-1}\right\|_{\mathrm{F}} < 1e^{-4}$ **then**
    break.
$\mathbf{X}^{(\mathrm{imp})} \leftarrow \mathbf{X}^t$.

C. Solution to the FMT problem

In this section, we detail the algorithm to solve the FMT problem to obtain the FMT discrepancy $\mathcal{P}^{\varepsilon,\kappa}$ in (3). The FMT problem retains structural similarities with the standard OT problem, allowing it to be solved following the Sinkhorn paradigm in [55]. Specifically, we initialize the dual potentials

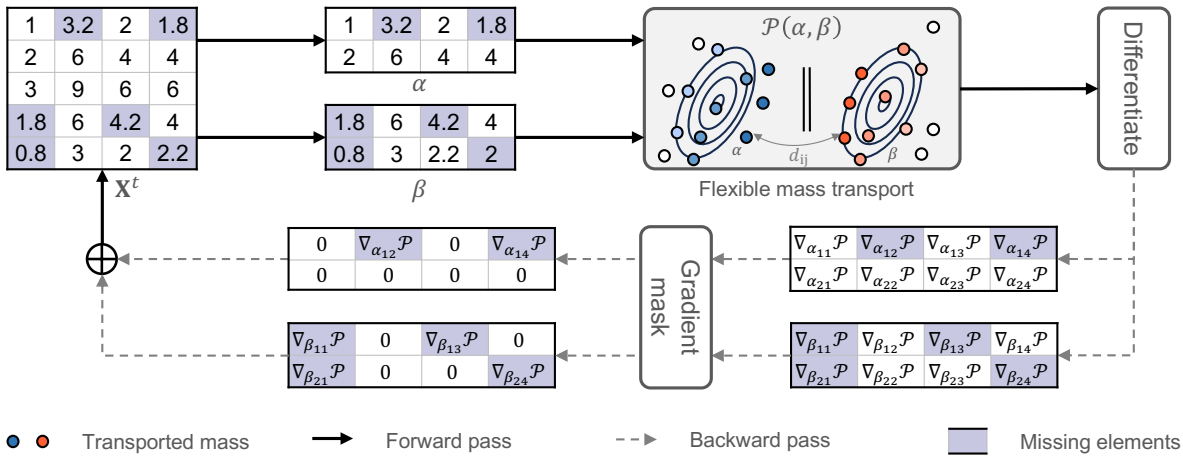


Fig. 2. Visualization of the workflow of FMT-I, where the dataset $\mathbf{X}$ contains 5 samples and 4 features with $\mathrm{p}_{\mathrm{miss}} = 0.3$. The batch size is set to 2.

($\mathbf{f}$ and $\mathbf{g}$) and sample masses ($\mathbf{a}$ and $\mathbf{b}$) with zero vectors and simplex vectors, respectively (steps 1-2). According to the first-order condition [57], the transport matrix can be represented with the dual potentials (steps 4-5) as follows:

$$\mathbf{P}_{ij} = \exp(\frac{\mathbf{f}_i + \mathbf{g}_j - \mathbf{D}_{ij}}{\varepsilon}). \tag{4}$$

The Fenchel-Legendre dual formulation of FMT (3) is:

$$\max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} -F^*(-\mathbf{f}) - G^*(-\mathbf{g}) - \varepsilon \sum_{i,j} \exp\left(\frac{-\mathbf{D}_{ij}}{\varepsilon}\right),$$

$$F^*(\mathbf{f}) = \max_{\mathbf{z}\in\mathbb{R}^n} \mathbf{z}^\top \mathbf{f} - \kappa \mathrm{D}_{\mathrm{KL}}(\mathbf{z}\|\mathbf{a}) = \kappa \left\langle e^{\mathbf{f}/\kappa}, \mathbf{a} \right\rangle - \mathbf{a}^\top \mathbf{1}_n,$$

$$G^*(\mathbf{g}) = \max_{\mathbf{z}\in\mathbb{R}^m} \mathbf{z}^\top \mathbf{g} - \kappa \mathrm{D}_{\mathrm{KL}}(\mathbf{z}\|\mathbf{b}) = \kappa \left\langle e^{\mathbf{g}/\kappa}, \mathbf{b} \right\rangle - \mathbf{b}^\top \mathbf{1}_m,$$

where $F^*(\cdot)$ and $G^*(\cdot)$ are Legendre transformations of the KL divergence. An equivalent minimization problem, disregarding irrelevant constants, is formulated as:

$$\min_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \varepsilon \sum_{i,j=1}^{n} \exp\left(\frac{\mathbf{Z}_{ij}}{\varepsilon}\right) + \kappa \left\langle e^{-\mathbf{f}/\kappa}, \mathbf{a} \right\rangle + \kappa \left\langle e^{-\mathbf{g}/\kappa}, \mathbf{b} \right\rangle.$$

According to the first-order condition, gradients at the minimum point should be zero. Thus, fixing $\mathbf{g}^\ell$ at the ℓ-th step, the updated $\mathbf{f}^{\ell+1}$ is determined by:

$$\exp\left(\frac{\mathbf{f}_i^{\ell+1}}{\varepsilon}\right) \sum_{j=1}^{n} \exp\left(\frac{\mathbf{g}_j^{\ell} - \mathbf{D}_{ij}}{\varepsilon}\right) = \exp\left(-\frac{\mathbf{f}_i^{\ell+1}}{\kappa}\right) \mathbf{a}_i,$$

Recalling that $\mathbf{a}' := \mathbf{P}\mathbf{1}_n$ with $\mathbf{P}_{ij} := \exp(\mathbf{f}_i^\ell + \mathbf{g}_j^\ell - \mathbf{D}_{ij})$, multiplying both sides by $\exp(\mathbf{f}_i^\ell/\varepsilon)$, we have:

$$\exp\left(\frac{\mathbf{f}_i^{\ell+1}}{\varepsilon}\right) \mathbf{a}_i' = \exp\left(\frac{\mathbf{f}_i^{\ell}}{\varepsilon}\right) \exp\left(-\frac{\mathbf{f}_i^{\ell+1}}{\kappa}\right) \mathbf{a}_i$$

Taking logarithms on both sides of the equation above yields:

$$\frac{\varepsilon + \kappa}{\varepsilon\kappa} \mathbf{f}_i^{\ell+1} = \left(\frac{\mathbf{f}_i^{\ell}}{\varepsilon}\right) + \log(\mathbf{a}_i) - \log(\mathbf{a}_i') \tag{5}$$

Similarly, with $\mathbf{f}$ fixed, the update for $\mathbf{g}^{\ell+1}$ is given by:

$$\frac{\varepsilon + \kappa}{\varepsilon\kappa} \mathbf{g}_j^{\ell+1} = \left(\frac{\mathbf{g}_j^{\ell}}{\varepsilon}\right) + \log(\mathbf{b}_j) - \log(\mathbf{b}_j') \tag{6}$$

where $\mathbf{b}' := \mathbf{P}^{\mathrm{T}}\mathbf{1}_m$. The equations in (5) and (6) constitute the iterative steps in Algorithm 1 (steps 6-11). After reaching the maximum number of iterations, the optimal transport matrix $\mathbf{P}^*$ is reconstructed from the dual potentials, and the FMT discrepancy is computed as the dot product of $\mathbf{P}^*$ and $\mathbf{D}$.

D. FMT-I: FMT for missing data imputation

In this section, we elaborate on the workflow of the FMT-I framework. The principal objective of FMT-I is to minimize the distributional discrepancy between arbitrary mini-batch samples, thereby ensuring that the statistical properties of the imputed values are consistent with those of the complete dataset. This discrepancy is quantified as FMT discrepancy in (3) to exploit its noise resilience. The workflow is outlined in Algorithm 2 with descriptions as follows.

The workflow starts with an initial imputation of the incomplete dataset $\mathbf{X}^{(\mathrm{obs})}$ using a basic KNN approach, wherein missing entries are estimated as the weighted average of the nearest neighbors based on observed features. This yields an initial imputation matrix $\mathbf{X}^{t=0}$ (steps 1-2). The imputed values are subsequently treated as learnable parameters, and their gradients are tracked throughout the optimization process.

The subsequent optimization consists of a forward phase and a backward phase, as illustrated in Fig. 2:

- In the forward phase, two mini-batches, denoted as $\alpha \in \mathbb{R}^{\mathrm{B}\times\mathrm{D}}$ and $\beta \in \mathbb{R}^{\mathrm{B}\times\mathrm{D}}$, are sampled from $\mathbf{X}^t$ (step 4). The FMT transport cost $\mathcal{P}^{\varepsilon,\kappa}$, which serves as a measure of the distributional discrepancy between α and β, is then computed in accordance with Algorithm 2 (steps 5-6).
- In the backward phase, the gradients of $\mathcal{P}^{\varepsilon,\kappa}$ with respect to α and β are obtained via automatic differentiation as:

$$\nabla_{\alpha_i}\mathcal{P} = \sum_{j=1}^{\mathrm{B}} \mathbf{P}_{i,j}^* \frac{\alpha_i - \beta_j}{\|\alpha_i - \beta_j\|_2},$$

$$\nabla_{\beta_j}\mathcal{P} = \sum_{i=1}^{\mathrm{B}} \mathbf{P}_{i,j}^* \frac{\beta_j - \alpha_i}{\|\alpha_i - \beta_j\|_2},$$

which are employed to iteratively update the imputed values corresponding to the missing entries in α and β at a learning

TABLE I
COMPARATIVE ANALYSIS OF FMT-I VERSUS BASELINE MODELS WITH A FIXED MISSING RATIO $\mathrm{p_{miss}} = 0.1$.

Metric	RMSE								MAE							
Model	U1	U2	U3	U4	U5	U6	U7	Y	U1	U2	U3	U4	U5	U6	U7	Y
Mean	0.095	0.043	0.205	0.117	0.116	0.167	0.172	0.164	0.069	0.025	0.155	0.093	0.084	0.144	0.150	0.116
Median	0.094	0.043	0.206	0.125	0.118	0.167	0.171	0.165	0.064	0.025	0.150	0.091	0.076	0.143	0.149	0.114
Mode	0.101	0.043	0.341	0.153	0.131	0.256	0.217	0.165	0.066	0.025	0.310	0.130	0.084	0.214	0.179	0.119
ICE	0.128	0.070	0.228	0.173	0.102	0.052	0.052	0.211	0.099	0.054	0.180	0.132	0.082	0.035	0.035	0.156
MICE	0.095	0.047	0.177	0.127	0.076	0.040	0.042	0.171	0.070	0.035	0.140	0.097	0.059	0.027	0.026	0.124
GAIN	0.096	0.046	0.226	0.133	0.090	0.086	0.086	0.159	0.070	0.031	0.190	0.102	0.066	0.071	0.070	0.110
Miss.D.	0.180	0.309	0.323	0.187	0.330	0.325	0.308	0.227	0.157	0.304	0.290	0.157	0.311	0.288	0.269	0.189
Miss.F.	0.076	0.041	0.173	0.102	0.075	0.078	0.080	0.145	0.051	0.025	0.134	0.075	0.055	0.050	0.051	0.101
Sinkhorn	0.092	0.042	0.198	0.113	0.103	0.143	0.147	0.162	0.066	0.025	0.150	0.088	0.073	0.124	0.128	0.114
TRPCA	0.076	0.082	0.166	0.104	0.077	0.045	0.046	0.144	0.051	0.064	0.132	0.076	0.056	0.034	0.032	0.099
SCPN	0.078	0.037	0.167	0.098	0.077	0.070	0.065	0.142	0.051	0.022	0.133	0.070	0.049	0.050	0.049	0.099
NewImp	0.083	0.028	0.184	0.104	0.083	0.070	0.071	0.157	0.055	0.021	0.135	0.073	0.052	0.057	0.058	0.105
SoftImp	0.094	0.095	0.188	0.116	0.093	0.086	0.083	0.176	0.071	0.074	0.149	0.084	0.067	0.068	0.066	0.124
MIWAE	0.094	0.046	0.203	0.127	0.099	0.104	0.111	0.177	0.065	0.027	0.156	0.089	0.065	0.058	0.062	0.125
Miracle	0.080	0.040	0.175	0.105	0.126	0.195	0.212	0.152	0.056	0.025	0.133	0.079	0.055	0.048	0.049	0.107
TDM	0.076	0.045	0.190	0.116	0.068	0.086	0.071	0.140	0.053	0.029	0.133	0.078	0.046	0.066	0.043	0.090
FMT-I	**0.061**	**0.026**	**0.120**	**0.074**	**0.038**	**0.030**	**0.029**	**0.103**	**0.037**	**0.019**	**0.082**	**0.048**	**0.027**	**0.023**	**0.022**	**0.063**

[1] *Kindly Note*: The presented results represent the average from several trials. **Bold** typeface highlights the top performance for each metric, while underlined text denotes the second-best results.

rate η (steps 7-8). The derivation of these gradients is provided in Appendix C. It is important to emphasize that observed (non-missing) values remain fixed throughout this process. To this end, we apply a gradient mask strategy, wherein the gradients corresponding to observed entries (i.e., those with $\mathbf{M}_{i,j} = 0$) are set to zero. The imputation matrix $\mathbf{X}^t$ is then updated by replacing the elements relevant to α and β with their updated counterparts, α' and β'.

Once the stopping criterion for the iterative procedure is satisfied (steps 3 and 11), the resulting matrix $\mathbf{X}^t$ is designated as the final imputation matrix $\mathbf{X}^{(\mathrm{imp})}$. The imputation quality is assessed using the modified Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), defined as follows:

$$\mathrm{MAE} := \frac{\sum_{n,d=1}^{\mathrm{N,D}} \left| \mathbf{X}_{n,d}^{(\mathrm{imp})} - \mathbf{X}_{n,d}^{(\mathrm{id})} \right| \odot \mathbf{M}_{n,d}}{\sum_{n,d=1}^{\mathrm{N,D}} \mathbf{M}_{n,d}},$$

$$\mathrm{RMSE} := \sqrt{\frac{\sum_{n,d=1}^{\mathrm{N,D}} \left\| \mathbf{X}_{n,d}^{(\mathrm{imp})} - \mathbf{X}_{n,d}^{(\mathrm{id})} \right\|_2^2 \odot \mathbf{M}_{n,d}}{\sum_{n,d=1}^{\mathrm{N,D}} \mathbf{M}_{n,d}}},$$

where $\mathbf{X}^{(\mathrm{id})}$ denotes the ideally complete dataset. The two metrics above are modified to exclude non-missing elements.

V. EMPIRICAL INVESTIGATION

A. Experimental setup

- **Datasets:** The experimental study is performed using the Debutanizer Column (DC), an exemplary industrial setting for desulfuration and naphtha separation. Precise measurement of butane concentration from the column's bottom flow (Y) is crucial for enhancing downstream production quality. To facilitate inferential sensor modeling, seven sensors (U1-U7) are strategically deployed. The DC dataset was chosen for its public availability, facilitating reproducibility, and large size relative to other public industrial datasets [59]. To simulate missing data scenarios, we employ a binary mask matrix using i.i.d. Bernoulli random variables with mean p_{miss}, corresponding to the missing ratio.
- **Baselines:** To situate the performance of FMT-I, we conduct a comparative analysis against a wide array of established methods. For discriminative methods, we select Mean, Median, Mode, ICE [11], MICE [11], Miss.F. [18], SoftImp [25], MIRACLE [27], Sinkhorn [50], TDM [49], TRPCA [30] and SCPN [31]; for generative methods, we select: GAIN [14], MIWAE [26], Miss.D [12] and NewImp [39].
- **Implementation details:** For consistency in experimental conditions, the batch size B is fixed at 512. The Adam optimizer, known for its effectiveness, is used for training, with a learning rate $\eta = 0.01$, following the specifications in [60]. To ensure convergence, we set $\mathrm{T_{max}} = 50$ and $\ell_{\max} = 1,000$. The key hyperparameters, namely κ and ε, are set to 1 and 0.05, respectively, determined by allocating 5% of the training data for validation and finetuning over the ranges $[5 \times 10^{-2}, 5]$ for κ and $[5 \times 10^{-4}, 5 \times 10^{-2}]$ for ε. The experiments are performed on a platform with two Intel(R) Xeon(R) Platinum 8383C CPUs @ 2.70GHz and a NVIDIA GeForce RTX 4090 GPU. All experiments are performed with multiple random seeds (1, 2, 42) to ensure the robustness of the results.

B. Overall performance

Table I presents the imputation results of FMT-I and baselines in the DC dataset, with a missing ratio $\mathrm{p_{miss}} = 0.1$. Key observations are summarized as follows:

- The baseline imputation methods demonstrate practical effectiveness in industrial contexts. Iterative methods, in particular, generally exhibit superior performance. For example, the Miss.F. method, which employs random forest as its base estimator, is especially well-suited for tabular data and achieves the lowest RMSE for three out of eight variables. Furthermore, discrepancy-based methods such as Sinkhorn and TDM, which incorporate standard OT, perform comparably to iterative methods, highlighting the potential of discrepancy-based approaches for the MDI task.
- FMT-I consistently achieves superior performance, yielding the lowest imputation errors across all variables. Notably,

TABLE II
COMPARATIVE ANALYSIS OF FMT-I VERSUS BASELINE MODELS UNDER DIVERSE MISSING RATIOS ($\mathrm{p_{miss}}$).

Missing ratio	$\mathrm{p_{miss}}=0.1$		$\mathrm{p_{miss}}=0.3$		$\mathrm{p_{miss}}=0.5$		$\mathrm{p_{miss}}=0.7$	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Mean	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$
Median	$0.103_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$
Mode	$0.142_{\pm 0.000}$	$0.197_{\pm 0.000}$	$0.140_{\pm 0.000}$	$0.195_{\pm 0.000}$	$0.151_{\pm 0.000}$	$0.208_{\pm 0.000}$	$0.153_{\pm 0.000}$	$0.216_{\pm 0.000}$
ICE	$0.095_{\pm 0.000}$	$0.142_{\pm 0.000}$	$0.108_{\pm 0.000}$	$0.156_{\pm 0.000}$	$0.121_{\pm 0.000}$	$0.170_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.191_{\pm 0.000}$
MICE	$0.071_{\pm 0.000}$	$0.109_{\pm 0.000}$	$0.081_{\pm 0.000}$	$0.121_{\pm 0.000}$	$0.090_{\pm 0.000}$	$\underline{0.131}_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.146_{\pm 0.000}$
GAIN	$0.088_{\pm 0.002}$	$0.126_{\pm 0.004}$	$0.092_{\pm 0.004}$	$0.129_{\pm 0.005}$	$0.140_{\pm 0.019}$	$0.185_{\pm 0.020}$	$0.338_{\pm 0.074}$	$0.400_{\pm 0.075}$
Miss.D.	$0.246_{\pm 0.081}$	$0.286_{\pm 0.094}$	$0.261_{\pm 0.031}$	$0.311_{\pm 0.035}$	$0.262_{\pm 0.020}$	$0.313_{\pm 0.022}$	$0.276_{\pm 0.020}$	$0.328_{\pm 0.022}$
Miss.F.	$0.067_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.079_{\pm 0.001}$	$0.118_{\pm 0.001}$	$\underline{0.089}_{\pm 0.003}$	$0.133_{\pm 0.003}$	$0.104_{\pm 0.002}$	$0.149_{\pm 0.002}$
Sinkhorn	$0.097_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.101_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.143_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.146_{\pm 0.000}$
TRPCA	$0.067_{\pm 0.009}$	$0.100_{\pm 0.013}$	$0.088_{\pm 0.004}$	$0.126_{\pm 0.006}$	$0.107_{\pm 0.008}$	$0.150_{\pm 0.013}$	$0.134_{\pm 0.022}$	$0.181_{\pm 0.031}$
SCPN	$\underline{0.065}_{\pm 0.000}$	$\underline{0.100}_{\pm 0.000}$	$0.081_{\pm 0.000}$	$0.118_{\pm 0.000}$	$0.092_{\pm 0.000}$	$0.133_{\pm 0.000}$	$\underline{0.098}_{\pm 0.000}$	$\underline{0.139}_{\pm 0.000}$
NewImp	$0.069_{\pm 0.000}$	$0.108_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.146_{\pm 0.000}$
SoftImp	$0.087_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.096_{\pm 0.000}$	$0.135_{\pm 0.000}$	$0.119_{\pm 0.000}$	$0.169_{\pm 0.000}$	$0.174_{\pm 0.000}$	$0.251_{\pm 0.000}$
MIWAE	$0.080_{\pm 0.000}$	$0.129_{\pm 0.003}$	$0.089_{\pm 0.008}$	$0.132_{\pm 0.006}$	$0.111_{\pm 0.000}$	$0.154_{\pm 0.001}$	$0.111_{\pm 0.000}$	$0.155_{\pm 0.000}$
Miracle	$0.069_{\pm 0.000}$	$0.148_{\pm 0.000}$	$\underline{0.074}_{\pm 0.000}$	$\underline{0.114}_{\pm 0.000}$	$0.090_{\pm 0.000}$	$0.132_{\pm 0.000}$	$0.128_{\pm 0.000}$	$0.164_{\pm 0.000}$
TDM	$0.067_{\pm 0.001}$	$0.109_{\pm 0.002}$	$0.083_{\pm 0.000}$	$0.126_{\pm 0.001}$	$0.094_{\pm 0.000}$	$0.136_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.001}$
FMT-I	$\mathbf{0.040}^{*}_{\pm 0.000}$	$\mathbf{0.068}^{*}_{\pm 0.001}$	$\mathbf{0.058}^{*}_{\pm 0.000}$	$\mathbf{0.097}^{*}_{\pm 0.000}$	$\mathbf{0.074}^{*}_{\pm 0.000}$	$\mathbf{0.118}^{*}_{\pm 0.000}$	$\mathbf{0.086}^{*}_{\pm 0.000}$	$\mathbf{0.130}^{*}_{\pm 0.000}$

Kindly Note: The results are presented as $\mathrm{mean}_{\pm \mathrm{std}}$. Bold typeface highlights the top performance for each metric, while underlined text denotes the second-best results. "*" marks the results where FMT-I significantly outperforms the second best method, with p-value<0.05 in a two-sample t-test.

FMT-I reduces the RMSE by more than 20% in seven out of eight cases relative to the less effective methods. For variables U2 and U5, the reductions are particularly pronounced, reaching 35% and 44.11%, respectively. This exceptional performance can be attributed to the strengths of discrepancy-based imputation and the capacity of FMT to accurately quantify distributional discrepancies.

C. Performance with different missing ratios

In this section, we evaluate the imputation capabilities of FMT-I relative to baselines across various missing ratios within the DC dataset. The results, summarized in Table II, present the average imputation errors for different features. FMT-I consistently surpasses all baselines in terms of performance, irrespective of the missing ratio. For example, at a missing ratio of $\mathrm{p_{miss}}=0.1$, FMT-I records the lowest mean absolute error (MAE) at 0.040, which represents a significant 40.29% reduction compared to the next most effective method (TDM). This advance continues at higher missing ratios demonstrating the robustness of FMT-I under diverse data missing ratios.

D. Performance with different noise ratios

In this section, we compare imputation performance under noisy observations. Specifically, a fraction $\mathrm{p_{noise}}$ of the observed data is perturbed by zero-mean Gaussian noise with standard deviation 0.2. This moderate noise level ensures the convergence of most imputation methods and enables meaningful comparison. The results are detailed in Table III. Key observations are summarized as follows:

- Increasing noise ratios generally degrades the performance of most methods, showcasing their vulnerability to noise interference. For example, the iterative method ICE's MAE increases from 0.111 at $\mathrm{p_{noise}}=0.01$ to 0.131 at $\mathrm{p_{noise}}=0.1$, marking an 18.1% deterioration. Notably, simple methods based on global statistics (Mean, Median) remain relatively stable, as the zero-mean noise does not affect these statistics.

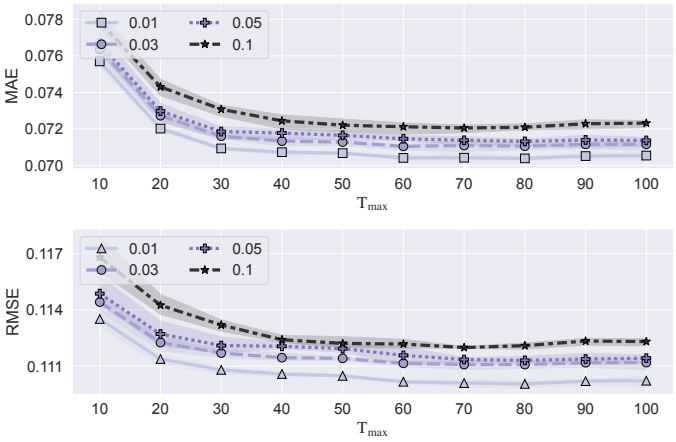


Fig. 3. Convergence analysis of FMT-I with a missing ratio $\mathrm{p_{miss}}=0.3$ and varying noisy ratios $\mathrm{p_{noise}}=0.01, 0.03, 0.05, 0.1$, shown with colored lines for means and shaded areas for 99.9% confidence intervals.

- Some methods explicitly modeling noisy observations demonstrate noise resilience. Specifically, SCPN's MAE increases only marginally from 0.082 to 0.083 as $\mathrm{p_{noise}}$ rises from 0.01 to 0.1. However, as a factorization-based approach, SCPN relies on a low-rank data structure, which may not be present in the DC dataset. Therefore, its performance is suboptimal, surpassed by other baseline methods.
- FMT-I consistently achieves the best performance across all noise ratios, with MAE changing minimally from 0.071 to 0.072 as $\mathrm{p_{noise}}$ increases from 0.01 to 0.1. Such noise resilience is attributed to FMT-I's flexible matching mechanism. This mechanism encourages the pairing of typical samples while disregarding noisy ones, making FMT-I exceptionally suitable for industrial scenarios plagued by high noise ratios.

E. Convergence analysis

In this section, we examine the convergence of FMT-I, focusing on its performance over various numbers of epochs

TABLE III
COMPARATIVE ANALYSIS OF FMT-I VERSUS BASELINE MODELS UNDER DIVERSE NOISE RATIOS ($\mathrm{p_{noise}}$).

Missing ratio	$\mathrm{p_{noise}}=0.01$		$\mathrm{p_{noise}}=0.03$		$\mathrm{p_{noise}}=0.05$		$\mathrm{p_{noise}}=0.1$	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Mean	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$
Median	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$
Mode	$0.140_{\pm 0.000}$	$0.195_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.211_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.211_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.211_{\pm 0.000}$
ICE	$0.111_{\pm 0.001}$	$0.159_{\pm 0.001}$	$0.117_{\pm 0.001}$	$0.164_{\pm 0.001}$	$0.121_{\pm 0.001}$	$0.168_{\pm 0.001}$	$0.131_{\pm 0.001}$	$0.177_{\pm 0.001}$
MICE	$0.083_{\pm 0.001}$	$0.124_{\pm 0.000}$	$0.086_{\pm 0.001}$	$0.127_{\pm 0.001}$	$0.088_{\pm 0.001}$	$0.129_{\pm 0.001}$	$0.094_{\pm 0.001}$	$0.134_{\pm 0.000}$
GAIN	$0.096_{\pm 0.011}$	$0.136_{\pm 0.011}$	$0.094_{\pm 0.010}$	$0.133_{\pm 0.013}$	$0.094_{\pm 0.009}$	$0.132_{\pm 0.010}$	$0.095_{\pm 0.008}$	$0.133_{\pm 0.010}$
Miss.D.	$0.261_{\pm 0.030}$	$0.311_{\pm 0.035}$	$0.262_{\pm 0.029}$	$0.312_{\pm 0.033}$	$0.262_{\pm 0.028}$	$0.312_{\pm 0.033}$	$0.262_{\pm 0.028}$	$0.311_{\pm 0.032}$
Miss.F	$0.079_{\pm 0.001}$	$0.119_{\pm 0.004}$	$0.078_{\pm 0.001}$	$0.118_{\pm 0.001}$	$0.079_{\pm 0.001}$	$0.120_{\pm 0.001}$	$0.080_{\pm 0.000}$	$0.119_{\pm 0.001}$
Sinkhorn	$0.102_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.101_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.139_{\pm 0.000}$
TRPCA	$0.089_{\pm 0.005}$	$0.127_{\pm 0.007}$	$0.093_{\pm 0.006}$	$0.129_{\pm 0.008}$	$0.095_{\pm 0.006}$	$0.131_{\pm 0.009}$	$0.098_{\pm 0.008}$	$0.135_{\pm 0.010}$
SCPN	$0.082_{\pm 0.000}$	$0.119_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.120_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.120_{\pm 0.000}$	$0.083_{\pm 0.000}$	$0.121_{\pm 0.001}$
NewImp	$0.082_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.083_{\pm 0.000}$	$0.124_{\pm 0.000}$
SoftImp	$0.097_{\pm 0.000}$	$0.136_{\pm 0.000}$	$0.098_{\pm 0.000}$	$0.137_{\pm 0.000}$	$0.098_{\pm 0.001}$	$0.138_{\pm 0.001}$	$0.101_{\pm 0.001}$	$0.142_{\pm 0.001}$
MIWAE	$0.091_{\pm 0.008}$	$0.134_{\pm 0.006}$	$0.096_{\pm 0.014}$	$0.139_{\pm 0.013}$	$0.101_{\pm 0.008}$	$0.143_{\pm 0.009}$	$0.109_{\pm 0.003}$	$0.152_{\pm 0.003}$
Miracle	$\underline{0.074}_{\pm 0.000}$	$\underline{0.114}_{\pm 0.000}$	$\underline{0.075}_{\pm 0.000}$	$\underline{0.115}_{\pm 0.000}$	$\underline{0.075}_{\pm 0.000}$	$\underline{0.116}_{\pm 0.000}$	$\underline{0.078}_{\pm 0.000}$	$\underline{0.118}_{\pm 0.001}$
TDM	$0.084_{\pm 0.001}$	$0.126_{\pm 0.001}$	$0.085_{\pm 0.001}$	$0.127_{\pm 0.001}$	$0.086_{\pm 0.001}$	$0.128_{\pm 0.001}$	$0.089_{\pm 0.000}$	$0.131_{\pm 0.001}$
FMT-I	$\mathbf{0.071}^{*}_{\pm 0.000}$	$\mathbf{0.111}^{*}_{\pm 0.000}$	$\mathbf{0.071}^{*}_{\pm 0.000}$	$\mathbf{0.111}^{*}_{\pm 0.000}$	$\mathbf{0.072}^{*}_{\pm 0.000}$	$\mathbf{0.112}^{*}_{\pm 0.000}$	$\mathbf{0.072}^{*}_{\pm 0.000}$	$\mathbf{0.112}^{*}_{\pm 0.000}$

Kindly Note: The results are presented as $\mathrm{mean}_{\pm \mathrm{std}}$. Bold typeface highlights the top performance for each metric, while underlined text denotes the second-best results. "*" marks the results where FMT-I significantly outperforms the second best method, with p-value< 0.05 in a two-sample t-test. We set $\mathrm{p_{miss}} = 0.1$; the results with different $\mathrm{p_{miss}}$ are reported in the Appendix.

and noise ratios ($\mathrm{p_{noise}}$), as illustrated in Fig. 3. Generally, increasing the maximum number of epochs ($\mathrm{T_{max}}$) tends to enhance the quality of imputation across different scenarios. However, performance improvements appear to plateau after reaching a certain number of epochs. Specifically, at a noise ratio of $\mathrm{p_{noise}} = 0.03$, the MAE and RMSE start at 0.0766 and 0.1145, respectively, with $\mathrm{T_{max}}$ set at 10 epochs. Extending $\mathrm{T_{max}}$ to 50 leads to a consistent reduction in MAE to 0.0717 and RMSE to 0.1116. Nevertheless, further increasing $\mathrm{T_{max}}$ to 100 results in only marginal changes and some fluctuations in both MAE and RMSE, suggesting that FMT-I reaches a convergence status beyond this point.

F. Noise-resilience analysis

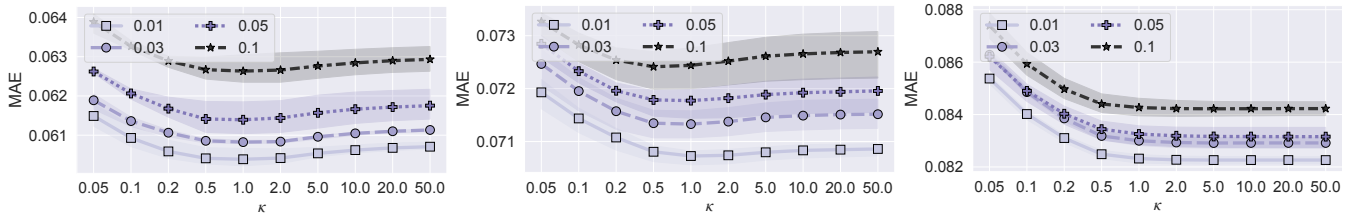
In this section, we analyze the utility of the selective matching mechanism in FMT-I to enhance noise resilience. To this end, we present the imputation performance given different mass preservation strengths (κ) and noisy ratios in Fig. 4. Key observations are summarized below:

- The utility of the selective matching mechanism varies with the missing ratios. For smaller missing ratios (e.g., $\mathrm{p_{miss}} = 0.1, 0.3$), reducing κ from ∞ to 1 consistently enhances imputation accuracy. However, excessively reducing κ towards zero impedes matching of typical samples and degrades imputation quality. In contrast, for larger missing ratios (e.g., $\mathrm{p_{miss}} = 0.5$), the benefits of reducing κ are less pronounced. This occurs because the selective matching mechanism, which disregards some samples from matching, leads to decreased sample efficiency—a critical factor in scenarios with high missing ratios. Nevertheless, FMT does not deteriorate performance in this case, alleviating concerns regarding its applicability in cases with high missing ratios.
- The optimal value of κ tends to decrease as noise ratios increase. For example, with a missing ratio of $\mathrm{p_{miss}} = 0.3$, the ideal κ progressively decreases from 1.0 at a $\mathrm{p_{noise}} = 0.01$ to 0.5 at a $\mathrm{p_{noise}} = 0.1$. As noise ratios rise, lowering κ further relaxes the mass preservation constraint, preventing forced matches between increased noisy samples and normal ones. Such flexibility underscores FMT-I's capacity to adapt to industrial environments with varying noise ratios.

G. Parameter sensitivity analysis

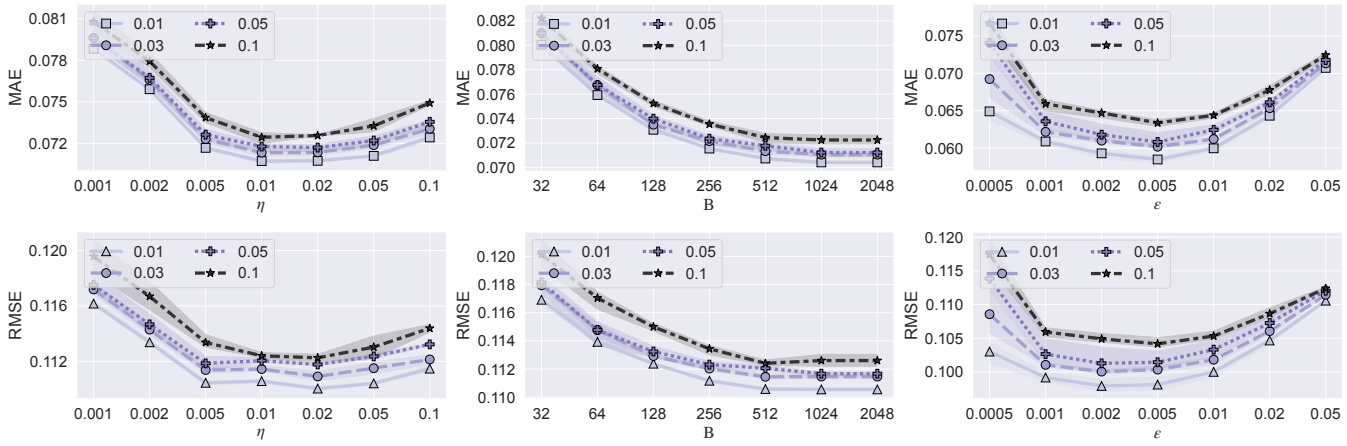
In this section, we examine the influence of critical hyperparameters on the performance of FMT-I in Fig. 5. Below are the key observations:

- The update rate (η) plays a pivotal role in controlling the volume of updates to the imputation matrix each epoch. As η is reduced from 0.1 to approximately 0.01, both MAE and RMSE decrease, indicating that a smaller η enhances update stability. However, further reduction of η to 0.001 results in increased MAE and RMSE, where the meaningful update direction becomes overshadowed by noise, preventing model convergence within the allocated epochs.
- The batch size (B) affects the scale of the problem in calculating discrepancies, with sizes ranging from 32 to 2048 examined. There is a consistent decrease in MAE and RMSE as the batch size increases to $\mathrm{B} = 512$, enhancing the reliability of discrepancy estimations. Increasing the batch size beyond this point yields diminishing returns and may lead to unnecessary computational overhead.
- The entropic regularization (ε) facilitates the convergence of FMT at the cost of blurring the transport matrix [57]. When ε is decreased from 0.05 to approximately 0.01, performance improves consistently, as the reduction in blurring diminishes irrelevant pairings. However, reducing ε further to 0.0005 increases MAE and RMSE, as an overly small ε, acting as denominators in the iterative steps of Algorithm 1, complicates FMT convergence. An optimal ε of around 0.01 balances accuracy and convergence effectively. Another interesting observation is that the optimal ε value reduces from 0.01 to 0.005 as the missing ratio decreases from 0.5 to 0.3. It suggests that a more blurred transport matrix can be beneficial in scenarios with high missing ratios by promoting matching to more samples to boost sample efficiency.

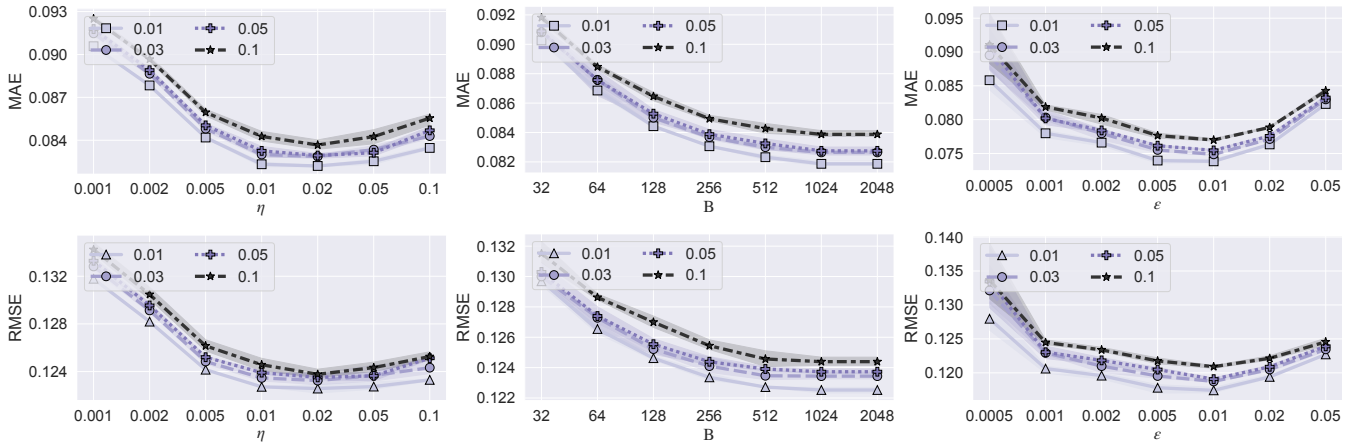


(a) Imputation performance with $\mathrm{p_{miss}} = 0.1$ (b) Imputation performance with $\mathrm{p_{miss}} = 0.3$ (c) Imputation performance with $\mathrm{p_{miss}} = 0.5$

Fig. 4. Imputation error for different mass preservation strengths (κ), shown with colored lines for means and shaded areas for 99.9% confidence intervals. Each color corresponds to a specific ratio of noisy samples ($\mathrm{p_{noise}}$).



(a) Imputation error given missing ratio $\mathrm{p_{miss}} = 0.3$.



(b) Imputation error given missing ratio $\mathrm{p_{miss}} = 0.5$.

Fig. 5. Imputation error for different values of update rate (η), batch size (B) and entropic regularization strength (ε), shown with colored lines for means and shaded areas for 99.9% confidence intervals. Each color corresponds to a specific ratio of noisy samples ($\mathrm{p_{noise}}$).

VI. Conclusion

This study introduces the FMT-I approach, applying an innovative OT perspective to address the MDI task amidst noise interference. At the heart of the FMT-I framework is the flexible mass transport (FMT) formulation, meticulously designed to match typical samples across distributions while excluding noise interference. Leveraging the selective matching mechanism of FMT, the FMT-I approach significantly enhances the resilience of imputation processes against noise, establishing itself as an effective tool for completing industrial datasets compromised by noise interference.

Limitations & future work. The present work does not explicitly account for temporal correlations in datasets, such as periodicity and trends, and instead treats each sample as independent. While this assumption is consistent with the majority of existing MDI methodologies [49, 50] and is appropriate for many industrial applications [61], temporal correlations can be beneficial for imputation tasks if they are available. Future research could extend the proposed OT-based framework to accommodate temporal correlations. For example, one could employ a sliding window approach to construct patches of adjacent observations, treating each patch

as a sample, thereby enabling the capturing and utilization of temporal correlations. Such extensions would enhance the applicability of the framework to industrial datasets where temporal correlation is significant.

REFERENCES

[1] H. Wang, Z. Wang, Y. Niu, Z. Liu, H. Li, Y. Liao, Y. Huang, and X. Liu, "An accurate and interpretable framework for trustworthy process monitoring," *IEEE Trans. Artif. Intell*, vol. 5, no. 5, pp. 2241–2252, 2023.

[2] Z. Chen, H. Wang, Z. Song, and Z. Ge, "Improving data-driven inferential sensor modeling by industrial knowledge: A bayesian perspective," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 55, no. 2, pp. 1043–1055, 2025.

[3] L. Pan, H. Wang, Z. Chen, Y. Huang, Y. Niu, Z. Liu, Q. He, and X. Liu, "Controllable mixture-of-experts for multivariate soft sensors," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 19 789–19 800, 2025.

[4] Z. Chen, Y. Zhang, O. Zhang, F. Wang, L. Yao, H. Wang, and Z. Song, "Blending data and knowledge for process industrial modeling under riemannian preconditioned bayesian framework," *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[5] Z. Zhang, S. Ye, Y. Zhang, W. Ding, and H. Wang, "Belief combination of classifiers for incomplete data," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 4, pp. 652–667, 2022.

[6] Z. Chen, H. Wang, G. Chen, Y. Ma, L. Yao, Z. Ge, and Z. Song, "Analyzing and improving supervised nonlinear dynamical probabilistic latent variable model for inferential sensors," *IEEE Trans. Ind. Informat.*, vol. 20, no. 11, pp. 13 296–13 307, 2024.

[7] H. Wang, Z. Chen, Z. Liu, L. Pan, H. Xu, Y. Liao, H. Li, and X. Liu, "Spot-i: Similarity preserved optimal transport for industrial iot data imputation," *IEEE Trans. Ind. Informat.*, vol. 20, no. 12, pp. 14 421–14 429, 2024.

[8] H. Wang, X. Liu, Z. Liu, H. Li, Y. Liao, Y. Huang, and Z. Chen, "Lspt-d: Local similarity preserved transport for direct industrial data imputation," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 9438–9448, 2025.

[9] D. Jarrett, B. Cebere, T. Liu, A. Curth, and M. van der Schaar, "Hyperimpute: Generalized iterative imputation with automatic model selection," in *Proc. Int. Conf. Mach. Learn.*, vol. 162, 2022, pp. 9916–9937.

[10] Z. Zhang, "Missing data imputation: focusing on single imputation," *Ann. Transl. Med*, vol. 4, no. 1, p. 9, 2016.

[11] P. Royston and I. R. White, "Multiple imputation by chained equations (mice): implementation in stata," *J. Statist. Softw.*, vol. 45, pp. 1–20, 2011.

[12] Y. Ouyang, L. Xie, C. Li, and G. Cheng, "Missdiff: Training diffusion models on tabular data with missing values," *arXiv preprint arXiv:2307.00467*, 2023.

[13] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "Csdi: Conditional score-based diffusion models for probabilistic time series imputation," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 24 804–24 816, 2021.

[14] J. Yoon, J. Jordon, and M. van der Schaar, "GAIN: missing data imputation using generative adversarial nets," in *Proc. Int. Conf. Mach. Learn.*, vol. 80, 2018, pp. 5675–5684.

[15] T. Wang, H. Ke, A. Jolfaei, S. Wen, M. S. Haghighi, and S. Huang, "Missing value filling based on the collaboration of cloud and edge in artificial intelligence of things," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5394–5402, 2022.

[16] H. Li, C. Zheng, W. Wang, H. Wang, F. Feng, and X.-H. Zhou, "Debiased recommendation with noisy feedback," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2024, p. 1576–1586.

[17] H. Wang, Z. Chen, Y. Shen, H. Zheng, D. Yang, D. Zhao, and B. Liang, "Unbiased recommender learning from implicit feedback via weakly supervised learning," in *IEEE Trans. Neural Netw. Learn. Syst.*, 2025.

[18] D. J. Stekhoven and P. Bühlmann, "Missforest—non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012.

[19] H. Chen, Y. Jiang, S. Guo, X. Mao, Y. Lin, and H. Wan, "Difflight: a partial rewards conditioned diffusion model for traffic signal control with missing data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 37, 2024, pp. 123 353–123 378.

[20] Y. Zhang, X. Kong, W. Zhou, J. Liu, Y. Fu, and G. Shen, "A comprehensive survey on traffic missing data imputation," *IEEE Trans. Intell. Transp. Syst.*, 2024.

[21] H. Li, Y. Liao, Z. Tian, Z. Liu, J. Liu, and X. Liu, "Bidirectional stackable recurrent generative adversarial imputation network for specific emitter missing data imputation," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2967–2980, 2024.

[22] Y. Du, C. Yang, N. Yu, W. Lin, Q. Zhao, and S. Wang, "Latent imputation before prediction: A new computational paradigm for de novo peptide sequencing," in *Proc. Int. Conf. Mach. Learn.*, 2025.

[23] D. Um, J. W. Yoon, S. J. Ahn, and Y. Yeo, "Gene-gene relationship modeling based on genetic evidence for single-cell rna-seq data imputation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 37, 2024, pp. 18 882–18 909.

[24] R. Malarvizhi and A. S. Thanamani, "K-nearest neighbor in missing data imputation," *Int. J. Eng. Res. Dev*, vol. 5, no. 1, pp. 5–7, 2012.

[25] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *J. Mach. Learn. Res.*, vol. 11, pp. 2287–2322, 2010.

[26] P. Mattei and J. Frellsen, "MIWAE: deep generative modelling and imputation of incomplete data sets," in *Proc. Int. Conf. Mach. Learn.*, vol. 97, 2019, pp. 4413–4423.

[27] T. Kyono, Y. Zhang, A. Bellot, and M. van der Schaar, "MIRACLE: causally-aware imputation via learning missing data mechanisms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 23 806–23 817.

[28] Y. Liu and A. Constantinou, "Improving the imputation of missing data with markov blanket discovery," in *Proc. Int. Conf. Learn. Represent.*, 2022.

[29] J. Fan, C. Yang, and M. Udell, "Robust non-linear matrix factorization for dictionary learning, denoising, and clustering," *IEEE Trans. Signal Process.*, vol. 69, pp. 1755–1770, 2021.

[30] Y. Hu and D. B. Work, "Robust tensor recovery with fiber outliers for traffic events," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 1, pp. 1–27, 2020.

[31] L. Hu, Y. Jia, W. Chen, L. Wen, and Z. Ye, "A flexible and robust tensor completion approach for traffic data recovery with low-rankness," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 3, pp. 2558–2572, 2023.

[32] X. Feng, H. Zhang, C. Wang, and H. Zheng, "Traffic data recovery from corrupted and incomplete observations via spatial-temporal trpca," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 17 835–17 848, 2022.

[33] J. Kim, K. Lee, and T. Park, "To predict or not to predict? proportionally masked autoencoders for tabular data imputation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 39, no. 17, 2025, pp. 17 886–17 894.

[34] T. Du, L. M. Melis, and T. Wang, "Remasker: Imputing tabular data with masked autoencoding," in *Proc. Int. Conf. Learn. Represent.*, 2024.

[35] Z. Pan, Y. Wang, K. Wang, H. Chen, C. Yang, and W. Gui, "Imputation of missing values in time series using an adaptive-learned median-filled deep autoencoder," *IEEE Trans. Cybern.*, vol. 53, no. 2, pp. 695–706, 2022.

[36] M. Kang, R. Zhu, D. Chen, X. Liu, and W. Yu, "Cm-gan: A cross-modal generative adversarial network for imputing completely missing data in digital industry," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 3, pp. 2917–2926, 2023.

[37] Z. Sun, H. Li, W. Wang, J. Liu, and X. Liu, "DTIN: dual transformer-based imputation nets for multivariate time series emitter missing data," *Knowl. Based Syst.*, vol. 284, p. 111270, 2024.

[38] H. Li, Y. Liao, Z. Tian, Z. Liu, J. Liu, and X. Liu, "Bidirectional stackable recurrent generative adversarial imputation network for specific emitter missing data imputation," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2967–2980, 2024.

[39] Z. Chen, H. Li, F. Wang, H. Zhang, H. Xu, X. Jiang, Z. Song, and H. Wang, "Rethinking the diffusion models for missing data imputation: A gradient flow perspective," in *Proc. Adv. Neural Inf. Process. Syst.*, 2024.

[40] C. Ou, H. Zhu, Y. A. Shardt, L. Ye, X. Yuan, Y. Wang, C. Yang, and W. Gui, "Missing-data imputation with position-encoding denoising auto-encoders for industrial processes," *IEEE Trans. Instrum. Meas.*, 2024.

[41] D. Liu, Y. Wang, C. Liu, K. Wang, X. Yuan, and C. Yang, "Blackout missing data recovery in industrial time series based on masked-former hierarchical imputation framework," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1138–1150, 2023.

[42] X.-Y. Li, Y. Xu, Q.-X. Zhu, and Y.-L. He, "Industrial data imputation based on multiscale spatiotemporal information embedding with asymmetrical transformer," *IEEE Trans. Neural Netw. Learn. Syst.*, 2025.

[43] X. Zhou, J. Wang, E. Sui, W. Wang, and J. Li, "Multi-dimensional attention-based gain for missing data imputation in multimode batch processes," *IEEE Trans. Instrum. Meas.*, 2025.

[44] Z. Yao and C. Zhao, "Figan: A missing industrial data imputation method customized for soft sensor application," *IEEE Trans. Autom.*

Sci. Eng., vol. 19, no. 4, pp. 3712–3722, 2021.
[45] Z. Zhan, D. Ma, X. Hu, and S. Zhang, "An online collaborative imputation method for industrial missing data based on multi-scale matgan in edge computing," *IEEE Internet Things J.*, 2025.
[46] B. Lu, Q. Miao, Y. Liu, T. S. Tamir, H. Zhao, X. Zhang, Y. Lv, and F.-Y. Wang, "A diffusion model for traffic data imputation," *IEEE/CAA J. Autom. Sinica*, vol. 12, no. 3, pp. 606–617, 2025.
[47] H. Xu, Z. Liu, H. Wang, C. Li, Y. Niu, W. Wang, and X. Liu, "Denoising diffusion straightforward models for energy conversion monitoring data imputation," *IEEE Trans. Ind. Informat.*, vol. 20, no. 10, pp. 11 987–11 997, 2024.
[48] D. Liu, Y. Wang, C. Liu, X. Yuan, K. Wang, and C. Yang, "Scope-free global multi-condition-aware industrial missing data imputation framework via diffusion transformer," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 6977–6988, 2024.
[49] H. Zhao, K. Sun, A. Dezfouli, and E. V. Bonilla, "Transformed distribution matching for missing value imputation," in *Proc. Int. Conf. Mach. Learn.*, vol. 202, 2023, pp. 42 159–42 186.
[50] B. Muzellec, J. Josse, C. Boyer, and M. Cuturi, "Missing data imputation using optimal transport," in *Proc. Int. Conf. Mach. Learn.*, vol. 119, 2020, pp. 7130–7140.
[51] G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Mem. Math. Phys. Acad. Royale Sci.*, pp. 666–704, 1781.
[52] L. V. Kantorovich, "On the translocation of masses," *J. Math. Sci.*, vol. 133, no. 4, pp. 1381–1382, 2006.
[53] Y. Disser and M. Skutella, "The simplex algorithm is np-mighty," *ACM Trans. Algorithms*, vol. 15, no. 1, pp. 5:1–5:19, 2019.
[54] K. Fatras, T. Séjourné, R. Flamary, and N. Courty, "Unbalanced mini-batch optimal transport; applications to domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 3186–3197.
[55] L. Chizat, G. Peyré, B. Schmitzer, and F. Vialard, "Scaling algorithms for unbalanced optimal transport problems," *Math. Comput.*, vol. 87, no. 314, pp. 2563–2609, 2018.
[56] T. Séjourné, J. Feydy, F. Vialard, A. Trouvé, and G. Peyré, "Sinkhorn divergences for unbalanced optimal transport," *CoRR*, vol. abs/1910.12958, 2019.
[57] J. M. Altschuler, J. Weed, and P. Rigollet, "Near-linear time approximation algorithms for optimal transport via sinkhorn iteration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1964–1974.
[58] P. E. Dvurechensky, A. V. Gasnikov, and A. Kroshnin, "Computational optimal transport: Complexity by accelerated gradient descent is better than by sinkhorn's algorithm," in *Proc. Int. Conf. Mach. Learn.*, vol. 80, 2018, pp. 1366–1375.
[59] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3235–3243, 2018.
[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–9.
[61] L. Yao and Z. Ge, "Industrial big data modeling and monitoring framework for plant-wide processes," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6399–6408, 2021.
[62] G. Peyré and M. Cuturi, "Computational optimal transport," *Found. Trends Mach. Learn.*, vol. 11, no. 5-6, pp. 355–607, 2019.
[63] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1853–1865, 2017.
[64] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich, "Displacement interpolation using lagrangian mass transport," *ACM Trans. Graph.*, vol. 30, no. 6, p. 158, 2011.
[65] H. Wang, Z. Chen, Z. Liu, X. Chen, H. Li, and Z. Lin, "Proximity matters: Local proximity enhanced balancing for treatment effect estimation," *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2025.
[66] H. Wang, Z. Chen, H. Zhang, Z. Li, L. Pan, H. Li, and M. Gong, "Debiased recommendation via wasserstein causal balancing," *ACM Trans. Inf. Syst.*, 2025.
[67] S. D. Marino and A. Gerolin, "An optimal transport approach for the schrödinger bridge problem and convergence of sinkhorn algorithm," *J. Sci. Comput.*, vol. 85, no. 2, p. 27, 2020.
[68] H. Xu, D. Luo, H. Zha, and L. Carin, "Gromov-wasserstein learning for graph matching and node embedding," in *Proc. Int. Conf. Mach. Learn.*, vol. 97, 2019, pp. 6932–6941.
[69] H. Wang, Z. Chen, J. Fan, H. Li, T. Liu, W. Liu, Q. Dai, Y. Wang, Z. Dong, and R. Tang, "Optimal transport for treatment effect estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 5404–5418.
[70] H. Wang, Z. Li, H. Li, X. Chen, M. Gong, B. Chen, and Z. Chen, "Optimal transport for time series imputation," in *Proc. Int. Conf. Learn. Represent.*, 2025, pp. 1–9.

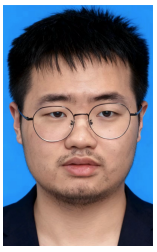
Hao Wang received the B. Eng. degree in detection, guidance, and control technology from the College of Aeronautics and Astronautics, Central South University, Hunan, China, in 2020. He is currently working toward the Ph.D. degree in electronic engineering with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China.

His research interests include process monitoring and time-series analysis.

Licheng Pan received the B.Eng. degree in automation from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2021 and is currently pursuing the Ph.D. degree in cyberspace security with the College of Computer Science and Technology, Zhejiang University, Hangzhou, China. His research interests include multi-task learning, time series analysis, and the construction of secure and trustworthy large language models.

Yuan Lu is currently the technical leader of the Agent Team at Xiaohongshu. Prior to joining Xiaohongshu, he worked at DiDi AI Lab and Alibaba DAMO Academy, where he focused on research and development in pre-trained models for over 7 years.

His research interests include representation learning, dialogue systems, graph neural networks, and pre-training for both language and multimodal models.



Zhengnan Li received his Bachelor's degree in Digital Media Arts from the Communication University of China and is currently pursuing his Master's degree in Data Science at The Chinese University of Hong Kong, China.

His research focuses on agent memory, symbolic logic, and reasoning in large language models, extending to related areas such as time series analysis and missing data imputation.

Shuting He received the BE degree from Xiamen University, China, in 2018, and the PhD degree from Zhejiang University, Hangzhou, China, in 2023. She was a research fellow with Nanyang Technological University, Singapore, from 2023 to 2024. She is currently a tenure-track assistant professor with the Shanghai University of Finance and Economics, Shanghai, China.

Her research interests include computer vision and machine learning. She is also an area chair of CVPR, BMVC, and ICLR.

Xiaoyu Jiang (Member, IEEE) received the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2023. He is currently an associate professor with the Hangzhou International Innovation Institute, Beihang University, Hangzhou, China.

His research interests include deep learning, industrial big data, process monitoring, and AI reliability.

Zhichao Chen received the Ph.D. degree in Control Science and Engineering from the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, China, in 2025. He is currently a Post-Doctoral Research Fellow with the School of Intelligence Science and Technology, Peking University, Beijing, China.

His research interests include process data analytics, both linear and nonlinear optimization, variational Bayesian inference, ordinary/partial/stochastic differential equation, and optimal control.

Xinggao Liu received the Ph.D. degree from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2000, respectively. Then, he was a Postdoctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China, from 2000 to 2002.

He is currently a full-time Professor with the College of Control Science and Engineering, Zhejiang University. His main research interests include signal processing, nuclear process monitoring, time series analysis, and dynamic optimization.

APPENDIX

A. Summary of symbols

The primary symbols used are summarized in Table IV.

TABLE IV
THE LIST OF IMPORTANT SYMBOLS

Symbol	Definition
Symbols of Data and Mask Matrice	
$\mathbf{X}^{(\mathrm{id})}$	The ideal complete data matrix.
$\mathbf{X}^{(\mathrm{obs})}$	The observed data matrix with missing entries.
$\mathbf{X}^{(\mathrm{imp})}$	The final imputed data matrix.
$\mathbf{X}^{t}$	The imputed data matrix at iteration t.
$\mathbf{M}$	The binary mask matrix.
Symbols of Parameters and Hyperparameters	
N	The number of samples in dataset.
D	The number of features/dimensions.
B	The batch size for mini-batch sampling.
η	The update rate (learning rate).
$\mathrm{T}_{\max}$	The maximum number of training epochs.
$\ell_{\max}$	The maximum number of iterations for the FMT solver.
$\mathrm{p}_{\mathrm{miss}}$	The missing ratio.
$\mathrm{p}_{\mathrm{noise}}$	The noise ratio.
Symbols of Optimal Transport	
α, β	The sampled mini-batches from dataset.
n, m	The number of samples in α and β, respectively.
$\mathbf{D}$	The pairwise Euclidean distance matrix.
$\mathbf{P}$	The transport matrix.
$\mathbf{P}^*$	The optimal transport matrix.
$\mathbf{a}, \mathbf{b}$	The mass vectors of α and β.
$\Delta_{\mathrm{n}}, \Delta_{\mathrm{m}}$	The simplex vectors with length n and m, respectively.
$\mathbf{1}_{\mathrm{n}}, \mathbf{1}_{\mathrm{m}}$	The vector of ones with length n and m, respectively.
$\Pi(\alpha, \beta)$	The constraint set for standard optimal transport.
$\mathcal{W}^{\varepsilon}(\alpha, \beta)$	The Wasserstein discrepancy.
$\mathcal{P}^{\varepsilon,\kappa}(\alpha, \beta)$	The FMT discrepancy.
$\mathbf{f}, \mathbf{g}$	The dual potentials of FMT.
ε	The entropic regularization strength.
κ	The mass-preservation strength.
Others	
$\|\cdot\|_2$	The Euclidean (ℓ_2) norm.
$\|\cdot\|_{\mathrm{F}}$	The Frobenius norm.
$\mathrm{D}_{\mathrm{KL}}(\cdot\|\cdot)$	The Kullback-Leibler divergence.
$\mathrm{H}(\cdot)$	The entropy function.

B. Background on discrete optimal transport

This section introduces the foundational concepts and algorithms necessary for computing OT between discrete measures. Consider a scenario involving n warehouses and m factories, where the i-th warehouse holds $\mathbf{a}_i$ units of material and the j-th factory requires $\mathbf{b}_j$ units of material [62]. The objective is to establish a mapping from warehouses to factories that: (1) completely allocates all warehouse materials, (2) fulfills all factory demands, and (3) restricts the transportation from each warehouse to at most one factory. Each potential mapping is evaluated based on a global cost, which aggregates the local costs incurred from transporting a unit of material from warehouse i to factory j.

Definition A.1. *The Monge problem for discrete measures, where $\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{\mathbf{x}_i}$ and $\beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{\mathbf{x}_j}$, seeks a mapping $\mathbb{T} : \{\mathbf{x}_i\}_{i=1}^{n} \to \{\mathbf{x}_j\}_{j=1}^{m}$ that optimally redistributes the mass from α to β. Specifically, for each j, it must hold that $\mathbf{b}_j = \sum_{i:\mathbb{T}(\mathbf{x}_i)=\mathbf{x}_j} \mathbf{a}_i$ and $\mathbb{T}_\sharp \alpha = \beta$. The goal is to minimize the transportation cost, represented by $c(x, y)$, leading to the following formulation:*

$$\min_{\mathbb{T}:\mathbb{T}_\sharp \alpha=\beta} \left\{ \sum_i c(\mathbf{x}_i, \mathbb{T}(\mathbf{x}_i)) \right\}.$$

The original Monge formulation does not guarantee the existence or uniqueness of solutions [62]. Therefore, Kantorovich [52] extended this framework by relaxing the one-to-one mapping constraint, allowing transportation from a single warehouse to multiple factories, and reformulated the problem as a linear programming challenge.

Definition A.2. *The Kantorovich problem for discrete measures α and β defines a cost-minimization task over feasible transport plans $\pi \in \mathbb{R}_+^{n\times m}$. The objective is to find a plan that minimizes the overall transport cost:*

$$\mathcal{W}(\alpha, \beta) := \min_{\pi \in \Pi(\alpha,\beta)} \langle \mathbf{D}, \pi \rangle,$$
$$\Omega(\alpha, \beta) := \left\{ \pi \in \mathbb{R}_+^{n\times m} : \pi \mathbf{1}_m = \mathbf{a}, \pi^{\mathrm{T}} \mathbf{1}_n = \mathbf{b} \right\},$$

where $\mathcal{W}(\alpha, \beta)$ represents the Wasserstein discrepancy between α and β; $\mathbf{D}$ denotes the distance matrix computed using the squared Euclidean metric [63], and $\mathbf{a}$, $\mathbf{b}$ are vectors describing the mass distribution in α and β, respectively.

The subsequent research primarily follows two trajectories. The first aims to reduce the computational complexity of solving OT problems. While exact solutions can be obtained through linear programming algorithms, these come with cubic complexity in relation to the number of samples [64]. To address this, various approximate algorithms for acceleration have been developed, such as the Sinkhorn and sliced OT algorithm [57] with quadratic and linear complexity, respectively. The second line of research focuses on modifying the transport problem to suit specific applications [8, 65, 66]. Examples include the weak transport problem in domain adaptation [55], the Schrödinger bridge problem in generative modeling [67], the Gromov problem in graph matching [68], the unbalanced transport problem in causal inference [69] and imputation [70].

C. Convergence analysis

Theorem A.1. *Let $\mathcal{P}(\zeta)$ be the FMT transport cost in (3) for the imputation values $\zeta = (\alpha, \beta)$, and let $\zeta_k = (\alpha_k, \beta_k)$ be the imputation values at the k-th iteration, ζ^* be the optimal imputation values that minimize $\mathcal{P}$. We assume that $\mathcal{P}$ is convex and has an L-Lipschitz continuous gradient (i.e., is L-smooth), and that the initial parameters ζ_0 are close to ζ^* such that $\|\zeta_0 - \zeta^*\|_2 \leq \epsilon$ for some constant $\epsilon > 0$. When optimized via gradient descent with a step size $\eta \leq 1/L$, the FMT-I framework guarantees that the optimality gap after K iterations is bounded by:*

$$\mathcal{P}(\alpha_k, \beta_k) - \mathcal{P}(\alpha^*, \beta^*) \leq \frac{1}{2\eta \mathrm{K}} \epsilon^2.$$

Proof. By the way of preface, we demonstrate that the objective function $\mathcal{P}$ is convex with respect to ζ. This convexity follows from its composition. The parameters ζ influences $\mathcal{P}$ solely through the term $\langle \mathbf{D}, \mathbf{P} \rangle$, where $\mathbf{D}_{i,j} = \|\alpha_i - \beta_j\|_2$. Since the Euclidean norm $\|\cdot\|_2$ and inner product $\langle \cdot \rangle$ are convex, according to the fact that the cascade of convex

functions is convex, we have $\langle \mathbf{D}, \mathbf{P} \rangle$ is convex to ζ=(α, β). Since $\langle \mathbf{D}, \mathbf{P} \rangle$ is the only part in $\mathcal{P}$ that is related to ζ, we derive $\mathcal{P}$ is convex to ζ.

Based on the convexity of $\mathcal{P}$, we assume $\mathcal{P}$ has L-Lipschitz continuous gradients, i.e., $\|\mathcal{P}(\zeta_{k+1}-\mathcal{P}(\zeta_k)\|_2 < \mathrm{L}\|\zeta_{k+1}-\zeta\|_2$. Under these conditions and assumptions, we can follow the standard practices to derive the convergence of FMT-I, which employs the gradient descent update rule:

$$\zeta_{k+1} = \zeta_k - \eta \nabla_\zeta \mathcal{P}(\zeta_k),$$

The convergence analysis proceeds by first establishing a key recurrence relation. By the descent lemma for L-smooth functions, we have: Since $\mathcal{P}$ has L-Lipschitz continuous gradients, we have:

$$\mathcal{P}(\zeta_{k+1}) \leq \mathcal{P}(\zeta_k) + \nabla_\zeta \mathcal{P}(\zeta_k)(\zeta_{k+1} - \zeta_k) + \frac{L}{2}\|\zeta_{k+1} - \zeta_k\|^2.$$

Substituting the gradient descent update rule into the inequality above, we have

$$\begin{aligned}\mathcal{P}(\zeta_{k+1}) &\leq \mathcal{P}(\zeta_k) - \eta\|\nabla_\zeta \mathcal{P}(\zeta_k)\|^2 + \frac{L\eta^2}{2}\|\nabla_\zeta \mathcal{P}(\zeta_k)\|^2. \\ &= \mathcal{P}(\zeta_k) - \eta(1 - \frac{L\eta}{2})\|\nabla_\zeta \mathcal{P}(\zeta_k)\|^2.\end{aligned}$$

By setting the step size $\eta \leq 1/\mathrm{L}$, we have $(1 - \frac{L\eta}{2}) > 1/2$, which simplifies the bound to

$$\mathcal{P}(\zeta_{k+1}) \leq \mathcal{P}(\zeta_k) - \frac{\eta}{2}\|\nabla_\zeta \mathcal{P}(\zeta_k)\|^2.$$

This result confirms that the value of $\mathcal{P}$ is non-increasing along the optimization process. Subsequently, we have

$$\mathcal{P}(\zeta_{k+1}) \leq \mathcal{P}(\zeta^*) + \nabla_\zeta \mathcal{P}(\zeta_k)(\zeta_k - \zeta^*) - \frac{\eta}{2}\|\nabla_\zeta \mathcal{P}(\zeta_k)\|^2.$$

which immediately follows from the convexity of $\mathcal{P}$: $\mathcal{P}(\zeta_k) \leq \mathcal{P}(\zeta^*)+\nabla_\zeta \mathcal{P}(\zeta_k)(\zeta_k-\zeta^*)$. We can further simplify this bound using simple algebra:

$$\begin{aligned}\mathcal{P}(\zeta_{k+1}) &\leq \mathcal{P}(\zeta^*) + \nabla_\zeta \mathcal{P}(\zeta_k)(\zeta_k - \zeta^*) - \frac{\eta}{2}\|\nabla_\zeta \mathcal{P}(\zeta_k)\|^2 \\ &= \mathcal{P}(\zeta^*) + \frac{1}{2\eta}\|\zeta_k - \zeta^*\|^2 - \frac{1}{2\eta}\|\zeta_k - \zeta^*\|^2 \\ &+ \nabla_\zeta \mathcal{P}(\zeta_k)(\zeta_k - \zeta^*) - \frac{\eta}{2}\|\nabla_\zeta \mathcal{P}(\zeta_k)\|^2 \\ &= \mathcal{P}(\zeta^*) + \frac{1}{2\eta}\|\zeta_k - \zeta^*\|^2 - \frac{1}{2\eta}\|\zeta_k - \zeta^* - \eta\nabla_\zeta \mathcal{P}(\zeta_k)\|^2 \\ &= \mathcal{P}(\zeta^*) + \frac{1}{2\eta}(\|\zeta_k - \zeta^*\|^2 - \|\zeta_{k+1} - \zeta^*\|^2),\end{aligned}$$

Rearranging this inequality gives $\mathcal{P}(\zeta_{k+1}) - \mathcal{P}(\zeta^*) \leq \frac{1}{2\eta}(\|\zeta_k - \zeta^*\|_2^2 - \|\zeta_{k+1} - \zeta^*\|_2^2)$. By summing this relation from $k = 0$ to $\mathrm{K} - 1$, we obtain a telescoping series:

$$\begin{aligned}\sum_{k=0}^{\mathrm{K}-1}(\mathcal{P}(\zeta_{k+1}) - \mathcal{P}(\zeta^*)) &\leq \sum_{k=0}^{\mathrm{K}-1}\frac{1}{2\eta}\left(\|\zeta_k - \zeta^*\|_2^2 - \|\zeta_{k+1} - \zeta^*\|_2^2\right) \\ &= \frac{1}{2\eta}\left(\|\zeta_0 - \zeta^*\|_2^2 - \|\zeta_K - \zeta^*\|_2^2\right) \\ &\leq \frac{1}{2\eta}\|\zeta_0 - \zeta^*\|_2^2.\end{aligned}$$

Since the sequence $\mathcal{P}(\zeta_k)$ is non-increasing, we have $\mathrm{K}(\mathcal{P}(\zeta_\mathrm{K}) - \mathcal{P}(\zeta^*)) \leq \sum_{k=1}^{\mathrm{K}}(\mathcal{P}(\zeta_k) - \mathcal{P}(\zeta^*)) = \sum_{k=0}^{\mathrm{K}-1}(\mathcal{P}(\zeta_{k+1}) - \mathcal{P}(\zeta^*))$. Combining these inequalities yields:

$$\mathcal{P}(\zeta_K) - \mathcal{P}(\zeta^*) \leq \frac{\|\zeta_0 - \zeta^*\|_2^2}{2\eta K} \leq \frac{\epsilon^2}{2\eta K}.$$

where we assumed that the initial parameters are close to ζ^*: $\|\zeta_0 - \zeta^*\|_2 \leq \epsilon$. □

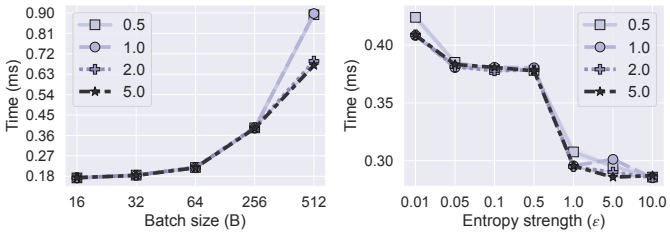


Fig. 6. Running time of solving FMT in varying B, ε andκ. Each color corresponds to a specific value of κ.

D. Gradient derivation

After acquiring the optimum transport matrix, denoted as $\mathbf{P}$, by solving the optimization problem in (3), the FMT is calculated as

$$\mathcal{P}(\alpha, \beta) := \langle \mathbf{D}, \mathbf{P} \rangle$$

which depends on α and β through the distance matrix $\mathbf{D}_{i,j} = \|\alpha_i - \beta_j\|_2 = \sqrt{(\alpha_i - \beta_j)^\top(\alpha_i - \beta_j)}$. Using the chain rule, we have:

$$\begin{aligned}\frac{\partial \mathcal{P}}{\partial \alpha_i} &= \sum_{j=1}^{\mathrm{m}} \frac{\partial \mathcal{P}}{\partial \mathbf{D}_{i,j}}\frac{\partial \mathbf{D}_{i,j}}{\partial \alpha_i} = \sum_{j=1}^{\mathrm{m}} \mathbf{P}_{i,j}\frac{\alpha_i - \beta_j}{\|\alpha_i - \beta_j\|_2}, \\ \frac{\partial \mathcal{P}}{\partial \beta_j} &= \sum_{i=1}^{\mathrm{n}} \frac{\partial \mathcal{P}}{\partial \mathbf{D}_{i,j}}\frac{\partial \mathbf{D}_{i,j}}{\partial \beta_j} = -\sum_{i=1}^{\mathrm{n}} \mathbf{P}_{i,j}\frac{\beta_j - \alpha_i}{\|\alpha_i - \beta_j\|_2}.\end{aligned}$$

E. Complexity analysis

In this section, we examine the computational requirements. Although we have previously discussed the amount of epochs where FMT-I converges, we have not yet investigated the computational costs associated with a single epoch. Therefore, we focus on the time needed for each epoch, specifically the execution time of Algorithm 1 for solving the FMT problem, under varying batch sizes (B), entropic regularization strengths (ε), and mass-preservation strengths (κ). The results are detailed in Fig. 6 and are outlined below.

- Increasing the batch size (B) raises the computational cost since it enlarges the transport matrix ($\mathbf{P}$) and thereby the scale of the optimization problem. Nonetheless, the running time remains practical, less than 1 ms even at the large batch size of 512. Thus, while larger batch sizes increase computational load, they do not significantly hinder practical applications.
- Raising the entropic regularization strength (ε) from 0.01 to 10 slightly decreases the computational cost. This occurs because a higher ε facilitates the convergence of the Sinkhorn-like algorithm [57]. In Algorithm 1, ε appears in the denominator in key iteration steps (steps 4, 5, 8, 10). An excessively small denominator can slow down

the convergence of the iterative process, requiring more iterations to reach plateau in Algorithm 1.

- Interestingly, the mass-preservation strength (κ) has a minimal impact on the running time compared to other hyperparameters. In Algorithm 1, κ functions as both the numerator and denominator in iteration steps (steps 8, 10). Therefore, increasing κ does not significantly alter the values of $\mathbf{f}$ and $\mathbf{g}$, nor does it impede the convergence of Algorithm 1.

F. Additional experimental results on different missing mechanisms

Real-world data missingness, especially in industrial contexts, is often complex and multifaceted. According to the mechanism of generating the missing data indicator matrix $\mathbf{O}$, data missingness can be categorized into three types [9]: Missing Completely At Random (MCAR), where the missingness is independent of any data values; Missing At Random (MAR), where missingness depends on observed values; and Missing Not At Random (MNAR), where missingness is influenced by the values that are missing themselves.

In this paper, our experimental focus is exclusively on the MCAR setting, which does not suffice to showcase the efficacy of FMT-I in real-world settings. Therefore, we extend the experiments to more complex MAR and MNAR patterns.

- To simulate MAR data, features are randomly divided into two categories: observable and missing. Missingness is determined by a logistic model influenced by observable features, parameterized by randomly selected weights and a bias adjusted to achieve the desired missing ratio.
- To simulate MNAR data, the non-missing features from the MAR simulation are performed with a secondary MCAR masking. This additional step creates dependencies between missing values and certain observable features, thus replicating the MNAR condition.

The results of these extended experiments are summarized in Table V. They show that FMT-I consistently outperforms comparative methods across all types of missing data patterns. In the MNAR setting, FMT-I notably achieved the lowest imputation error across various missing ratios. Similarly, in the MAR scenario, FMT-I demonstrated a reduction in average MAE by over 15% relative to other leading baselines. Therefore, FMT-I excels at handling complex and varied data missingness patterns, making it a superior choice for industrial applications where such challenges are prevalent.

G. Additional experiments on missing ratios and noisy conditions

In real-world industrial applications, missing data imputation is often complicated by the presence of noisy observations, which can distort the distribution of observed data and challenge the effectiveness of imputation models. Therefore, it is necessary to compare the baseline models and FMT-I in the cases where noisy observations exist. In this paper, our experiments were limited to a fixed missing ratio of $\mathrm{p_{miss}} = 0.1$. In this section, we expand our analysis to include a broader range of missing ratios $\{0.1, 0.3, 0.5, 0.7\}$ and noise ratios $\{0.01, 0.03, 0.05, 0.1\}$, providing a more comprehensive evaluation of baseline models and FMT-I under varied conditions. The results of these expanded experiments are detailed in Table III, with key observations summarized below:

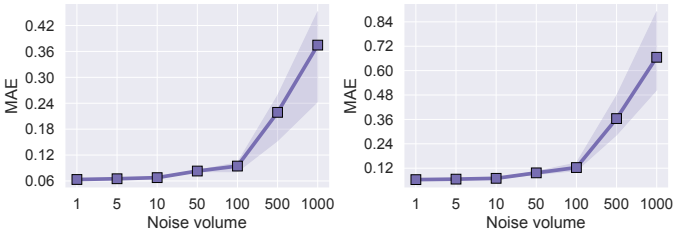


Fig. 7. Performance of FMT-I in varying noise volume, shown with lines for means and shaded areas for 99.9% confidence intervals. The missing ratio is 0.3. The noise ratio is 0.02 (left panel) and 0.05 (right panel).

- As noise ratios increase, most baselines show significant performance degradation, highlighting their susceptibility to noisy data. An exception is observed with baselines that rely on global statistics, such as the Mean and Median, which maintain stable performance regardless of the noise and missing ratios. The Mean approach for instance performs optimally given $\mathrm{p_{miss}} = 0.7$ and $\mathrm{p_{noise}} = 0.7$. This stability is attributed to the zero-mean characteristic of the Gaussian noise used in our experiments, which does not alter the overall statistics of the dataset as noise ratios increase. However, this resilience is contingent on the nature of the noise; with non-zero mean noise, these methods would likely falter as global data statistics shift.
- In scenarios with low missing ratios, FMT-I consistently outperforms other models across all noise levels, showing minimal impact from increased noise. For instance, even as $\mathrm{p_{noise}}$ increases from 0.01 to 0.1 at a missing ratio of $\mathrm{p_{miss}} = 0.1$, FMT-I exhibits only a slight change in MAE. This resilience is largely due to its selective matching mechanism, which enables FMT-I to effectively pair accurate samples while ignoring outliers and disturbances.
- As the missing ratio increases, the advantage of FMT-I over simpler methods like mean imputation diminishes. At $\mathrm{p_{miss}} = 0.7$, for example, FMT-I is surpassed by Mean in 2 out of 8 metrics. This phenomenon can be attributed to the decreasing availability of observations necessary for constructing a reliable transport matrix, a central component of the FMT-I's methodology. The selective matching mechanism, while beneficial in filtering out less reliable data, exacerbates this issue by further limiting the usable data. Therefore, the selective matching mechanism should be conservatively employed given high missing ratios, with the mass-preservation strength κ tuned towards 1.

H. Additional experiments on noise volumes

In this section, we examine the performance of FMT-I given varying noise volumes, complementing our previous analysis on varying noise ratios with fixed noise volume (Section V-D). Specifically, we fix the noise ratio at 0.02 and 0.05, and progressively increase the noise volume from 1 to 1000.

TABLE V
COMPARATIVE STUDY OF FMT-I VERSUS OTHER BASELINE METHODS GIVEN DIFFERENT MISSING RATIOS $\mathrm{p_{miss}}$.

Missing ratio	$\mathrm{p_{miss}}=0.1$		$\mathrm{p_{miss}}=0.3$		$\mathrm{p_{miss}}=0.5$		$\mathrm{p_{miss}}=0.7$	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Missing mechanism: MCAR								
Mean	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$
Median	$0.103_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$
Mode	$0.142_{\pm 0.000}$	$0.197_{\pm 0.000}$	$0.140_{\pm 0.000}$	$0.195_{\pm 0.000}$	$0.151_{\pm 0.000}$	$0.208_{\pm 0.000}$	$0.153_{\pm 0.000}$	$0.216_{\pm 0.000}$
ICE	$0.095_{\pm 0.000}$	$0.142_{\pm 0.000}$	$0.108_{\pm 0.000}$	$0.156_{\pm 0.000}$	$0.121_{\pm 0.000}$	$0.170_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.191_{\pm 0.000}$
MICE	$0.071_{\pm 0.000}$	$0.109_{\pm 0.000}$	$0.081_{\pm 0.000}$	$0.121_{\pm 0.000}$	$0.090_{\pm 0.000}$	$\underline{0.131}_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.146_{\pm 0.000}$
GAIN	$0.088_{\pm 0.002}$	$0.126_{\pm 0.004}$	$0.092_{\pm 0.004}$	$0.129_{\pm 0.005}$	$0.140_{\pm 0.019}$	$0.185_{\pm 0.020}$	$0.338_{\pm 0.074}$	$0.400_{\pm 0.075}$
Miss.D.	$0.246_{\pm 0.081}$	$0.286_{\pm 0.094}$	$0.261_{\pm 0.031}$	$0.311_{\pm 0.035}$	$0.262_{\pm 0.020}$	$0.313_{\pm 0.022}$	$0.276_{\pm 0.020}$	$0.328_{\pm 0.022}$
Miss.F.	$0.067_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.079_{\pm 0.001}$	$0.118_{\pm 0.001}$	$\underline{0.089}_{\pm 0.003}$	$0.133_{\pm 0.003}$	$0.104_{\pm 0.002}$	$0.149_{\pm 0.002}$
Sinkhorn	$0.097_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.101_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.143_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.146_{\pm 0.000}$
TRPCA	$0.067_{\pm 0.009}$	$0.100_{\pm 0.013}$	$0.088_{\pm 0.004}$	$0.126_{\pm 0.006}$	$0.107_{\pm 0.008}$	$0.150_{\pm 0.013}$	$0.134_{\pm 0.022}$	$0.181_{\pm 0.031}$
SCPN	$\underline{0.065}_{\pm 0.000}$	$\underline{0.100}_{\pm 0.000}$	$0.081_{\pm 0.000}$	$0.118_{\pm 0.000}$	$0.092_{\pm 0.000}$	$0.133_{\pm 0.000}$	$\underline{0.098}_{\pm 0.000}$	$\underline{0.139}_{\pm 0.000}$
NewImp	$0.069_{\pm 0.000}$	$0.108_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.146_{\pm 0.000}$
SoftImp	$0.087_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.096_{\pm 0.000}$	$0.135_{\pm 0.000}$	$0.119_{\pm 0.000}$	$0.169_{\pm 0.000}$	$0.174_{\pm 0.000}$	$0.251_{\pm 0.000}$
MIWAE	$0.080_{\pm 0.000}$	$0.129_{\pm 0.003}$	$0.089_{\pm 0.008}$	$0.132_{\pm 0.006}$	$0.111_{\pm 0.000}$	$0.154_{\pm 0.001}$	$0.111_{\pm 0.000}$	$0.155_{\pm 0.000}$
Miracle	$0.069_{\pm 0.000}$	$0.148_{\pm 0.000}$	$\underline{0.074}_{\pm 0.000}$	$\underline{0.114}_{\pm 0.000}$	$0.090_{\pm 0.000}$	$0.132_{\pm 0.000}$	$0.128_{\pm 0.000}$	$0.164_{\pm 0.000}$
TDM	$0.067_{\pm 0.001}$	$0.109_{\pm 0.002}$	$0.083_{\pm 0.000}$	$0.126_{\pm 0.001}$	$0.094_{\pm 0.000}$	$0.136_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.001}$
FMT-I	$\mathbf{0.040}^{*}_{\pm 0.000}$	$\mathbf{0.068}^{*}_{\pm 0.001}$	$\mathbf{0.058}^{*}_{\pm 0.000}$	$\mathbf{0.097}^{*}_{\pm 0.000}$	$\mathbf{0.074}^{*}_{\pm 0.000}$	$\mathbf{0.118}^{*}_{\pm 0.000}$	$\mathbf{0.086}^{*}_{\pm 0.000}$	$\mathbf{0.130}^{*}_{\pm 0.000}$
Missing mechanism: MAR								
Mean	$0.108_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.142_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.143_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.146_{\pm 0.000}$
Median	$0.105_{\pm 0.000}$	$0.142_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.149_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.151_{\pm 0.000}$	$0.111_{\pm 0.000}$	$0.155_{\pm 0.000}$
Mode	$0.145_{\pm 0.000}$	$0.197_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.196_{\pm 0.000}$	$0.140_{\pm 0.000}$	$0.190_{\pm 0.000}$	$0.134_{\pm 0.000}$	$0.184_{\pm 0.000}$
ICE	$0.086_{\pm 0.000}$	$0.123_{\pm 0.000}$	$0.088_{\pm 0.000}$	$0.124_{\pm 0.000}$	$0.088_{\pm 0.000}$	$0.126_{\pm 0.000}$	$0.092_{\pm 0.000}$	$0.132_{\pm 0.000}$
MICE	$0.067_{\pm 0.000}$	$0.099_{\pm 0.000}$	$0.069_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.068_{\pm 0.000}$	$0.100_{\pm 0.000}$	$0.071_{\pm 0.000}$	$0.105_{\pm 0.000}$
GAIN	$0.079_{\pm 0.010}$	$0.111_{\pm 0.013}$	$0.076_{\pm 0.004}$	$0.106_{\pm 0.005}$	$0.101_{\pm 0.012}$	$0.132_{\pm 0.013}$	$0.182_{\pm 0.081}$	$0.238_{\pm 0.111}$
Miss.D.	$0.301_{\pm 0.068}$	$0.357_{\pm 0.067}$	$0.303_{\pm 0.013}$	$0.363_{\pm 0.013}$	$0.341_{\pm 0.016}$	$0.398_{\pm 0.015}$	$0.345_{\pm 0.013}$	$0.400_{\pm 0.012}$
Miss.F.	$0.065_{\pm 0.000}$	$0.092_{\pm 0.000}$	$0.063_{\pm 0.001}$	$\underline{0.093}_{\pm 0.001}$	$\underline{0.063}_{\pm 0.001}$	$\underline{0.094}_{\pm 0.001}$	$\underline{0.064}_{\pm 0.000}$	$\underline{0.095}_{\pm 0.000}$
Sinkhorn	$0.097_{\pm 0.000}$	$0.127_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.137_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.141_{\pm 0.000}$
TRPCA	$\underline{0.056}_{\pm 0.006}$	$\underline{0.084}_{\pm 0.007}$	$0.064_{\pm 0.005}$	$0.094_{\pm 0.006}$	$0.076_{\pm 0.007}$	$0.104_{\pm 0.007}$	$0.097_{\pm 0.015}$	$0.125_{\pm 0.016}$
SCPN	$0.063_{\pm 0.000}$	$0.094_{\pm 0.000}$	$0.066_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.072_{\pm 0.000}$	$0.103_{\pm 0.000}$	$0.085_{\pm 0.000}$	$0.117_{\pm 0.000}$
NewImp	$0.070_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.075_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.083_{\pm 0.000}$	$0.116_{\pm 0.000}$	$0.093_{\pm 0.000}$	$0.126_{\pm 0.000}$
SoftImp	$0.081_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.085_{\pm 0.000}$	$0.115_{\pm 0.000}$	$0.089_{\pm 0.000}$	$0.118_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.131_{\pm 0.000}$
MIWAE	$0.066_{\pm 0.001}$	$0.104_{\pm 0.001}$	$0.064_{\pm 0.001}$	$0.102_{\pm 0.000}$	$0.065_{\pm 0.000}$	$0.102_{\pm 0.001}$	$0.066_{\pm 0.001}$	$0.102_{\pm 0.001}$
Miracle	$0.058_{\pm 0.000}$	$0.090_{\pm 0.000}$	$\underline{0.063}_{\pm 0.000}$	$0.096_{\pm 0.000}$	$0.072_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.112_{\pm 0.000}$	$0.145_{\pm 0.000}$
TDM	$0.070_{\pm 0.002}$	$0.104_{\pm 0.003}$	$0.082_{\pm 0.001}$	$0.119_{\pm 0.001}$	$0.090_{\pm 0.000}$	$0.128_{\pm 0.001}$	$0.097_{\pm 0.000}$	$0.137_{\pm 0.000}$
FMT-I	$\mathbf{0.039}^{*}_{\pm 0.001}$	$\mathbf{0.063}^{*}_{\pm 0.001}$	$\mathbf{0.046}^{*}_{\pm 0.000}$	$\mathbf{0.076}^{*}_{\pm 0.000}$	$\mathbf{0.050}^{*}_{\pm 0.000}$	$\mathbf{0.082}^{*}_{\pm 0.000}$	$\mathbf{0.056}^{*}_{\pm 0.000}$	$\mathbf{0.089}^{*}_{\pm 0.000}$
Missing mechanism: MNAR								
Mean	$0.108_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.111_{\pm 0.000}$	$0.150_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.150_{\pm 0.000}$	$0.111_{\pm 0.000}$	$0.152_{\pm 0.000}$
Median	$0.106_{\pm 0.000}$	$0.149_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.153_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.111_{\pm 0.000}$	$0.157_{\pm 0.000}$
Mode	$0.150_{\pm 0.000}$	$0.206_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.191_{\pm 0.000}$	$0.143_{\pm 0.000}$	$0.199_{\pm 0.000}$	$0.135_{\pm 0.000}$	$0.190_{\pm 0.000}$
ICE	$0.097_{\pm 0.000}$	$0.142_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.160_{\pm 0.000}$	$0.119_{\pm 0.000}$	$0.168_{\pm 0.000}$	$0.140_{\pm 0.000}$	$0.192_{\pm 0.000}$
MICE	$0.075_{\pm 0.000}$	$0.115_{\pm 0.000}$	$0.084_{\pm 0.000}$	$0.125_{\pm 0.000}$	$\underline{0.093}_{\pm 0.000}$	$\underline{0.133}_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$
GAIN	$0.099_{\pm 0.010}$	$0.139_{\pm 0.008}$	$0.110_{\pm 0.005}$	$0.153_{\pm 0.005}$	$0.180_{\pm 0.038}$	$0.229_{\pm 0.041}$	$0.325_{\pm 0.072}$	$0.383_{\pm 0.075}$
Miss.D.	$0.300_{\pm 0.026}$	$0.352_{\pm 0.028}$	$0.333_{\pm 0.031}$	$0.388_{\pm 0.030}$	$0.335_{\pm 0.048}$	$0.389_{\pm 0.048}$	$0.315_{\pm 0.016}$	$0.369_{\pm 0.016}$
Miss.F.	$0.071_{\pm 0.000}$	$0.107_{\pm 0.001}$	$\underline{0.077}_{\pm 0.000}$	$\underline{0.115}_{\pm 0.000}$	$0.100_{\pm 0.003}$	$0.143_{\pm 0.003}$	$\underline{0.103}_{\pm 0.004}$	$0.148_{\pm 0.007}$
Sinkhorn	$0.100_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.109_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.110_{\pm 0.000}$	$0.151_{\pm 0.000}$
TRPCA	$0.071_{\pm 0.008}$	$0.107_{\pm 0.012}$	$0.098_{\pm 0.004}$	$0.140_{\pm 0.007}$	$0.118_{\pm 0.009}$	$0.162_{\pm 0.015}$	$0.150_{\pm 0.025}$	$0.199_{\pm 0.035}$
SCPN	$0.071_{\pm 0.000}$	$0.108_{\pm 0.000}$	$0.089_{\pm 0.000}$	$0.128_{\pm 0.000}$	$0.099_{\pm 0.000}$	$0.141_{\pm 0.000}$	$0.104_{\pm 0.000}$	$\underline{0.146}_{\pm 0.000}$
NewImp	$0.075_{\pm 0.000}$	$0.115_{\pm 0.000}$	$0.087_{\pm 0.000}$	$0.126_{\pm 0.000}$	$0.101_{\pm 0.000}$	$0.141_{\pm 0.000}$	$0.108_{\pm 0.000}$	$0.150_{\pm 0.000}$
SoftImp	$0.093_{\pm 0.000}$	$0.130_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.150_{\pm 0.000}$	$0.129_{\pm 0.000}$	$0.189_{\pm 0.000}$	$0.195_{\pm 0.000}$	$0.279_{\pm 0.000}$
MIWAE	$0.082_{\pm 0.001}$	$0.138_{\pm 0.009}$	$0.093_{\pm 0.007}$	$0.137_{\pm 0.005}$	$0.114_{\pm 0.000}$	$0.157_{\pm 0.001}$	$0.115_{\pm 0.000}$	$0.159_{\pm 0.000}$
Miracle	$\underline{0.066}_{\pm 0.000}$	$\underline{0.106}_{\pm 0.000}$	$0.080_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.096_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.135_{\pm 0.000}$	$0.174_{\pm 0.000}$
TDM	$0.075_{\pm 0.002}$	$0.118_{\pm 0.001}$	$0.090_{\pm 0.001}$	$0.134_{\pm 0.001}$	$0.101_{\pm 0.000}$	$0.143_{\pm 0.001}$	$0.109_{\pm 0.000}$	$0.153_{\pm 0.000}$
FMT-I	$\mathbf{0.046}^{*}_{\pm 0.000}$	$\mathbf{0.079}^{*}_{\pm 0.000}$	$\mathbf{0.067}^{*}_{\pm 0.000}$	$\mathbf{0.111}^{*}_{\pm 0.000}$	$\mathbf{0.080}^{*}_{\pm 0.000}$	$\mathbf{0.125}^{*}_{\pm 0.000}$	$\mathbf{0.094}^{*}_{\pm 0.000}$	$\mathbf{0.138}^{*}_{\pm 0.000}$

Kindly Note: The results are presented as $\mathrm{mean}_{\pm \mathrm{std}}$. Bold typeface highlights the top performance for each metric, while underlined text denotes the second-best results. "*" marks the results where FMT-I significantly outperform the second best method, with p-value< 0.05 in a two-sample t-test.

The resulting imputation performance is illustrated in Fig. 7. **Overall, even as the noise volume increases by a factor of 1000, the imputation error increases by less than a factor of 10.** For instance, at a noise ratio of 0.02, the MAE rises from 0.06 (noise volume = 1) to only 0.36 (noise volume = 1000). It showcases FMT-I's resilience to noise with increased volume.

I. Additional experiments on a temporally contiguous missing scenario

In this section, we examine the performance of FMT-I on a temporally contiguous missing scenario. Specifically, we randomly mask out a contiguous sequence of samples in each feature, maintaining the overall missing ratio as $\mathrm{p_{miss}}$. The resulting imputation performance is available in Table VII, where FMT-I exhibits the best overall performance, showcasing its applicability to scenarios where missingness is temporally contiguous.

J. Additional experiments on other industrial datasets

In this section, we evaluate the performance of FMT-I on an additional dataset collected from a real-world Primary Reformer (PR) unit [4]. In this process, Natural Gas (NG) and high-pressure steam are fed into an array of parallel fixed-tube reactors. The primary objective is to generate Hydrogen

TABLE VI
PERFORMANCE COMPARISON OF FMT-I AND BASELINE MODELS WITH DIFFERENT MISSING RATIOS (p_{miss}) AND NOISE RATIOS (p_{noise}).

Missing ratio	$p_{noise}=0.01$		$p_{noise}=0.03$		$p_{noise}=0.05$		$p_{noise}=0.1$	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Missing ratio: $p_{miss}=0.1$								
Mean	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$
Median	$0.103_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.103_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.103_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.103_{\pm 0.000}$	$0.145_{\pm 0.000}$
Mode	$0.142_{\pm 0.000}$	$0.197_{\pm 0.000}$	$0.156_{\pm 0.000}$	$0.213_{\pm 0.000}$	$0.156_{\pm 0.000}$	$0.213_{\pm 0.000}$	$0.156_{\pm 0.000}$	$0.213_{\pm 0.000}$
ICE	$0.101_{\pm 0.001}$	$0.151_{\pm 0.001}$	$0.108_{\pm 0.001}$	$0.157_{\pm 0.001}$	$0.113_{\pm 0.001}$	$0.161_{\pm 0.001}$	$0.124_{\pm 0.001}$	$0.171_{\pm 0.000}$
MICE	$0.075_{\pm 0.001}$	$0.118_{\pm 0.003}$	$0.079_{\pm 0.001}$	$0.120_{\pm 0.002}$	$0.082_{\pm 0.001}$	$0.121_{\pm 0.002}$	$0.088_{\pm 0.000}$	$0.127_{\pm 0.001}$
GAIN	$0.090_{\pm 0.007}$	$0.130_{\pm 0.010}$	$0.089_{\pm 0.002}$	$0.127_{\pm 0.002}$	$0.087_{\pm 0.004}$	$0.124_{\pm 0.005}$	$0.093_{\pm 0.004}$	$0.133_{\pm 0.002}$
Miss.D.	$0.246_{\pm 0.081}$	$0.286_{\pm 0.093}$	$0.244_{\pm 0.078}$	$0.284_{\pm 0.091}$	$0.245_{\pm 0.079}$	$0.285_{\pm 0.092}$	$0.246_{\pm 0.080}$	$0.286_{\pm 0.092}$
Miss.F	$0.068_{\pm 0.000}$	$0.105_{\pm 0.001}$	$0.068_{\pm 0.001}$	$0.105_{\pm 0.001}$	$0.068_{\pm 0.001}$	$0.106_{\pm 0.001}$	$0.070_{\pm 0.001}$	$0.108_{\pm 0.001}$
Sinkhorn	$0.097_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.133_{\pm 0.000}$
TRPCA	$0.069_{\pm 0.009}$	$0.102_{\pm 0.013}$	$0.072_{\pm 0.011}$	$0.104_{\pm 0.015}$	$0.074_{\pm 0.012}$	$0.105_{\pm 0.016}$	$0.076_{\pm 0.015}$	$0.108_{\pm 0.018}$
SCPN	$\underline{0.066}_{\pm 0.000}$	$\underline{0.100}_{\pm 0.000}$	$\underline{0.066}_{\pm 0.000}$	$\underline{0.101}_{\pm 0.000}$	$\underline{0.066}_{\pm 0.000}$	$\underline{0.101}_{\pm 0.000}$	$\underline{0.067}_{\pm 0.000}$	$\underline{0.103}_{\pm 0.001}$
NewImp	$0.070_{\pm 0.000}$	$0.108_{\pm 0.000}$	$0.070_{\pm 0.000}$	$0.109_{\pm 0.000}$	$0.070_{\pm 0.000}$	$0.109_{\pm 0.000}$	$0.072_{\pm 0.000}$	$0.111_{\pm 0.001}$
SoftImp	$0.088_{\pm 0.000}$	$0.124_{\pm 0.001}$	$0.089_{\pm 0.000}$	$0.125_{\pm 0.001}$	$0.089_{\pm 0.001}$	$0.126_{\pm 0.001}$	$0.091_{\pm 0.001}$	$0.129_{\pm 0.001}$
MIWAE	$0.080_{\pm 0.001}$	$0.132_{\pm 0.004}$	$0.080_{\pm 0.003}$	$0.128_{\pm 0.006}$	$0.081_{\pm 0.001}$	$0.130_{\pm 0.003}$	$0.084_{\pm 0.004}$	$0.129_{\pm 0.002}$
Miracle	$0.069_{\pm 0.001}$	$0.141_{\pm 0.006}$	$0.069_{\pm 0.001}$	$0.125_{\pm 0.008}$	$0.068_{\pm 0.001}$	$0.113_{\pm 0.002}$	$0.070_{\pm 0.001}$	$0.111_{\pm 0.004}$
TDM	$0.068_{\pm 0.001}$	$0.111_{\pm 0.002}$	$0.072_{\pm 0.001}$	$0.115_{\pm 0.001}$	$0.074_{\pm 0.000}$	$0.117_{\pm 0.002}$	$0.077_{\pm 0.001}$	$0.121_{\pm 0.002}$
FMT-I	$\mathbf{0.060}^{*}_{\pm 0.000}$	$\mathbf{0.096}^{*}_{\pm 0.000}$	$\mathbf{0.061}^{*}_{\pm 0.000}$	$\mathbf{0.097}^{*}_{\pm 0.000}$	$\mathbf{0.061}^{*}_{\pm 0.000}$	$\mathbf{0.098}^{*}_{\pm 0.001}$	$\mathbf{0.062}^{*}_{\pm 0.000}$	$\mathbf{0.099}^{*}_{\pm 0.001}$
Missing ratio: $p_{miss}=0.3$								
Mean	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$
Median	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$
Mode	$0.140_{\pm 0.000}$	$0.195_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.211_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.211_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.211_{\pm 0.000}$
ICE	$0.111_{\pm 0.001}$	$0.159_{\pm 0.001}$	$0.117_{\pm 0.001}$	$0.164_{\pm 0.001}$	$0.121_{\pm 0.001}$	$0.168_{\pm 0.001}$	$0.131_{\pm 0.001}$	$0.177_{\pm 0.001}$
MICE	$0.083_{\pm 0.001}$	$0.124_{\pm 0.000}$	$0.086_{\pm 0.001}$	$0.127_{\pm 0.001}$	$0.088_{\pm 0.001}$	$0.129_{\pm 0.001}$	$0.094_{\pm 0.001}$	$0.134_{\pm 0.000}$
GAIN	$0.096_{\pm 0.011}$	$0.136_{\pm 0.011}$	$0.094_{\pm 0.010}$	$0.133_{\pm 0.013}$	$0.094_{\pm 0.009}$	$0.132_{\pm 0.010}$	$0.095_{\pm 0.008}$	$0.133_{\pm 0.010}$
Miss.D.	$0.261_{\pm 0.030}$	$0.311_{\pm 0.035}$	$0.262_{\pm 0.029}$	$0.312_{\pm 0.033}$	$0.262_{\pm 0.028}$	$0.312_{\pm 0.033}$	$0.262_{\pm 0.028}$	$0.311_{\pm 0.032}$
Miss.F	$0.079_{\pm 0.001}$	$0.119_{\pm 0.004}$	$0.078_{\pm 0.001}$	$0.118_{\pm 0.001}$	$0.079_{\pm 0.001}$	$0.120_{\pm 0.001}$	$0.080_{\pm 0.000}$	$0.119_{\pm 0.001}$
Sinkhorn	$0.102_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.101_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.102_{\pm 0.000}$	$0.139_{\pm 0.000}$
TRPCA	$0.089_{\pm 0.005}$	$0.127_{\pm 0.007}$	$0.093_{\pm 0.006}$	$0.129_{\pm 0.008}$	$0.095_{\pm 0.006}$	$0.131_{\pm 0.009}$	$0.098_{\pm 0.008}$	$0.135_{\pm 0.010}$
SCPN	$0.082_{\pm 0.000}$	$0.119_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.120_{\pm 0.000}$	$0.082_{\pm 0.000}$	$0.120_{\pm 0.000}$	$0.083_{\pm 0.000}$	$0.121_{\pm 0.001}$
SoftImp	$0.097_{\pm 0.000}$	$0.136_{\pm 0.000}$	$0.098_{\pm 0.000}$	$0.137_{\pm 0.000}$	$0.098_{\pm 0.001}$	$0.138_{\pm 0.001}$	$0.101_{\pm 0.001}$	$0.142_{\pm 0.001}$
MIWAE	$0.091_{\pm 0.008}$	$0.134_{\pm 0.006}$	$0.096_{\pm 0.014}$	$0.139_{\pm 0.013}$	$0.101_{\pm 0.008}$	$0.143_{\pm 0.009}$	$0.109_{\pm 0.003}$	$0.152_{\pm 0.003}$
Miracle	$\underline{0.074}_{\pm 0.000}$	$\underline{0.114}_{\pm 0.000}$	$\underline{0.075}_{\pm 0.000}$	$\underline{0.115}_{\pm 0.000}$	$\underline{0.075}_{\pm 0.000}$	$\underline{0.116}_{\pm 0.000}$	$\underline{0.078}_{\pm 0.000}$	$\underline{0.118}_{\pm 0.001}$
TDM	$0.084_{\pm 0.001}$	$0.126_{\pm 0.001}$	$0.085_{\pm 0.001}$	$0.127_{\pm 0.001}$	$0.086_{\pm 0.001}$	$0.128_{\pm 0.001}$	$0.089_{\pm 0.000}$	$0.131_{\pm 0.001}$
FMT-I	$\mathbf{0.071}^{*}_{\pm 0.000}$	$\mathbf{0.111}^{*}_{\pm 0.000}$	$\mathbf{0.071}^{*}_{\pm 0.000}$	$\mathbf{0.111}^{*}_{\pm 0.000}$	$\mathbf{0.072}^{*}_{\pm 0.000}$	$\mathbf{0.112}^{*}_{\pm 0.000}$	$\mathbf{0.072}^{*}_{\pm 0.000}$	$\mathbf{0.112}^{*}_{\pm 0.000}$
Missing ratio: $p_{miss}=0.5$								
Mean	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.145_{\pm 0.000}$
Median	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.147_{\pm 0.000}$
Mode	$0.151_{\pm 0.000}$	$0.208_{\pm 0.000}$	$0.151_{\pm 0.000}$	$0.208_{\pm 0.000}$	$0.151_{\pm 0.000}$	$0.208_{\pm 0.000}$	$0.151_{\pm 0.000}$	$0.208_{\pm 0.000}$
ICE	$0.124_{\pm 0.001}$	$0.172_{\pm 0.001}$	$0.129_{\pm 0.001}$	$0.176_{\pm 0.002}$	$0.132_{\pm 0.001}$	$0.179_{\pm 0.001}$	$0.140_{\pm 0.001}$	$0.187_{\pm 0.000}$
MICE	$0.091_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.094_{\pm 0.001}$	$0.135_{\pm 0.001}$	$0.096_{\pm 0.001}$	$0.137_{\pm 0.001}$	$0.100_{\pm 0.001}$	$0.141_{\pm 0.000}$
GAIN	$0.130_{\pm 0.024}$	$0.174_{\pm 0.025}$	$0.138_{\pm 0.039}$	$0.186_{\pm 0.049}$	$0.133_{\pm 0.014}$	$0.179_{\pm 0.017}$	$0.168_{\pm 0.065}$	$0.227_{\pm 0.082}$
Miss.D.	$0.262_{\pm 0.019}$	$0.312_{\pm 0.021}$	$0.263_{\pm 0.019}$	$0.314_{\pm 0.021}$	$0.263_{\pm 0.018}$	$0.314_{\pm 0.021}$	$0.262_{\pm 0.019}$	$0.314_{\pm 0.021}$
Miss.F	$0.092_{\pm 0.000}$	$0.135_{\pm 0.002}$	$0.093_{\pm 0.003}$	$0.138_{\pm 0.006}$	$0.092_{\pm 0.001}$	$0.136_{\pm 0.001}$	$0.092_{\pm 0.001}$	$0.138_{\pm 0.002}$
Sinkhorn	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.144_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.143_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.143_{\pm 0.000}$
TRPCA	$0.109_{\pm 0.009}$	$0.150_{\pm 0.014}$	$0.111_{\pm 0.010}$	$0.152_{\pm 0.015}$	$0.113_{\pm 0.010}$	$0.154_{\pm 0.015}$	$0.116_{\pm 0.012}$	$0.157_{\pm 0.016}$
SCPN	$0.092_{\pm 0.000}$	$0.133_{\pm 0.000}$	$0.093_{\pm 0.000}$	$0.134_{\pm 0.000}$	$0.093_{\pm 0.000}$	$0.134_{\pm 0.000}$	$0.094_{\pm 0.000}$	$0.135_{\pm 0.000}$
NewImp	$0.097_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.138_{\pm 0.000}$	$0.097_{\pm 0.000}$	$0.138_{\pm 0.000}$
SoftImp	$0.119_{\pm 0.000}$	$0.170_{\pm 0.000}$	$0.120_{\pm 0.000}$	$0.172_{\pm 0.000}$	$0.121_{\pm 0.000}$	$0.173_{\pm 0.000}$	$0.124_{\pm 0.000}$	$0.177_{\pm 0.000}$
MIWAE	$0.111_{\pm 0.000}$	$0.154_{\pm 0.001}$	$0.111_{\pm 0.000}$	$0.155_{\pm 0.001}$	$0.112_{\pm 0.000}$	$0.155_{\pm 0.000}$	$0.112_{\pm 0.000}$	$0.156_{\pm 0.000}$
Miracle	$\underline{0.090}_{\pm 0.000}$	$\underline{0.132}_{\pm 0.000}$	$\underline{0.091}_{\pm 0.000}$	$\underline{0.133}_{\pm 0.001}$	$\underline{0.091}_{\pm 0.000}$	$\underline{0.133}_{\pm 0.001}$	$\underline{0.092}_{\pm 0.001}$	$\underline{0.133}_{\pm 0.001}$
TDM	$0.094_{\pm 0.000}$	$0.136_{\pm 0.001}$	$0.095_{\pm 0.001}$	$0.137_{\pm 0.001}$	$0.097_{\pm 0.001}$	$0.139_{\pm 0.001}$	$0.099_{\pm 0.001}$	$0.141_{\pm 0.001}$
FMT-I	$\mathbf{0.082}^{*}_{\pm 0.000}$	$\mathbf{0.123}^{*}_{\pm 0.000}$	$\mathbf{0.083}^{*}_{\pm 0.000}$	$\mathbf{0.123}^{*}_{\pm 0.000}$	$\mathbf{0.083}^{*}_{\pm 0.000}$	$\mathbf{0.124}^{*}_{\pm 0.000}$	$\mathbf{0.084}^{*}_{\pm 0.000}$	$\mathbf{0.125}^{*}_{\pm 0.001}$
Missing ratio: $p_{miss}=0.7$								
Mean	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$
Median	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$
Mode	$0.153_{\pm 0.000}$	$0.215_{\pm 0.000}$	$0.153_{\pm 0.000}$	$0.215_{\pm 0.000}$	$0.153_{\pm 0.000}$	$0.215_{\pm 0.000}$	$0.153_{\pm 0.000}$	$0.215_{\pm 0.000}$
ICE	$0.141_{\pm 0.001}$	$0.192_{\pm 0.001}$	$0.144_{\pm 0.001}$	$0.195_{\pm 0.001}$	$0.147_{\pm 0.000}$	$0.197_{\pm 0.001}$	$0.153_{\pm 0.000}$	$0.203_{\pm 0.000}$
MICE	$0.105_{\pm 0.000}$	$0.147_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.148_{\pm 0.001}$	$0.108_{\pm 0.000}$	$0.149_{\pm 0.000}$	$0.111_{\pm 0.001}$	$0.152_{\pm 0.000}$
GAIN	$0.328_{\pm 0.072}$	$0.383_{\pm 0.070}$	$0.357_{\pm 0.120}$	$0.415_{\pm 0.121}$	$0.362_{\pm 0.121}$	$0.418_{\pm 0.128}$	$0.373_{\pm 0.131}$	$0.423_{\pm 0.136}$
Miss.D.	$0.275_{\pm 0.020}$	$0.328_{\pm 0.022}$	$0.276_{\pm 0.019}$	$0.329_{\pm 0.021}$	$0.277_{\pm 0.019}$	$0.330_{\pm 0.021}$	$0.277_{\pm 0.019}$	$0.330_{\pm 0.021}$
Miss.F	$0.105_{\pm 0.003}$	$0.150_{\pm 0.002}$	$\underline{0.101}_{\pm 0.002}$	$\underline{0.145}_{\pm 0.002}$	$0.108_{\pm 0.005}$	$0.155_{\pm 0.006}$	$0.110_{\pm 0.006}$	$0.159_{\pm 0.009}$
Sinkhorn	$0.106_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.146_{\pm 0.000}$
TRPCA	$0.136_{\pm 0.023}$	$0.182_{\pm 0.032}$	$0.137_{\pm 0.023}$	$0.184_{\pm 0.032}$	$0.139_{\pm 0.024}$	$0.185_{\pm 0.032}$	$0.142_{\pm 0.024}$	$0.188_{\pm 0.033}$
SCPN	$\underline{0.102}_{\pm 0.000}$	$\underline{0.146}_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.147_{\pm 0.000}$	$\underline{0.104}_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.000}$	$0.150_{\pm 0.001}$
NewImp	$0.104_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.146_{\pm 0.000}$	$0.104_{\pm 0.000}$	$\underline{0.146}_{\pm 0.000}$	$\underline{0.105}_{\pm 0.000}$	$\underline{0.146}_{\pm 0.000}$
SoftImp	$0.175_{\pm 0.000}$	$0.252_{\pm 0.000}$	$0.176_{\pm 0.000}$	$0.253_{\pm 0.001}$	$0.178_{\pm 0.000}$	$0.255_{\pm 0.001}$	$0.181_{\pm 0.001}$	$0.259_{\pm 0.002}$
MIWAE	$0.112_{\pm 0.000}$	$0.155_{\pm 0.000}$	$0.112_{\pm 0.000}$	$0.155_{\pm 0.000}$	$0.112_{\pm 0.000}$	$0.155_{\pm 0.000}$	$0.113_{\pm 0.000}$	$0.156_{\pm 0.000}$
Miracle	$0.128_{\pm 0.000}$	$0.165_{\pm 0.000}$	$0.128_{\pm 0.001}$	$0.165_{\pm 0.001}$	$0.128_{\pm 0.001}$	$0.164_{\pm 0.000}$	$0.128_{\pm 0.001}$	$0.164_{\pm 0.001}$
TDM	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.105_{\pm 0.001}$	$0.148_{\pm 0.001}$	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.106_{\pm 0.000}$	$0.149_{\pm 0.000}$
FMT-I	$\mathbf{0.100}^{*}_{\pm 0.000}$	$\mathbf{0.140}^{*}_{\pm 0.000}$	$\mathbf{0.100}_{\pm 0.000}$	$\mathbf{0.140}^{*}_{\pm 0.000}$	$\mathbf{0.100}^{*}_{\pm 0.000}$	$\mathbf{0.140}^{*}_{\pm 0.000}$	$\mathbf{0.101}^{*}_{\pm 0.000}$	$\mathbf{0.141}^{*}_{\pm 0.000}$

Kindly Note: The results are presented as $\mathrm{mean}_{\pm \mathrm{std}}$. Bold typeface highlights the top performance for each metric, while underlined text denotes the second-best results. "*" marks the results where FMT-I significantly outperform the second best method, with p-value< 0.05 in a two-sample t-test.

TABLE VII
PERFORMANCE OF FMT-I VERSUS BASELINE MODELS IN TEMPORALLY CONTIGUOUS MISSING SCENARIO.

Missing ratio	p = 0.1		p = 0.3		p = 0.5		p = 0.7	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Mean	$0.131_{\pm 0.000}$	$0.175_{\pm 0.000}$	$0.123_{\pm 0.000}$	$0.172_{\pm 0.000}$	$0.118_{\pm 0.000}$	$0.167_{\pm 0.000}$	$0.124_{\pm 0.000}$	$0.168_{\pm 0.000}$
Median	$0.135_{\pm 0.000}$	$0.182_{\pm 0.000}$	$0.130_{\pm 0.000}$	$0.182_{\pm 0.000}$	$0.124_{\pm 0.000}$	$0.174_{\pm 0.000}$	$0.127_{\pm 0.000}$	$0.173_{\pm 0.000}$
Mode	$0.167_{\pm 0.000}$	$0.221_{\pm 0.000}$	$0.156_{\pm 0.000}$	$0.213_{\pm 0.000}$	$0.158_{\pm 0.000}$	$0.209_{\pm 0.000}$	$0.216_{\pm 0.000}$	$0.283_{\pm 0.000}$
ICE	$0.103_{\pm 0.000}$	$0.157_{\pm 0.000}$	$0.121_{\pm 0.000}$	$0.173_{\pm 0.000}$	$0.132_{\pm 0.000}$	$0.186_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.199_{\pm 0.000}$
MICE	$0.091_{\pm 0.000}$	$0.141_{\pm 0.000}$	$0.104_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.113_{\pm 0.000}$	$0.163_{\pm 0.000}$	$0.125_{\pm 0.000}$	$0.173_{\pm 0.000}$
GAIN	$0.091_{\pm 0.002}$	$0.135_{\pm 0.008}$	$0.116_{\pm 0.005}$	$0.163_{\pm 0.005}$	$0.274_{\pm 0.020}$	$0.340_{\pm 0.023}$	$0.235_{\pm 0.047}$	$0.298_{\pm 0.043}$
Miss.D.	$0.217_{\pm 0.001}$	$0.252_{\pm 0.015}$	$0.237_{\pm 0.029}$	$0.280_{\pm 0.042}$	$0.244_{\pm 0.034}$	$0.293_{\pm 0.048}$	$0.256_{\pm 0.019}$	$0.314_{\pm 0.027}$
Miss.F	$0.082_{\pm 0.000}$	$\underline{0.125}_{\pm 0.001}$	$\underline{0.093}_{\pm 0.000}$	$\mathbf{0.142}_{\pm 0.001}$	$0.113_{\pm 0.001}$	$0.159_{\pm 0.000}$	$0.132_{\pm 0.012}$	$0.183_{\pm 0.017}$
Sinkhorn	$0.120_{\pm 0.000}$	$0.163_{\pm 0.000}$	$0.117_{\pm 0.001}$	$0.166_{\pm 0.001}$	$0.116_{\pm 0.000}$	$0.164_{\pm 0.000}$	$0.123_{\pm 0.000}$	$0.167_{\pm 0.000}$
SoftImp	$0.105_{\pm 0.000}$	$0.148_{\pm 0.000}$	$0.123_{\pm 0.000}$	$0.177_{\pm 0.000}$	$0.142_{\pm 0.000}$	$0.200_{\pm 0.000}$	$0.154_{\pm 0.000}$	$0.208_{\pm 0.000}$
MIWAE	$0.103_{\pm 0.015}$	$0.151_{\pm 0.014}$	$0.110_{\pm 0.005}$	$0.160_{\pm 0.005}$	$\underline{0.110}_{\pm 0.012}$	$\underline{0.159}_{\pm 0.013}$	$0.124_{\pm 0.005}$	$0.170_{\pm 0.005}$
Miracle	$\underline{0.078}_{\pm 0.000}$	$0.141_{\pm 0.000}$	$0.109_{\pm 0.000}$	$0.256_{\pm 0.000}$	$0.145_{\pm 0.000}$	$0.283_{\pm 0.000}$	$0.266_{\pm 0.000}$	$0.450_{\pm 0.000}$
TDM	$0.108_{\pm 0.001}$	$0.165_{\pm 0.002}$	$0.111_{\pm 0.001}$	$0.167_{\pm 0.001}$	$0.112_{\pm 0.001}$	$0.163_{\pm 0.000}$	$0.122_{\pm 0.000}$	$0.167_{\pm 0.000}$
TRPCA	$0.086_{\pm 0.000}$	$0.135_{\pm 0.000}$	$0.118_{\pm 0.000}$	$0.172_{\pm 0.000}$	$0.139_{\pm 0.000}$	$0.192_{\pm 0.000}$	$0.169_{\pm 0.000}$	$0.220_{\pm 0.000}$
SCPN	$0.081_{\pm 0.000}$	$0.127_{\pm 0.000}$	$0.107_{\pm 0.000}$	$0.166_{\pm 0.000}$	$0.123_{\pm 0.000}$	$0.176_{\pm 0.000}$	$0.120_{\pm 0.000}$	$\underline{0.167}_{\pm 0.000}$
NewImp	$0.094_{\pm 0.006}$	$0.164_{\pm 0.028}$	$0.104_{\pm 0.001}$	$0.171_{\pm 0.007}$	$0.112_{\pm 0.001}$	$0.169_{\pm 0.004}$	$\underline{0.118}_{\pm 0.000}$	$0.168_{\pm 0.000}$
FMT-I	$\mathbf{0.076}_{\pm 0.002}$	$\mathbf{0.122}_{\pm 0.002}$	$\mathbf{0.093}_{\pm 0.000}$	$\underline{0.149}_{\pm 0.001}$	$\mathbf{0.101}_{\pm 0.001}$	$\mathbf{0.154}_{\pm 0.001}$	$\mathbf{0.106}_{\pm 0.000}$	$\mathbf{0.154}_{\pm 0.000}$

[1] *Kindly Note*: The presented results represent the average from several trials. **Bold** typeface highlights the top performance for each metric, while underlined text denotes the second-best results.

TABLE VIII
PERFORMANCE OF FMT-I VERSUS BASELINE MODELS WITH FIXED MISSING RATIO $p_{miss} = 0.1$ AND NOISE RATIO $p_{noise} = 0.1$ ON PR DATASET.

Model	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	Y
Metric: RMSE														
Mean	0.150	0.132	0.123	0.167	0.136	0.164	0.149	0.133	0.109	0.158	0.117	0.156	0.170	0.156
Median	0.149	0.132	0.119	0.167	0.137	0.167	0.150	0.134	0.108	0.158	0.117	0.155	0.169	0.157
Mode	0.158	0.162	0.126	0.475	0.181	0.207	0.152	0.152	0.145	0.179	0.135	0.158	0.338	0.466
ICE	0.637	0.678	0.707	0.631	0.642	0.644	0.638	0.667	0.654	0.626	0.665	0.695	0.666	0.651
MICE	0.362	0.373	0.358	0.369	0.390	0.371	0.370	0.377	0.343	0.367	0.353	0.416	0.390	0.371
GAIN	0.231	0.191	0.237	0.295	0.245	0.294	0.237	0.350	0.243	0.267	0.264	0.300	0.371	0.342
Miss.D.	0.274	0.267	0.272	0.246	0.250	0.302	0.269	0.248	0.279	0.307	0.263	0.258	0.245	0.237
Miss.F	0.175	0.163	0.165	0.197	0.193	0.181	0.167	0.171	0.163	0.180	0.148	0.193	0.164	0.182
Sinkhorn	0.149	0.125	0.124	0.167	0.141	0.159	0.147	0.132	0.119	0.147	0.123	0.157	0.160	0.168
SoftImp	0.306	0.316	0.327	0.267	0.321	0.280	0.266	0.232	0.325	0.269	0.403	0.612	0.208	0.262
MIWAE	0.284	0.321	0.313	0.279	0.293	0.298	0.293	0.234	0.286	0.344	0.277	0.346	0.281	0.353
Miracle	0.592	0.605	0.634	0.573	0.651	0.468	0.588	0.441	0.601	0.583	0.559	0.532	0.451	0.603
TDM	0.146	0.134	0.126	0.177	0.144	0.151	0.145	0.130	0.118	0.149	0.125	0.145	0.147	0.177
TRPCA	0.213	0.189	0.190	0.192	0.155	0.213	0.193	0.168	0.216	0.218	0.193	0.199	0.163	0.198
SCPN	0.081	0.066	0.081	0.090	0.081	0.098	0.090	0.081	0.083	0.075	0.086	0.086	0.081	0.120
NewImp	0.305	0.246	0.361	0.320	0.528	0.243	0.304	0.202	0.403	0.406	0.295	0.400	0.265	0.451
FMT-I	**0.072**	**0.056**	**0.066**	**0.086**	**0.080**	**0.089**	**0.082**	**0.071**	**0.080**	**0.071**	**0.077**	**0.077**	**0.081**	**0.117**
Metric: MAE														
Mean	0.121	0.089	0.091	0.138	0.107	0.125	0.117	0.100	0.082	0.111	0.093	0.127	0.140	0.123
Median	0.121	0.088	0.088	0.138	0.107	0.125	0.117	0.097	0.081	0.106	0.093	0.126	0.139	0.125
Mode	0.123	0.115	0.092	0.444	0.136	0.177	0.123	0.111	0.118	0.117	0.108	0.127	0.297	0.439
ICE	0.505	0.554	0.564	0.493	0.511	0.523	0.501	0.534	0.514	0.498	0.536	0.550	0.548	0.516
MICE	0.284	0.292	0.281	0.288	0.307	0.296	0.284	0.300	0.262	0.286	0.285	0.323	0.317	0.302
GAIN	0.177	0.126	0.154	0.203	0.187	0.213	0.163	0.290	0.160	0.193	0.193	0.210	0.313	0.260
Miss.D.	0.237	0.240	0.248	0.206	0.224	0.266	0.233	0.218	0.257	0.275	0.239	0.220	0.208	0.202
Miss.F	0.132	0.101	0.109	0.151	0.124	0.133	0.124	0.109	0.100	0.123	0.104	0.140	0.128	0.137
Sinkhorn	0.120	0.086	0.090	0.137	0.111	0.121	0.113	0.093	0.085	0.106	0.092	0.124	0.129	0.131
SoftImp	0.175	0.143	0.153	0.178	0.184	0.169	0.144	0.115	0.152	0.151	0.190	0.288	0.153	0.174
MIWAE	0.172	0.136	0.145	0.166	0.160	0.165	0.173	0.111	0.137	0.166	0.149	0.201	0.182	0.204
Miracle	0.233	0.203	0.220	0.235	0.256	0.190	0.206	0.146	0.214	0.212	0.187	0.210	0.161	0.275
TDM	0.109	0.081	0.088	0.140	0.104	0.104	0.109	0.093	0.082	0.103	0.095	0.107	0.113	0.136
TRPCA	0.162	0.157	0.150	0.147	0.117	0.170	0.154	0.127	0.174	0.187	0.161	0.161	0.133	0.150
SCPN	0.037	0.024	0.033	**0.043**	0.043	0.040	0.039	0.031	0.037	0.036	**0.031**	0.036	**0.036**	**0.062**
NewImp	0.111	0.084	0.123	0.130	0.197	0.092	0.107	0.064	0.147	0.144	0.102	0.144	0.096	0.198
FMT-I	**0.034**	**0.024**	**0.030**	0.045	**0.043**	**0.039**	**0.037**	**0.030**	**0.037**	**0.036**	0.032	**0.036**	0.039	0.064

[1] *Kindly Note*: The presented results represent the average from several trials. **Bold** typeface highlights the top performance for each metric, while underlined text denotes the second-best results.

through the following reforming reactions:

$$\begin{cases} \mathrm{C}_n\mathrm{H}_{2n+2} + n\mathrm{H}_2\mathrm{O} \overset{\text{Heat}}{\leftrightarrow} n\mathrm{CO}_2 + (2n+1)\mathrm{H}_2 \\ \mathrm{CH}_4 + \mathrm{H}_2\mathrm{O} \overset{\text{Heat}}{\leftrightarrow} \mathrm{CO} + 3\mathrm{H}_2 \\ \mathrm{CO} + \mathrm{H}_2\mathrm{O} \overset{\text{Heat}}{\leftrightarrow} \mathrm{CO}_2 + \mathrm{H}_2 \end{cases}, \quad (7)$$

where $\mathrm{C}_n\mathrm{H}_{2n+2}$ indicates alkane, CH_4 is methane, and CO is carbon monoxide. Real-time monitoring of the oxygen content at the outlet of the low-temperature bed is needed in this process. To this end, a dedicated monitoring dataset was constructed, comprising a total of 3,200 samples and 14 process features.

The comparison of imputation performance on this dataset is available in Table VIII, where both missing ratio and noise ratio are set to 0.1. FMT-I still outperforms other baseline methods and achieves the best performance, which further showcases the utility of FMT-I in real-world industrial datasets.

K. Additional description of baseline methods

In this section, we provide a brief overview of the baseline methods used in this study.

- **Mean, Median, Mode** fill missing values with the mean, median, or mode of the observed values, respectively.
- **ICE** [11] treats each feature with missing data as a target variable, modeling it as a function of other observed features using iterative regression.
- **MICE** [11] extends ICE by involving multiple imputation techniques.
- **Miss.F.** [18] is an iterative imputation method that employs random forest models, leveraging the capacity of tree-based models to capture nonlinear feature dependencies.
- **SoftImp** [25] performs matrix completion via soft-thresholded singular value decomposition, regularized by a nuclear norm.
- **MIRACLE** [27] is a causally-aware imputation method that refines imputations by simultaneously learning the causal graph structure and the missing data mechanism, ensuring consistency with the underlying data generation process.
- **Sinkhorn** [50] is a discrepancy-based imputation method that minimizes distributional discrepancy between mini-batch data, and the discrepancy is quantified using standard optimal transport.
- **TDM** [49] builds upon the Sinkhorn framework by utilizing reversible neural networks (e.g., Normalizing Flows) to perform distribution alignment in a latent space.
- **TRPCA** [30] and **SCPN** [31] decompose the data matrix into low-rank and sparse components to isolate outliers, and utilize the low-rank components to perform imputation.
- **GAIN** [14] performs adversarial training between a generator that imputes missing values and a discriminator that distinguishes observed from imputed entries.
- **MIWAE** [26] integrates deep latent variable models for achieving multiple imputation.
- **Miss.D.** [12] models imputation as a reverse diffusion process, gradually denoising random noise conditioned on the observed data.
- **NewImp** [39] is an advanced diffusion-based model that enhances imputation quality by eliminating the implicit entropic regularization term, thereby reducing the uncertainty associated with the imputed values.