

## Research Article

# Regularized Multioutput Gaussian Convolution Process for Chemical Contents Data Imputation in Sintering Raw Materials

Wei Liu <sup>1,2,3</sup> Cailian Chen <sup>1,2,3</sup> Junpeng Li,<sup>4,5</sup> and Xinping Guan<sup>1,2,3</sup>

<sup>1</sup>Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

<sup>3</sup>Shanghai Engineering Research Center of Intelligent Control and Management, Shanghai 200240, China

<sup>4</sup>Engineering Research Center of Intelligent Control System and Intelligent Equipment Ministry of Education of China, Qinhuangdao 066004, China

<sup>5</sup>Faculty of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China

Correspondence should be addressed to Cailian Chen; [cailianchen@sjtu.edu.cn](mailto:cailianchen@sjtu.edu.cn)

Received 31 May 2023; Revised 7 August 2023; Accepted 23 August 2023; Published 23 October 2023

Academic Editor: Sourabh Sahu

Copyright © 2023 Wei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chemical contents, the important quality indicators are crucial for the modeling of sintering process. However, the lack of these data can result in the biasedness of state estimation in sintering process. It, thus, greatly reduces the accuracy of modeling. Although there are some general imputation methods to tackle the data lackness, they rarely consider the interoutputs correlation and the negative impacts caused by incorrect prefilling. In this article, a novel sparse multioutput Gaussian convolution process (MGCP) modeling framework is proposed for data imputation. MGCP can flexibly mine the relationships between the outputs by a convolution of a sharing latent function and different Gaussian kernels. Moreover, the penalization terms are designed to weaken the false relationship between these outputs. To generalize the MGCP to a long-period case, dynamic time warping term is introduced to keep the global similarity between the original and estimated time series. Compared with several existing methods, the proposed method shows great superiority in sintering raw material contents estimation with real-world data.

## 1. Introduction

Sintering process is widely used to manufacture sinter ore, which is one of the main raw materials for pig iron. As illustrated in Figure 1, the iron ore mix, dolomite, limestone, and return sinter are put together as the raw material of sintering process. These materials are mixed, ignited, sintered, crushed, and sorted in this process [1]. As the foundation of the iron making, the automation of sintering process is of great significance.

Since sintering process is a complex thermochemical process, the control process heavily relies on the operators' experiments. In order to achieve the automatic sintering, some expert systems have been first implemented to simulate the experts' decisions and experience. For example, Kawasaki Steel Company in Japan developed a diagnostic expert system to predict the burning through point based on the permeability data of raw material [1]. This approach greatly

improves the stability of sintering process, and the fluctuation in the determination of the fluctuation of the sintering endpoint position has decreased from 7% to 3%. Since manual water flooding brings a great variation in the moisture content of the raw materials, Jiang et al. [2] combined the deep learning and autoregressive model to predict the moisture efficiently. Additionally, Chen et al. [3] proposed a factor dynamic autoregressive hidden variables model to monitor the states of raw materials and detect the process abnormalities. However, due to the complex process of sintering process, it still remain large quantities of issues for further improvements, such as high-precision modeling with a large amount of imperfect data.

The chemical contents of the raw materials are measured and used to predict some key quality indicators of the sinter [4]. The accurate acquisition of the chemical contents helps to provide more reasonable dosing scheme, thereby improving the sintering quality and reducing the procurement cost

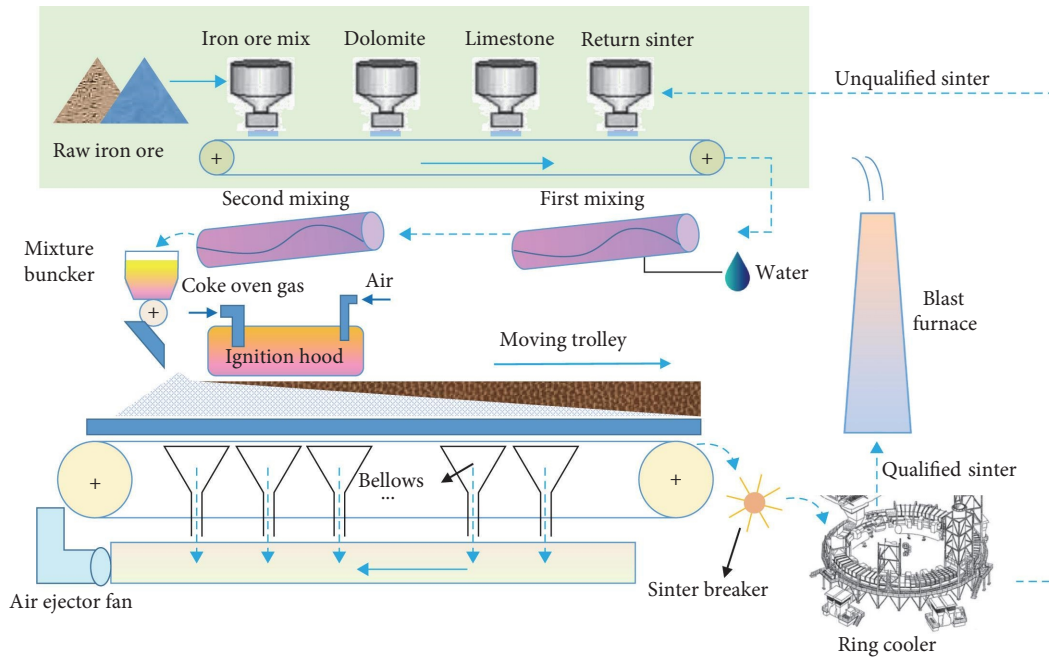


FIGURE 1: Schematic diagram of sintering process.

of sintering raw materials. However, the lack of raw material data is inevitable in actual production. These collected data of the raw materials are measured with different time intervals by different sampling frequencies. Moreover, the delayed uploading of these data often occurs due to the large amount of test samples. These unknown raw material contents may lead to poor quality of sinter ore and excessive emissions of pollution gases.

The lackness of chemical contents data can be filled with some substituted values, known as data imputation. In the past few decades, many data-driven methods have been proposed to address the imputation of missing values in industrial and engineering problems [5–9]. The most simple solutions are to impute the missing values with some statistical values, such as mean, median, and the last observation [10]. But these approaches ignore the inner correlation among these samples, it can not provide good performance when the missing situation is complex. In order to make full use of the global consistency among the time series, smoothing methods [11, 12] are proposed with easy implement. For multivariate time series, the relations between variables are utilized to recover the missing data. Based on the assumption that the missing values are closer to those samples with higher similarity, K-nearest neighbor (KNN) is adopted. Stockmann et al. [13] imputed the single missing data from  $k$  most similar samples of a multidimensional time series. This method was used for the imputation of time delay estimation and the effectiveness was demonstrated. Zhang et al. [14] proposed a matrix/tensor factorization method for the imputation of multidimensional data in structural health monitoring. These high-dimensional data were decomposed into low-dimensional factors so that the computation efficiency was highly promoted. Moreover, the prefilling-prediction methods are proposed for the data

imputation by prefilling these missing values, and then the missing values are refreshed through the prediction model until the prediction errors meet the requirements [15, 16]. Compared with other approaches, this method takes full use of the relationships among these inputs and outputs instead of making no distinction among these features. In Che et al.'s [16] study, the mean value and the last observation of the time series data are merged and regarded as the initial missing values, and gate recurrent unit-based approach was developed to predict the target labels. However, many of these works assumed that the missing rate is relatively small. Additionally, the patterns of missingness greatly influence the imputation accuracy, such as sparsity and sequence length, and these characteristics need to be better studied.

Although these methods provide the insights into the imputation of the chemical contents, it has to deal with some challenges:

- (1) The missing ratio of chemical contents of sinter raw materials can be over 60%, while most studies focus on the situations where the miss rate is below 50%.
- (2) The sintering process is a complex nonlinear process. Interfeature relationships are difficult to determine.
- (3) The interval length between two missing data points may vary.

Wu et al. [17] and Hu et al. [18] took into consider the issue of missing data on these chemical contents when they established the burdening optimization model for sintering. The missing values were substituted by the last observed value because this method is easy to implement. However, the above approach ignores the relationship among these variables. From the sintering process, the chemical contents are related to each other. In order to keep the high-basidity

environment, there is a proportional constraint between these alkaline substances and silica ( $\text{SiO}_2$ ). In addition, the sinter quality indicators are related to these chemical contents of raw materials. If these variables are not taken into consideration, the accuracy is difficult to make sure. With the development of advanced intelligent algorithms, the missing data in industry are gradually receiving more attention [19–23].

To this end, the multioutput Gaussian process (MGP) approaches are explored for dealing with the data imputation. The inspiration comes from the empirical rule that the richer information can be obtained with more related outputs when the observations of the current output are rare. GP regression model is a nonparametric method and shows excellent fitting capability for various functions [24, 25]. Recently, MGP has attracted increasing attention used for a multitask learning model with the ability to describe the correlation among different outputs [26], which has been widely employed in the capacity forecast of battery cells [27], traffic modeling [28], WiFi sensor system [29], and multienvironmental trial [30]. For instance, Rodrigues et al. [28] proposed the Bayesian nonparametric formalism-based MGP to model the complex spatial and temporal relationships in crowd-sourced traffic data. Hori et al. [30] combined MGP model with self-measuring similarity kernels to reflect correlation among genotypes, traits, and environments. The performance of MGP behaved better than regularized PCA. However, the majority of these approaches assume the correlations between variables are very strong, which are incapable of solving fragile relationships between variables introduced by large number of missing output values.

In this article, we propose a new regularized MGP with convolution covariance (MGCP) framework for the imputation of chemical contents of sintering raw materials. First, the prefilling strategy based on GP is adopted to obtain the missing values of the related outputs. Second, two novel MGCP methods are proposed to establish the models between the inputs and multiple outputs. In addition, the performance of the proposed methods are validated with the simulation case and the chemical contents dataset measured from the actual operating sintering plant. To the best of my knowledge, we are the first to systematically study the imputation of these missing values. The contributions of this article can be summarized as follows:

- (1) A new regularized MGCP-based structure is proposed for data imputation with high missing ratio, including the prefilling and reconstruction of the outputs, the model establishment based on regularized MGCP, and the implement of the corresponding optimization algorithm. This approach is applied to impute the missing outputs values and make full use of the relationships between the outputs.
- (2) The proposed MGCP approach improves the model accuracy with the knowledge transferred from the related outputs and greatly reduces the negative impact caused by filling a large number of missing values. Moreover, the dynamic time warping (DTW) method is applied to maintain high similarity with the measured time series, so as to avoid the overfitting.
- (3) The performance of the proposed method is verified by the chemical contents of sintering raw materials measured from a running steel industry. Compared with other imputation methods, the new method achieves the best hit rate.

The rest of this article is organized as follows. We briefly introduce some basic knowledge about MGCP. Then, a novel MGCP modeling framework for the imputation is presented. The proposed model is verified by the simulated data and the actual sintering raw materials data. Finally, the conclusion of this article is given.

## 2. Preliminaries of MGCP

In this section, we introduce the basic knowledge of MGCP. A GP is defined as a collection of random variables sampled from a joint Gaussian distribution model [25]. Given the input variables  $\mathbf{x} \in \mathbb{R}$  and the covariance function  $\Theta(\mathbf{x}, \mathbf{x}')$ , a GP can be described as  $\Phi(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \Theta(\mathbf{x}, \mathbf{x}'))$ , in which a GP is regarded as a model established by recovering the shared underlying function parameterized through a covariance matrix. Considering that the convolution between a Gaussian white noise process  $R(\mathbf{x})$  and a smoothing kernel  $Q(\mathbf{x})$  is still a GP, a Gaussian convolution process (GCP) is reconstructed through a CP. The covariance function of a GCP is as follows:

$$\Phi(\mathbf{x}) = Q(\mathbf{x}) \star R(\mathbf{x}) = \int_{-\infty}^{\infty} Q(\mathbf{x} - \mathbf{u})R(\mathbf{u})d\mathbf{u} , \quad (1)$$

where  $\star$  is a convolution operation. The parameters of the covariance can be estimated by the following optimization problem:

$$\min_{\theta, \sigma_\epsilon} \mathcal{L}(\mathbf{y}|\mathbf{x}; \theta, \sigma_\epsilon) = -\log P(\mathbf{y}|\mathbf{x}; \theta, \sigma_\epsilon). \quad (2)$$

While we extend to solve the problems of the multiple outputs, GP models constructed with the cross-covariance among different outputs are required. However, the covariance structure design for MGPs remains challenging [31]. The models with a separable covariance structure are the most common approaches, including linear model of coregionalization [32], and intrinsic coregionalization model [33]. However, the expression of the covariance is not suitable for borrowing strength from different outputs since they have the same covariance structure.

Nonseparable models are designed to overcome this limitation. The CP is applied to construct the MGCP with its flexible structure, denoted as MGCP. MGCP has two parts, including the shared latent process and different smoothing kernels. Compared to separable models, the structure of the MGCP is more flexible since the interaction between different outputs is available by different smoothing kernels, which make it possible to capture the intrinsic relatedness and bias to improve the accuracy. Suppose that we have the number of latent variables  $l$ , a base GP  $R(\mathbf{u})$  and the kernel  $Q(\mathbf{x})$ , and then the  $j$ th covariance is as follows:

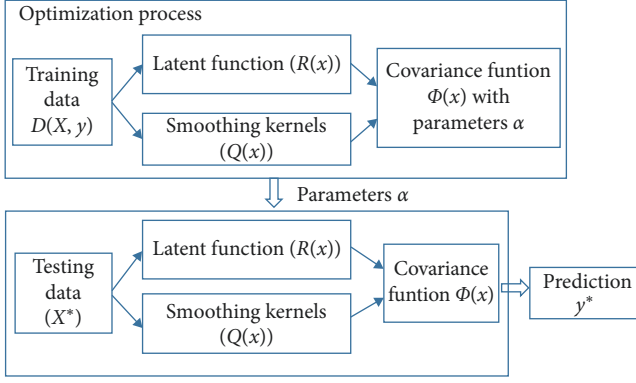


FIGURE 2: The framework of MGCP.

$$\Phi_j(\mathbf{x}) = \sum_{i=1}^q \int_{-\infty}^{\infty} Q_{ji}(\mathbf{x} - \mathbf{u}) R_i(\mathbf{u}) d\mathbf{u} . \quad (3)$$

The framework of MGCP are illustrated in Figure 2. The basic idea of MGCP is to construct the covariance by the shared latent process and different smoothing kernels. In the training process, the latent function and smoothing kernels are calculated according to their definitions. The latent function is a Gaussian white noise function, and the kernel can be a Gaussian kernel function. Furthermore, the covariance can be obtained by Equation (3). The objection function (Equation (2)) is implemented to realize the identification of different parameters. For the new input point  $x_*$ , based on multivariate normal theory, the posterior distribution of the target output  $f_t(x_*)$  can be expressed as follows [34]:

$$f_t(x_*) | \mathbf{X}, y \sim \mathcal{N}(\mu(x_*), \Phi_j(x_*)) , \quad (4)$$

where  $\mu(x_*)$  is the mean value of prediction and  $\Phi_j(x_*)$  is the variance.

Kasarla et al. [35] considered the complexity of the outputs, and proposes a MGCP model that each output can have its own unique feature. Therefore, they separate each output into the shared part and the independent part. This method improve the accuracy of the model, but the computation burden also increases. Wang et al. [34] established a MGCP to deal with the inconsistent input domain. The method marginalizes the inconsistent features to realize the domain adaptation. However, these methods do not consider the existence of large quantities of missing values, which will be discussed in the following sections.

### 3. Novel MGCP Model for Data Imputation

In this section, we proposed a sparse MGCP model used for data imputation. First, a MGCP model framework is established with a sparse covariance structure, and the covariance is calculated by the characteristics of the latent functions and kernels. Then, the proposed MGCP is applied for the imputation of missing values. The corresponding optimization problem of the proposed model is effectively completed.

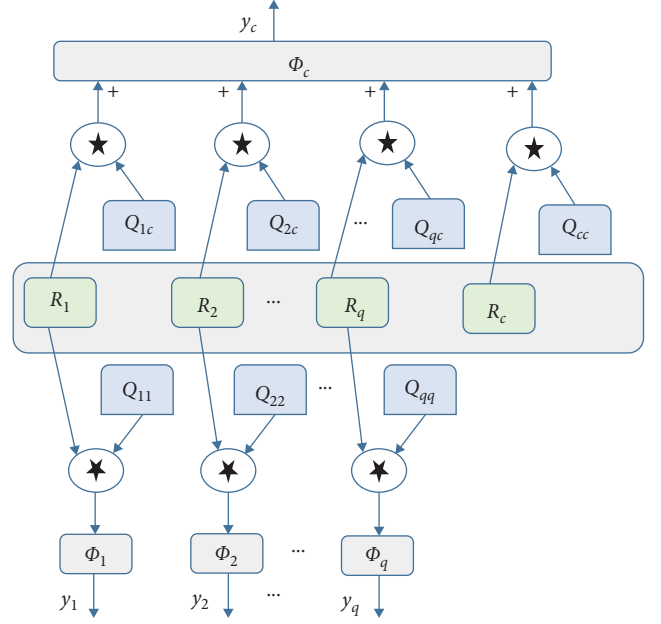


FIGURE 3: The structure of sparse MGCP.

#### 3.1. Sparse MGCP Model Framework

**3.1.1. Sparse MGCP Model.** Given the set of training samples  $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N \in \mathbb{R}^p \times \mathbb{R}^{q+1}$ , the input variables with  $p$  dimensions and the outputs with  $(q+1)$  dimensions are denoted as  $\mathbf{x}_i \in \mathbb{R}^p$  and  $\mathbf{y}_i \in \mathbb{R}^{q+1}$ , respectively. The dimension of the current working output  $y_c(\mathbf{x})$  is one, and the remaining outputs  $\mathbf{y}_j(\mathbf{x}) = \{y_1(\mathbf{x}), y_2(\mathbf{x}), \dots, y_q(\mathbf{x})\}^T$  are the auxiliary outputs. Let  $L = \{1, 2, \dots, q, c\}$  and  $L^{\mathcal{A}} = \{1, 2, \dots, q\}$  denote the index sets for all outputs and all auxiliary outputs, respectively.

Because the ingredients of raw iron materials are related, we obtain the modeling of the current chemical with the assistant of the other chemicals. First, the equation of the output is determined as follows:

$$\mathbf{y}(\mathbf{x}) = \begin{pmatrix} y_1(\mathbf{x}) \\ \vdots \\ y_q(\mathbf{x}) \\ y_c(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \mu_c \end{pmatrix} + \begin{pmatrix} \Phi_1(\mathbf{x}) \\ \vdots \\ \Phi_q(\mathbf{x}) \\ \Phi_c(\mathbf{x}) \end{pmatrix} + \begin{pmatrix} \epsilon_1(\mathbf{x}) \\ \vdots \\ \epsilon_q(\mathbf{x}) \\ \epsilon_c(\mathbf{x}) \end{pmatrix} = \mu + \Phi(\mathbf{x}) + \epsilon(\mathbf{x}), \quad (5)$$

where  $\mu$  is the mean value of the historical data,  $\Phi(\mathbf{x})$  is a MGP with a mean zero and the covariance  $\text{cov}_{ij}^{\Phi}(\mathbf{x}, \mathbf{x}')$ , and  $\epsilon(\mathbf{x})$  is the measurement noise which is independent with  $\Phi(\mathbf{x})$ . The modeling of the current output fully makes use of the transfer knowledge of the auxiliary outputs. Second, the covariance structure of the outputs is constructed.

The structure of sparse MGCP is illustrated in Figure 3, the covariance structure has the ability to convey the knowledge from the original data into the different outputs through the same latent function ( $R_j(\mathbf{x})$ ) and different

smoothing kernels ( $Q_{jj}(\mathbf{x})$  and  $Q_{jc}(\mathbf{x})$ ). More specifically, for the  $j$ th output, the correlation between all these outputs in MGCP is described by a Gaussian white noise process  $R_j(\mathbf{x})$ , and the special characteristic of the output is denoted by the smoothing kernels ( $Q_{jj}(\mathbf{x})$  and  $Q_{jc}(\mathbf{x})$ ). According to the description of MGCP before, the equation of MGCP can be  $\Phi_j(\mathbf{x}) = \sum_{l=1}^q Q_{jl}(\mathbf{x}) \star R_l(\mathbf{x}) = \sum_{l=1}^q \int_{-\infty}^{\infty} Q_{jl}(\mathbf{x} - \mathbf{u}) R_l(\mathbf{u}) d\mathbf{u}$ . On the other hand, the above MGCP have to compute every covariance with high computational complexity. To reduce the computation complexity of our model, a sparse MGCP structure is proposed with the assumption that the auxiliary outputs of our method are independent with each other. Then, the outputs are expressed as follows:

$$\begin{aligned} y_j(\mathbf{x}) &= \mu_j + \Phi_j(\mathbf{x}) + \epsilon_j(\mathbf{x}) \\ &= \mu_j + Q_{jj}(\mathbf{x}) \star R_j(\mathbf{x}) + \epsilon_j(\mathbf{x}), j \in L^{\mathcal{R}} \\ y_c(\mathbf{x}) &= \mu_c + \Phi_c(\mathbf{x}) + \epsilon_c(\mathbf{x}) \\ &= \mu_c + \sum_{k \in L} Q_{kc}(\mathbf{x}) \star R_k(\mathbf{x}) + \epsilon_c(\mathbf{x}). \end{aligned} \quad (6)$$

From Equation (6), we choose to get the accurate approximation of the auxiliary outputs. By taking this approach, we can both ensure the precision of the current output and simplify the computation. According to the covariance structure illustrated in Figure 3, we can obtain the sparse covariance matrix of our proposed MGCP:

$$\Theta_y = \begin{bmatrix} \Theta_{11}^y & \mathbf{0} & \dots & \mathbf{0} & \Theta_{1c}^y \\ \mathbf{0} & \Theta_{22}^y & \dots & \mathbf{0} & \Theta_{2c}^y \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Theta_{qq}^y & \Theta_{qc}^y \\ \Theta_{c1}^y & \Theta_{c2}^y & \dots & \Theta_{cq}^y & \Theta_{cc}^y \end{bmatrix} = \begin{bmatrix} P_1 & P_2 \\ P_2^T & P_3 \end{bmatrix}, \quad (7)$$

where  $\Theta_{kj}^y = \Theta_{kj}^y(\mathbf{x}, \mathbf{x}') = \text{cov}(y_k(\mathbf{x}), y_j(\mathbf{x}'))$ ,  $k, j \in L$ . The covariance matrix is divided into four matrix blocks, where  $P_1$  is the covariance matrix between auxiliary outputs,  $P_2$  represents the covariance matrix between the auxiliary output and the current output, and  $P_3$  represents the covariance matrix between the current outputs. The calculation of covariance  $\Theta_{kj}^y(\mathbf{x}, \mathbf{x}')$  and optimization process of the parameters are represented in the following sections.

Given the training data  $\{\mathbf{X}, \mathbf{y}\}$ , we can calculate the empirical best linear unbiased predictor (EBLUP) with our estimated parameters at the new inputs  $\mathbf{x}_*$ , and the predictive mean  $\hat{y}_c(\mathbf{x}_*)$  and the variance  $V_c(\mathbf{x}_*)$  are as follows:

$$\begin{aligned} \hat{y}_c(\mathbf{x}_*) &= \Theta_{*c}^T \Theta_y^{-1} \mathbf{y} \\ V_c(\mathbf{x}_*) &= \Theta_{cc}^{\Phi} - \Theta_{*c}^T \Theta_y^{-1} \Theta_{*c}, \end{aligned} \quad (8)$$

where  $\Theta_{*c}$  is the covariance between the new inputs and the historical data,  $\Theta_{cc}^{\Phi}$  is the covariance of the new inputs. Therefore, the mean prediction of the new inputs can be

regarded as the linear combination of the historical observations.

**3.1.2. Covariance of Sparse MGCP.** Based on the model structure in Equation (5) and the assumption that the independence between  $\Phi(\mathbf{x})$  and measurement noise  $\epsilon(\mathbf{x})$ , the covariance between any two outputs can be expressed as follows:

$$\begin{aligned} \Theta_{kj}^y(\mathbf{x}, \mathbf{x}') &= \text{cov}(y_k(\mathbf{x}), y_j(\mathbf{x}')) \\ &= \text{cov}(\Phi_k(\mathbf{x}), \Phi_j(\mathbf{x}')) + \text{cov}(\epsilon_k(\mathbf{x}), \epsilon_j(\mathbf{x}')) \\ &= \Theta_{kj}^{\Phi}(\mathbf{x}, \mathbf{x}') + \Theta_{kj}^{\epsilon}(\mathbf{x}, \mathbf{x}') \tau_{kj}, k, j \in L, \end{aligned} \quad (9)$$

where  $\Theta_{kj}^{\Phi}(\mathbf{x}, \mathbf{x}')$  is the covariance between  $\Phi_k(\mathbf{x})$  and  $\Phi_j(\mathbf{x}')$ , and  $\Theta_{kj}^{\epsilon}(\mathbf{x}, \mathbf{x}')$  is the covariance of the measurement noise. The Kronecker delta function  $\tau_{kj}$  has the characteristic that the value is equal to 1 while  $k=j$ , and the value is zero otherwise.

For the computation of covariance, we select the Gaussian white noise as the latent function ( $R_j$ ) and the Gaussian kernels as the smoothing kernels ( $Q_{kj}$ ). The Gaussian kernel has the metric that it can complete the modeling of various spatial features with few hyperparameters. The equation of the Gaussian kernel is as follows:

$$Q_{kj}(x) = \alpha_{kj} \pi^{-\frac{1}{4}} |\mu_{kj}|^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2\mu_{kj}^2}\right), k, j \in L, \quad (10)$$

where  $\alpha_{kj}$  is the scaling parameter factor used to vary the cross correlation between two outputs, and  $\mu_{kj}$  is the length scale for the inputs. Combined with the characteristics of the white noise Gaussian and Gaussian kernel, we obtain the following equation:

$$\Theta_{kj}^y(\mathbf{x}, \mathbf{x}') = \text{cov}_{kj}^y(\mathbf{x}, \mathbf{x}') = \mathbf{0}, k, j \in L^{\mathcal{R}}, k \neq j \& k, j \in L^R, \quad (11)$$

$$\Theta_{jc}^y(\mathbf{x}, \mathbf{x}') = \alpha_{jj} \alpha_{jc} \frac{2^{\frac{1}{2}} |\mu_{jj} \mu_{jc}|^{\frac{1}{2}}}{|\mu_{jj}^2 + \mu_{jc}^2|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{x} - \mathbf{x}')^2}{2(\mu_{jj}^2 + \mu_{jc}^2)}\right] + \Sigma \tau_{jc}, \quad (12)$$

$$\Theta_{jj}^y(\mathbf{x}, \mathbf{x}') = \alpha_{jj}^2 \exp\left[-\frac{(\mathbf{x} - \mathbf{x}')^2}{4\mu_{jj}^2}\right] + \Sigma \tau_{jj}, \quad (13)$$

$$\Theta_{cc}^y(\mathbf{x}, \mathbf{x}') = \sum_{l \in L} \alpha_{lc}^2 \exp\left[-\frac{(\mathbf{x} - \mathbf{x}')^2}{4\mu_{lc}^2}\right] + \Sigma \tau_{lc}, \quad (14)$$

where  $\Sigma$  is a diagonal matrix and represents the measurement noise. Equation (11) is based on the assumption that the auxiliary outputs are independent of each other. The derivation of Equation (12) is illustrated in the Appendix.

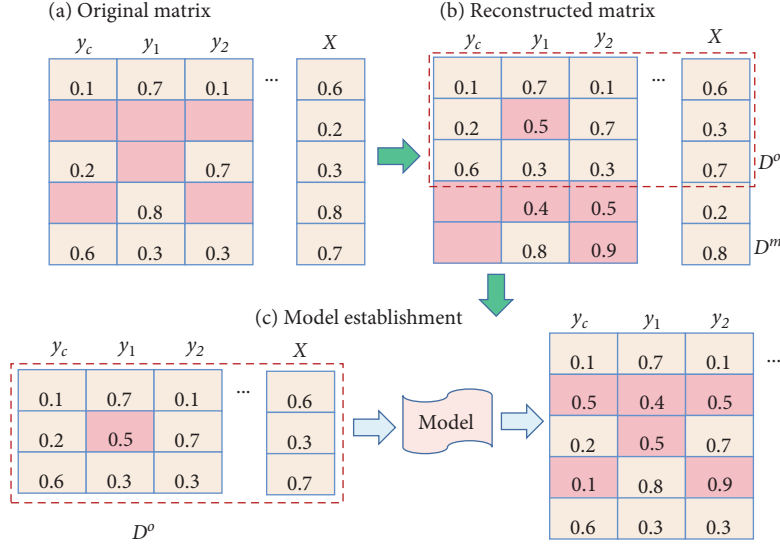


FIGURE 4: The process of reconstructing the dataset. (a) The original matrix is shown, in which each column stands for a feature, and the missing values are marked in red. (b) The missing values of the auxiliary outputs are estimated and the matrix is reconstructed. (c) The model is established by the observed dataset.

The parameter  $\alpha_{j_c}, j \in L^{\mathcal{R}}$  describes the correlation between  $\Phi_j(\mathbf{x})$  and  $\Phi_c(\mathbf{x})$ .

### 3.2. Regularized Sparse MGCP for Data Imputation

**3.2.1. Missing Values Prefilling.** The previous situations are based on the assumption that all the outputs are complete. However, the incomplete outputs with a large amount of missing data are a common issue in industrial applications. The rough discard of these incomplete data will greatly decrease the accuracy of the modeling. Therefore, we proposed an effective approach to prefill the missing values of outputs.

The basic idea of our method is to first impute missing values of the auxiliary outputs, then establish our MGCP model with the part of the observed data. The process is illustrated in Figure 4. First, the missing values of the auxiliary outputs are filled with the estimated values. In our work, GP is adopted as the method because it is easy to implement and no prior information is needed. The reconstructed matrix can be expressed as follows:

$$\mathcal{D}^{\text{new}} = (\mathbf{x}_i, \mathbf{y}'_i)_{i=1}^N, \quad (15)$$

where  $\mathbf{y}' = \{y'_1, y'_2, \dots, y'_q\}$ . Suppose that the number of the observations  $y_c^o$  is  $N^o$  and the number of the missing values  $y_c^m$  is  $N^m = N - N^o$ . Then, the reconstructed matrix is separated as the observed dataset  $\mathcal{D}^o = (\mathbf{x}_i, \mathbf{y}_i)_{i=1}^{N^o}$  and the missing part  $\mathcal{D}^m = (\mathbf{x}_i, \mathbf{y}_i)_{i=1}^{N^m}$  according to the missing situation of the current output. The model is established by our proposed MGCP with the observed dataset  $\mathcal{D}^o$ . After that, we can compute the missing values of the current output  $\hat{y}_c$  by the established model.

**3.2.2. Two Regularized MGCP for Data Imputation.** Since we prefill a large number of the missing values of the auxiliary

outputs to finish the modeling process, it is inevitable to introduce some false data. Furthermore, it may create incorrect connection between the auxiliary output and the current output. To avoid this negative influence, we proposed regularized MGCP based on the previous results.

From the structure of the covariance in Equation (7), the cross-covariance between the auxiliary output  $y_j$  and the current output  $y_c$  is  $\Theta_{j_c}^y$ . If the estimated parameter  $\alpha_{j_c}$  is equal to zero, cross-covariance  $\Theta_{j_c}^y$  is also equal to zero, which means there is no correlation between these two outputs. By introducing the  $l_0$ -norm over parameter  $\alpha_{j_c}$ , we can obtain the following optimization problem:

$$\min_{\theta, \sigma_e} -\log P(\mathbf{y}|\mathbf{x}; \theta, \sigma_e) + \lambda_1 \sum_{j \in L^{\mathcal{R}}} \|\alpha_{j_c}\|_0, \quad (16)$$

where  $-\log P(\mathbf{y}|\mathbf{x}; \theta, \sigma_e)$  denotes the negative log-likelihood function of MGCP, and  $\lambda_1$  is a nonnegative penalty hyperparameter. The optimization of  $\lambda_1$  can be conducted by cross-validation method.

It is a NP-hard problem because the  $l_0$ -norm function is not continuous and convex. By replacing the  $l_0$ -norm function with  $l_1$ -norm function, we achieve the relaxation of this item. This method is named MGCP-R and the expression is as follows:

$$\min_{\theta, \sigma_e} -\log P(\mathbf{y}|\mathbf{x}; \theta, \sigma_e) + \lambda_1 \sum_{j \in L^{\mathcal{R}}} \|\alpha_{j_c}\|_1. \quad (17)$$

Moreover, we expect the estimated missing values to have a high similarity with the actual time series trajectory if the special missing part is large. Therefore, we proposed an modeling method based on DTW, which is named as MGCP-R-D. Then, we have the following function:

$$\min_{\theta, \sigma_\epsilon} -\log P(\mathbf{y}|\mathbf{x}; \theta, \sigma_\epsilon) + \lambda_1 \sum_{j \in L^{\mathcal{D}}} \|\alpha_{jc}\|_1 + \lambda_2 \text{DTW}(\hat{\mathbf{y}}_c, \mathbf{y}_c^o), \quad (18)$$

where  $\text{DTW}(\hat{\mathbf{y}}_c, \mathbf{y}_c^o)$  is the distortion loss of our model based on DTW in order to obtain the global similarity between the prediction  $\hat{\mathbf{y}}_c$  and the observations  $\mathbf{y}_c^o$ .

**3.3. Parameter Identification.** For inputs variables  $\mathbf{x}$ , the outputs  $\mathbf{y}$  can be divided into the auxiliary outputs  $\mathbf{y}_s = \{y_1^T, y_2^T, \dots, y_q^T\}^T$  and the current output  $y_c$ . The mean values of the auxiliary outputs and the current output are  $\mu_s$  and  $\mu_c$ , respectively. From Equation (12), the parameters includes the kernel parameters  $\theta = \{\alpha_{jj}, \alpha_{jc}, \mu_{jj}, \mu_{jc} | j \in L\}$  and the measurement noise  $\sigma_\epsilon$ . For the convenience of expression, we omit the notation  $\theta$  and  $\sigma_\epsilon$  later. Based on our covariance structure, we can modify the objective function to reduce the computation complexity and make it suitable to the optimization methods, such as L-BFGS method.

**3.3.1. Maximum Likelihood Item.** In our model, there is no correlation between any two auxiliary outputs. We can decompose the first item of the objective function as follows:

$$\begin{aligned} P(\mathbf{y}; \mathbf{x}) &= P(y_1, y_2, \dots, y_q, y_c; \mathbf{x}) \\ &= P(y_c | y_1, y_2, \dots, y_q; \mathbf{x}) P(y_1, y_2, \dots, y_q; \mathbf{x}) \\ &= P(y_c | y_1, y_2, \dots, y_q; \mathbf{x}) \times \prod_{j \in L^{\mathcal{D}}} P(y_j; \mathbf{x}). \end{aligned} \quad (19)$$

Combined with the the properties of the Schur complement and the operation of multivariate normal distribution, the negative log-likelihood function is decomposed:

$$\begin{aligned} -\log P(\mathbf{y}|\mathbf{x}) &= \frac{1}{2} [(\mathbf{y}_c - \mathbf{A})^T \mathbf{B}^{-1} (\mathbf{y}_c - \mathbf{A}) \\ &\quad + \sum_{j \in L^{\mathcal{D}}} \mathbf{y}_j^T \left[ \boldsymbol{\Theta}_{jj}^y(\mathbf{x}, \mathbf{x}') \right]^{-1} \mathbf{y}_j] + \frac{1}{2} [\log |\mathbf{B}| \\ &\quad + \sum_{j \in L^{\mathcal{D}}} \log \left| \boldsymbol{\Theta}_{jj}^y(\mathbf{x}, \mathbf{x}') \right|] + \frac{N}{2} \log(2\pi), \end{aligned} \quad (20)$$

where  $\mathbf{A} = \mu_c + P_2^T P_1^{-1} (\mathbf{y}_s - \mu_s)$ ,  $\mathbf{B} = P_3 - P_2^T P_1^{-1} P_2$ ,  $\boldsymbol{\Theta}_{jj}^y(\mathbf{x}, \mathbf{x}') = \text{cov}(y_j(\mathbf{x}), y_j(\mathbf{x}'))$ , and the notations  $P_1, P_2$ , and  $P_3$  denote the partitions of  $\boldsymbol{\Theta}_y$  expressed in Equation (7).

When we have  $N$  samples for these  $q+1$  outputs, the complexity of this part is reduced from  $O([(q+1)N]^3)$  to  $O[(q+1)N^3]$  with our special covariance structure.

**3.3.2. Sparsity Item.** It is difficult to tackle the  $l_0$ -norm function due to its noncontinuity and nonconvexity. The most common strategy is to relax  $l_0$ -norm with  $l_1$ -norm. However, the  $l_1$ -norm can still not meet the requirement of smoothness at point zero. Then, a Huber smooth approximation is adopted for this problem [36]:

**Input:**  $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N, \lambda_1, \lambda_2, a, \gamma$

- 1: Prefill the missing values of the auxiliary outputs  $\mathcal{D}^{\text{new}} = (\mathbf{x}_i, \mathbf{y}'_i)_{i=1}^N$  and reconstruct the dataset into observed part  $\mathcal{D}^o = (\mathbf{x}_i, \mathbf{y}_i)_{i=1}^{N^o}$  and missing part  $\mathcal{D}^m = (\mathbf{x}_i, \mathbf{y}'_i)_{i=1}^{N^m}$  according to the missing states of the current output.
- 2: Initiate the parameter  $\theta = \{\alpha_{jj}, \alpha_{jc}, \mu_{jj}, \mu_{jc} | j \in L\}, \sigma_\epsilon$
- 3: For  $\mathcal{D}^o$ , calculate  $\hat{\theta}$  and  $\hat{\sigma}_\epsilon$  by solving the optimization problem Equations (18), (20)–(22), where the covariance matrixes are obtained by Equations (11)–(13)
- 4: Calculate the missing values of the current output using Equation (8) with  $\hat{\theta}$  and  $\hat{\sigma}_\epsilon$ .
- 5: Return  $\hat{\theta}, \hat{\sigma}_\epsilon$

ALGORITHM 1: Pseudo code of MGCP-R-D.

$$\sum_{j \in L^{\mathcal{D}}} \|\alpha_{jc}\|_1, \|\alpha_{jc}\|_1 = \begin{cases} \frac{1}{2a} \alpha_{jc}^2, & |\alpha_{jc}| \leq a \\ |\alpha_{jc}| - \frac{a}{2}, & |\alpha_{jc}| \geq a \end{cases}, \quad (21)$$

where  $a$  is a small constant, e.g.,  $10^{-4}$ .

**3.3.3. Shape Distortion Item.** DTW can be formulated as the optimization problem:

$$\text{DTW}(\hat{\mathbf{y}}_c, \mathbf{y}_c^o) = \min_{W \in \mathcal{W}_{m,n}} \langle W, \eta(\hat{\mathbf{y}}_c, \mathbf{y}_c^o) \rangle, \quad (22)$$

where  $W$  is the warping path to align all the points between the predictions  $\hat{\mathbf{y}}_c$  and the observers  $\mathbf{y}_c^o$ ,  $\mathcal{W}_{m,n}$  is the set of all possible warping paths,  $\eta(\hat{\mathbf{y}}_c, \mathbf{y}_c^o)$  is a similarity,  $m$  and  $n$  are the length of  $\hat{\mathbf{y}}_c$  and  $\mathbf{y}_c^o$ , respectively. The smooth min operator is applied to make the DTW differentiable. The equation is expressed as follows [37]:

$$\text{DTW}_\gamma(\hat{\mathbf{y}}_c, \mathbf{y}_c^o) = -\gamma \log \left[ \sum_{W \in \mathcal{W}_{m,n}} \exp \left( -\frac{\langle W, \eta(\hat{\mathbf{y}}_c, \mathbf{y}_c^o) \rangle}{\gamma} \right) \right], \quad (23)$$

where  $\gamma$  approaches zero, Equations (22) and (23) are equivalent.

The complexity of this part will not exceed  $O(N^3)$  for each output, which is smaller than the complexity of MGCP construction process in our method.

Finally, we merge Equations (20)–(22) into the optimization function Equation (11) to achieve the parameter identification. Since this objective function is not convex, we initiate the parameter values for several times to avoid the multiple local optima, which will slightly decrease the prediction precision of our model [31]. The selection of hyperparameters  $\lambda_1$  and  $\lambda_2$  is conducted by a grid search with cross-validation strategy. The implement of MGCP-R-D is shown in Algorithm 1.

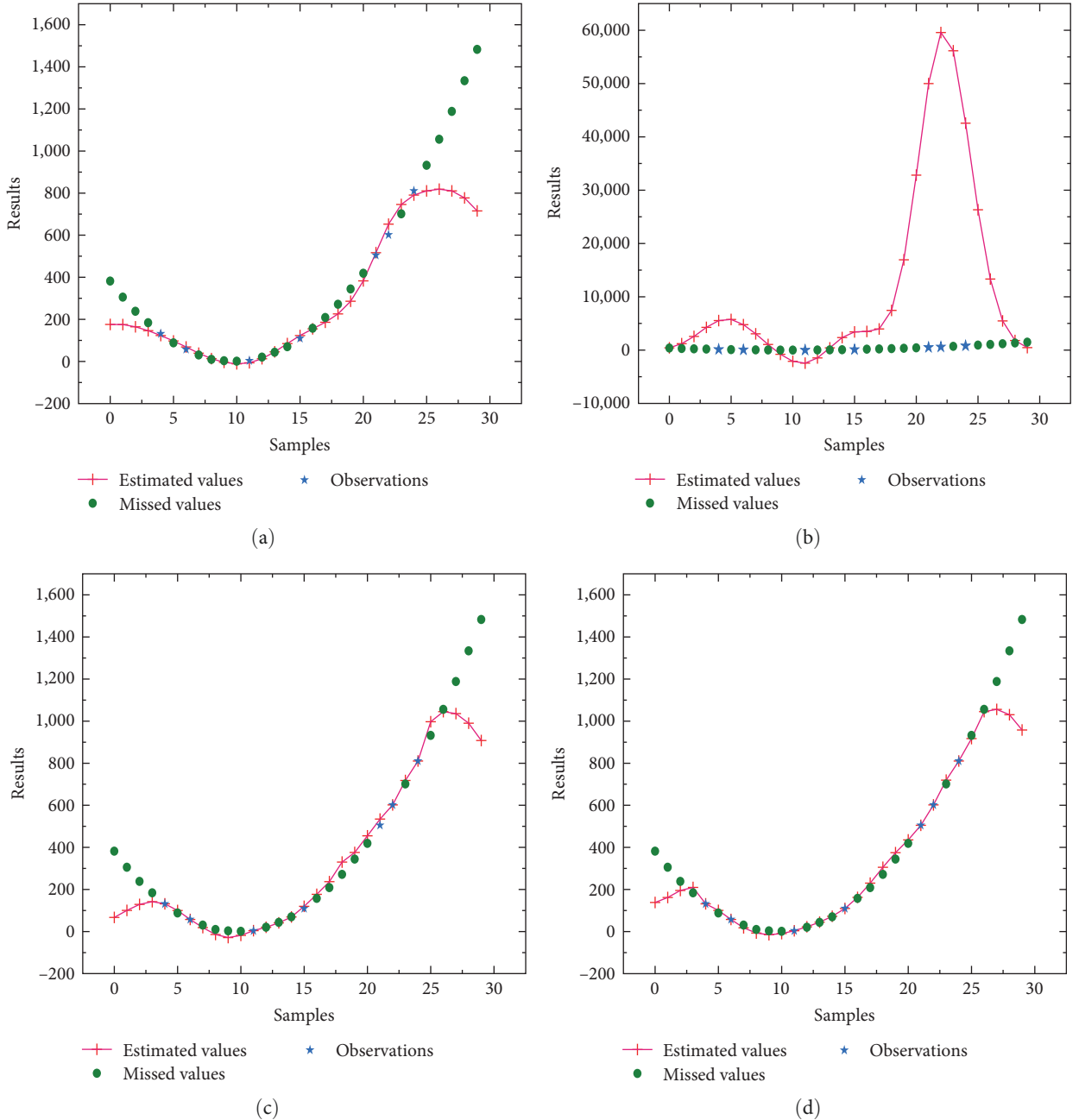


FIGURE 5: Performance comparison with 70% missing ratio in numerical study. (a) GCP, (b) MGCP, (c) MGCP-R, and (d) MGCP-R-D.

## 4. Case Application

**4.1. Numerical Case Study Using Simulated Data.** A numerical example is constructed to compare the performance and study the characteristics of our method. The hyperparameters, including  $\theta = \{\alpha_{ij}, \alpha_{jc}, \mu_{ij}, \mu_{jc} | j \in L\}$  and  $\sigma_\epsilon$ , are optimized with L-BFGS method in GPflow [38]. For the constants for the regularization items, we have  $a = 10^{-4}$  and  $\gamma = 10^{-5}$ . Three methods are introduced as benchmarks:

- (i) The single GP model with a convolution kernel, denoted as GCP. It only uses the current output for training and testing.

- (ii) A nonregularized multioutput GCP model with the same covariance structure as ours, denoted as MGCP.

An numerical example is shown in Equation (24), which has the inputs  $x_1 \in [-30, 30]$  and  $x_2 \in [-20, 40]$ , the current output  $f_c$  and three auxiliary outputs  $f_1, f_2, f_3$ . The expressions are expressed as follows:

$$\begin{aligned}
 f_1(x_1) &= 3 \sin(x_1) + r_1 \\
 f_2(x_1, x_2) &= 3 \cos(2x_1) + x_2^2 + x_2 + r_2 \\
 f_3(x_1, x_2) &= 2 \sin(2x_1) + x_2^2 + r_3 \\
 f_c(x_1, x_2) &= 2 \sin(x_1) + x_2^2 + x_2 + r_4,
 \end{aligned} \tag{24}$$



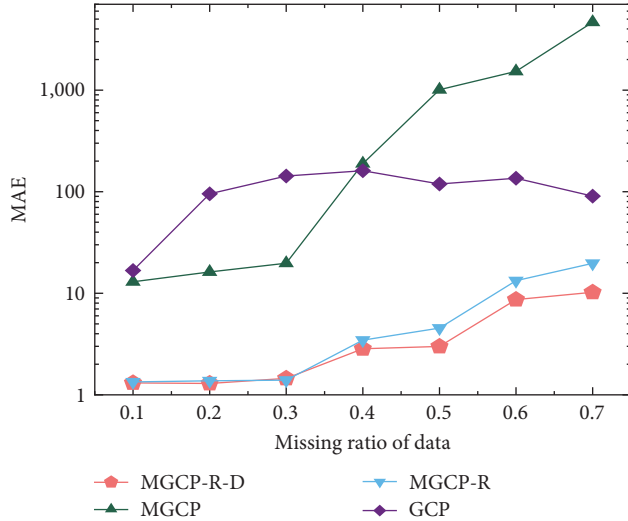


FIGURE 6: Comparison of MAE with different missing ratios.

TABLE 1: Pearson correlation between the auxiliary outputs and the current output with 70% missing ratio in numerical study.

	f1 : fc	f2 : fc	f3 : fc
Correlation	0.0647	0.981	0.965

TABLE 2: Part of parameters estimation with 70% missing ratio in numerical study.

	MGCP-R-D	MGCP-R	MGCP
$\alpha_{1c}$	0	0	12.2
$\alpha_{2c}$	15.1	1.26	69.3
$\alpha_{3c}$	3.49	5.21	31.3

where  $r_1$ ,  $r_2$ ,  $r_3$ , and  $r_4$  are the measurement noise with standard deviation  $\sigma = 0.04$ . We repeat each method for 50 times and the performance of these methods are evaluated by mean absolute error (MAE). Moreover, in order to avoid the influence of unbalanced data missing, the middle stable 20 points are used for evaluation.

The results are illustrated in Figures 5 and 6. From Figure 6, when the missing ratio is lower than 40%, all the models with MGCP structure have the superior performance than GCP due to the benefits of borrowing strength from other outputs. However, the performance of MGCP deteriorates dramatically with the increase of missing ratio especially when the missing ratio is greater than 50%. We believe that this is because the false connections between these outputs decrease the model accuracy for prediction. This is also verified by the subsequent correlation analysis. The PCs and the scaling parameters between the auxiliary outputs and the current output are shown in Tables 1 and 2, respectively. In Table 2, both MGCP-R-D and MGCP-R methods reduce the scaling parameter  $\alpha_{1c}$  to zero, which is consistent with the actual situation because the correlation between these two variables is low. Therefore, these results demonstrate the effectiveness of our models.

**4.2. Case Study with Air Quality Data.** The public dataset from KDD CUP Challenge 2018 contains the historical air quality data of Beijing from January 1, 2017 to December 30, 2017 [39]. The pressure, humidity, wind direction, and wind speed recorded by Zhaitang station are used as the input variables. The air quality indexes are used as the outputs, including PM2.5, NO<sub>2</sub>, O<sub>3</sub>, and SO<sub>2</sub>. In this section, PM2.5 is the current output needed to impute while NO<sub>2</sub>, O<sub>3</sub>, and SO<sub>2</sub> are the auxiliary outputs. Moreover, 60% of these outputs are randomly discarded. In the preprocessing process, the outliers of the dataset are detected by box plot and pre-filled by mean values, these input variables are standardized into the range between 0 and 1, and the missing output values are pre-filled by GP.

In the case of this public dataset, a substantial amount of data has been discarded, making it impractical to retrieve all the missing values. Therefore, it is imperative to choose an effective method for evaluating the performance of the proposed algorithm. We obtain the accuracy of the established model by the error between the prediction values and the observed values. There are 80% of the total dataset that are the training dataset and the rest data are regarded as the testing dataset. In the training process, the parameters are optimized by Algorithm 1. Then, in the testing process, the values of the current output are estimated with the new input variables. The mean square error (MSE) and MAE between estimated value and observations are calculated to evaluate the performance of the method.

Different data imputation methods are introduced for performance evaluation. Miss forest (MISS-F) method uses the other features to predict and impute the current value and improves its precision by combining many weak trees [40]. The expectation maximization (EM) estimates the incomplete value by calculating the maximum likelihood [41]. This algorithm uses only the current output to obtain its imputation values. The K-nearest neighbors (KNN) uses observations to find the KNN of missing values, and the average of nonmissing values is adopted as the missing values [8]. The matrix factorization (MF) reconstructed the original values by matrix decomposition [42]. The multiple imputation by chained equations (MICE) obtains the missing values by combining different simple imputation methods through the regression model [43]. For EM and KNN, we implement them with impute package with the recommended parameters. For MICE and MF, the fancyimpute package is applied to implement these two algorithms [44]. All the experiments are run 30 times and the mean values are taken as the final result to reduce the influence of contingency. The hyperparameters involved in all algorithms are tuned followed by 10 fold cross-validation strategy and the best hyperparameters are applied. The parameter of MGCP-R-D is determined to be  $\lambda_1 = 10$  and  $\lambda_2 = 100$ .

The results of seven approaches are illustrated in Figure 7 and Table 3. Except for these above approaches, the performance of MGCP and MGCP-R is also tested, but the results are far worse than the listed methods. It is observed that the proposed MGCP-R-D behaves well. Obviously, the results of the selected methods are related to the data structure of the

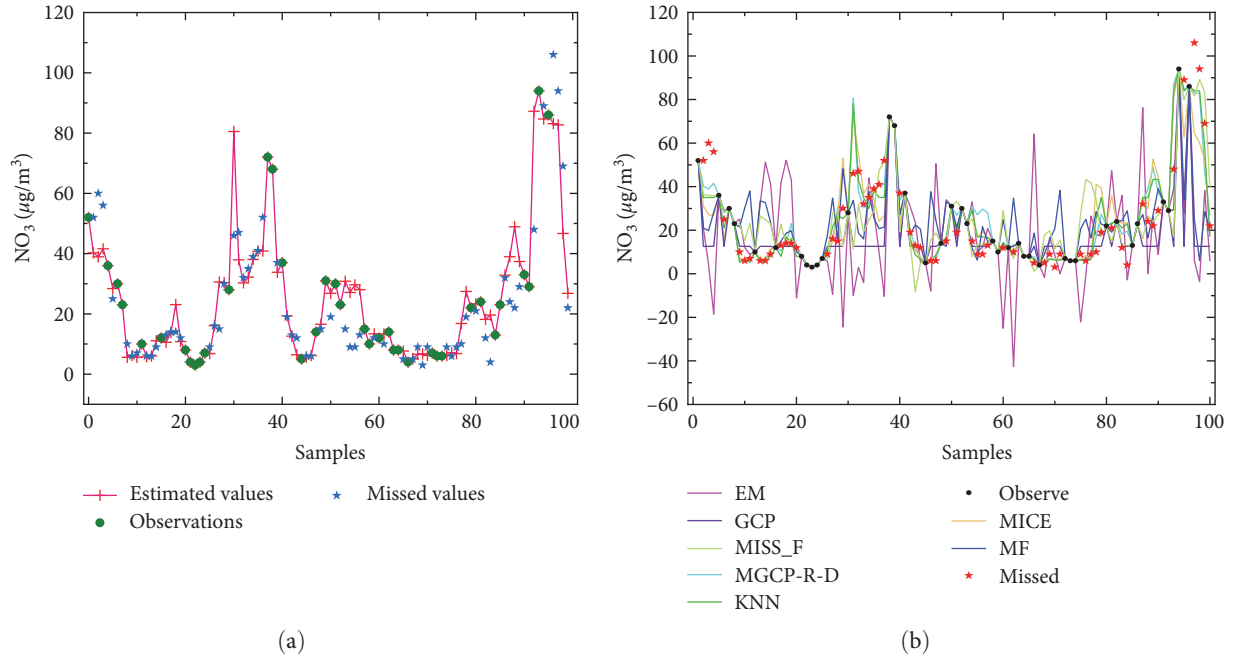


FIGURE 7: Results of different approaches. Performance of (a) MGCP-R-D and (b) different approaches.

TABLE 3: The performance comparison of different approaches.

Methods	MAE + STD	MSE + STD
KNN	4.61 + 0.65	74 + 5.7
MGCP-R-D	<b>4.77 + 0.42</b>	<b>81 + 5.4</b>
MICE	6.19 + 0.58	137 + 17.8
MISS-F	7.25 + 0.83	133 + 12.2
GCP	9.57 + 1.13	405 + 30.3
MF	10.63 + 2.11	359 + 31.5
EM	16.3 + 2.52	745 + 80.3

Bold values signify that the MGCP-R-D is our proposed method.

dataset. By enhance the strength with the related outputs, MGCP-R-D could obtain more intrinsically useful information for data imputation.

**4.3. Case Study with Sintering Raw Materials Data.** As shown in Figure 8, the sintering process with a 360 m<sup>2</sup> sintering machine is studied. The workshop of sintering plant includes mixing, sintering, crushing, cooling machine, and other equipments. The data of the variables are collected by distributed control system and are sent to the database. The complement of the chemical contents is vital for the prediction and control of the sintering process. However, the missing of the chemical contents are heavy due to the delayed measurement and refreshment. For example, the delayed refreshment of the raw material caused the excess emissions of sulfur content, and even leads to the failure of sintering sulfur treatment system, which is a real production accident occurred in running steel plant.

In Figure 8, the missing situations of these input material variables  $Y_c$ ,  $Y_1$ ,  $Y_2$ ,  $Y_3$ , and  $Y_4$  are presented, and the introduction of these variables is shown in Table 4. The blue lines denote the observations while the white space denotes these

missing values. It is observed that the pattern of the missing values is missing at random because these chemical contents are related to the dosing scheme of the sintering process and the quality of the sinter product. The characteristics of the chemical contents are presented as follows:

- (1) Irregular missing intervals: Since the chemical contents of different raw materials are measured with different frequencies, the missing situation for different varies greatly. Moreover, the sampling intervals for the same variable can vary according to the actual measurement situation. The irregular data missingness have increased the difficulty of data imputation.
- (2) Complex relationships between missing values: The missing values of the same materials are determined by the dosing scheme. In order to produce qualified sinter product, there are some constraints among these input raw materials. Moreover, the quality indexes of produced sinter product are related to these chemical contents. However, due to the complex chemical reaction in the sintering process, this modeling for these relationships is difficult to establish.

In our study, the raw data are collected from the actual running sintering plant of a steel company from January 27, 2022 to April 3, 2022. Due to the long sampling period, there are total 1,000 sets of data collected. Among them, 900 sets of samples are treated as training data, while the remaining 100 sets of samples are testing data. According to the analysis of the data, there are quantities of missing values included in the samples, about 50%–70%. The missing ratio is calculated by the proportion of the missing values in the number of the total values of the selected chemical contents [45, 46]. For the testing data, there are only 32 sets of samples remaining.

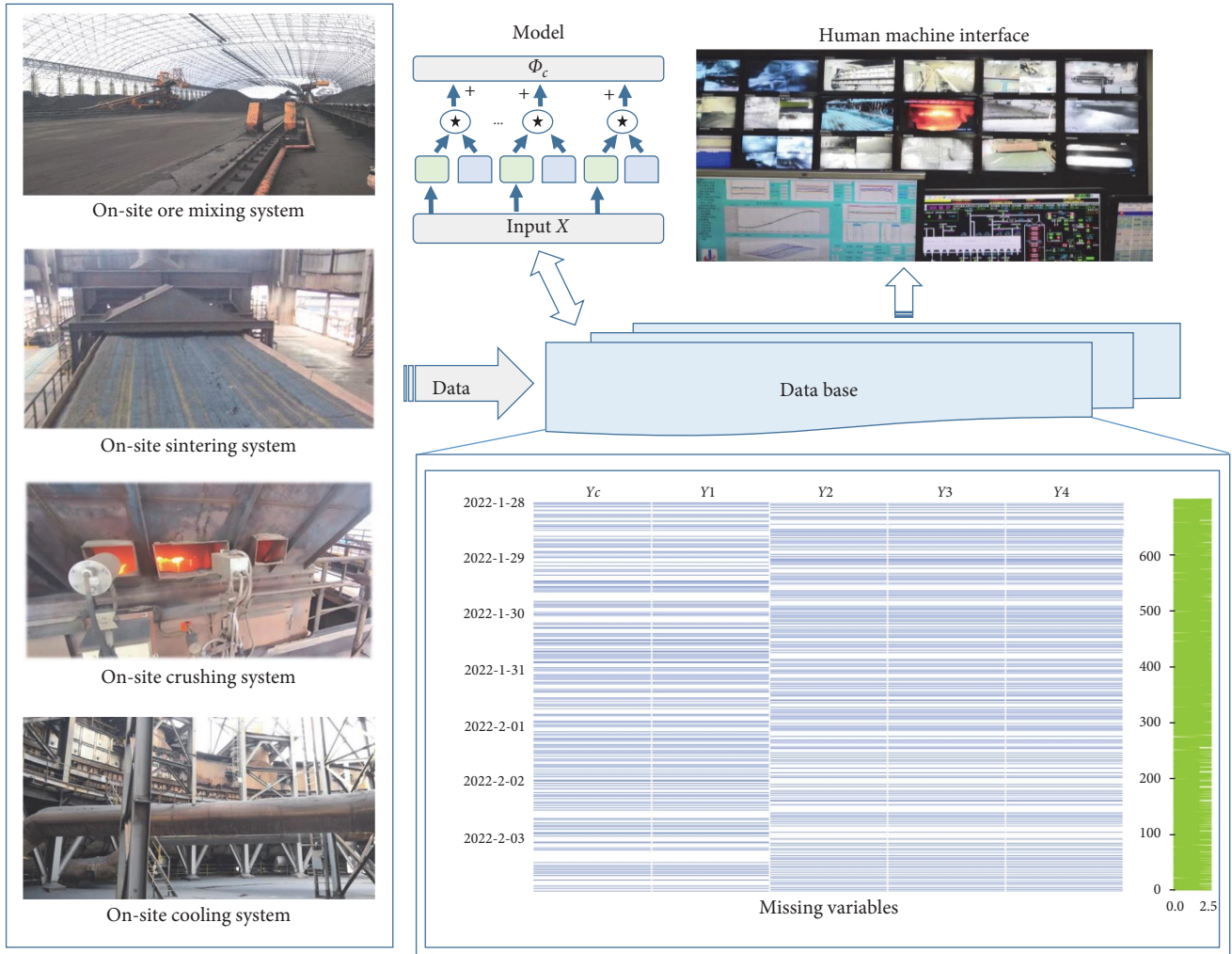


FIGURE 8: Time series of sintering raw materials with missing data measured and stored in the database. Based on our model, the processed data are returned to the database and the human machine interface (HMI) can display the processed data.

TABLE 4: Selected variables for imputation.

Variables	Materials	Symbol	Min	Max	Unit
$X_1$	Sinter	CaO	11.12	11.89	%
$X_2$	Sinter	R	2.01	2.25	—
$X_3$	Sinter	SiO <sub>2</sub>	4.97	5.67	%
$X_4$	Sinter	TFe	55.21	55.82	%
$Y_1$	Mixed iron ore	TFe	62.03	62.55	%
$Y_2$	Quicklime powder	CaO	80.47	85.51	%
$Y_3$	Dolomite powder	CaO	48.33	54.26	%
$Y_4$	Dolomite powder	S	0.131	0.174	%
$Y_c$	Mixed iron ore	SiO <sub>2</sub>	5.24	5.66	%

The variables of the model are determined according to the Pearson coefficient (PC) method and the mechanism of the sintering process mentioned previously, which is described in Table 4.

These existing quality indexes of the sinter quality data are used as the input variables, and these chemical contents of the raw materials are utilized as the outputs. For the input variables, the outliers are detected by box plot, and the

outliers are replaced by the average value of before and after two values of these values. These inputs variables are further standardized to [0, 1]. On the other hand, for these outputs, the Gaussian process is applied to prefill the missing values, as illustrated in Figure 9. Since all these data are incomplete, this prefilling process is used as the preprocess for establishing the model. These outputs are further separated as the current output and the auxiliary outputs. The

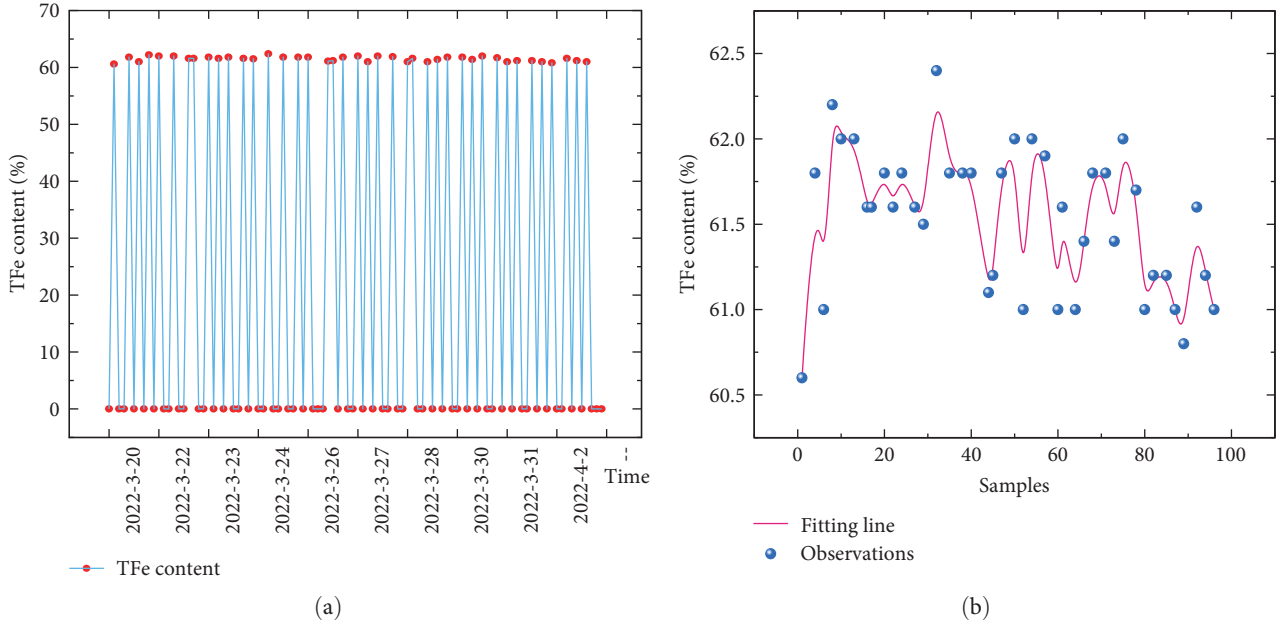


FIGURE 9: Preprocessing process of output data (TFe). (a) Original time series data for TFe and (b) fitting results for TFe by Gaussian process.

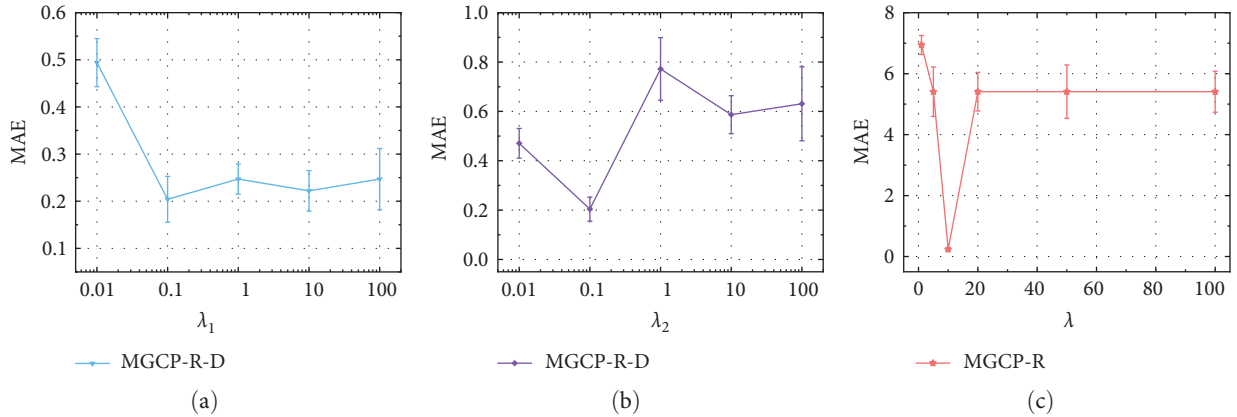


FIGURE 10: MAE of MGCP-R-D and MGCP-R with different parameters. The effects of (a)  $\lambda_1$  with  $\lambda_2 = 0.1$  on MGCP-R-D, (b)  $\lambda_2$  with  $\lambda_1 = 0.1$  on MGCP-R-D, and (c)  $\lambda$  on MGCP-R.

algorithm for obtaining the missing values of the current output is implemented by Algorithm 1.

In our experiments, 11 data imputation methods are introduced for performance evaluation, including these above methods, long short term memories (LSTM) and last observation filling (LOF). LSTM methods are used to predict and update missing values, which are prefilled with the mean values. LOF method applies the last observation as the missing values [17]. These experiments are carried out independently for 30 times. According to the description of the introduction section, these methods are the mainstream methods to compare the performance of the proposed method in different fields. The hyperparameters involved in all algorithms are tuned followed by 10 fold cross-validation strategy and the best hyperparameters are applied. As illustrated in Figure 10, the parameter of MGCP-R-D is

determined to be  $\lambda_1 = 0.1$  and  $\lambda_2 = 0.1$ , and the penalty factor  $\lambda$  of MGCP-R is set as 10. Compare to MGCP-R, MGCP-R-D is less sensitive to the selection of parameters due to its larger number of penalty coefficients, which makes it more robust to parameter selection.

For the performance index, the hit rate  $HR(\mathcal{M})$  is also defined to evaluate the performance of all algorithms:

$$HR(\mathcal{M}) = \frac{1}{n} \sum_{i=1}^n \mathcal{E}_i(\mathcal{M}) \times 100\%, \quad (25)$$

where  $\mathcal{E}_i$  is required as follows:

$$\mathcal{E}_i(\mathcal{M}) = \begin{cases} 1, & \text{if } |y_i - \hat{y}_i| < \mathcal{M}\% \\ 0, & \text{if } |y_i - \hat{y}_i| > \mathcal{M}\% \end{cases}, \quad (26)$$

TABLE 5: The performance comparison of different methods.

Methods	MAE + STD	HR1 (%)	HR2 (%)
MGCP-R-D	<b>0.204 + 0.0295</b>	<b>81.2</b>	<b>96.9</b>
MICE	0.238 + 0.0292	76.1	93.3
MGCP-R	<b>0.247 + 0.0487</b>	<b>66.2</b>	<b>93.1</b>
MF	0.248 + 0.0368	65.6	92.9
MISS-F	0.250 + 0.0325	66.7	93.2
KNN	0.259 + 0.0171	59.4	89.1
EM	0.274 + 0.0167	62.5	87.5
LOF	0.328 + 0.1292	61.7	81.3
GCP	0.693 + 0.302	43.7	59.4
LSTM	5.5 + 1.0292	57.5	46.8
MGCP	170,941 + 17,095	0	0

Bold values signify that the MGCP-R-D and MGCP-R are the proposed approaches in this paper.

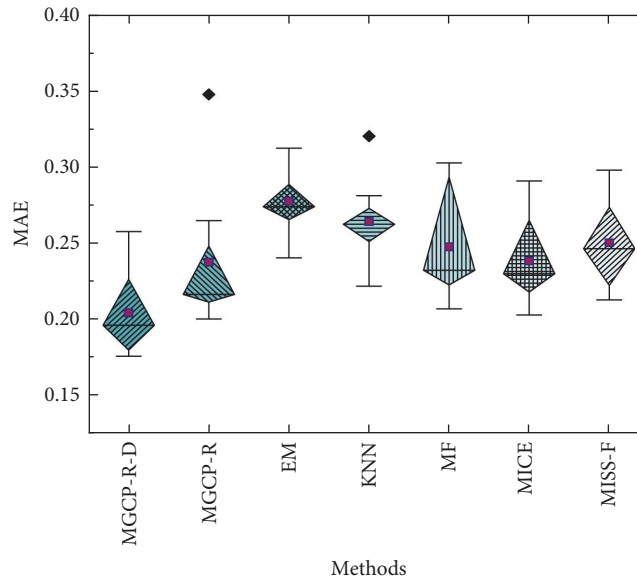


FIGURE 11: Comparison results of different methods.

where  $y_i$  is the true value and  $\hat{y}_i$  is the predicted value, respectively, and the total number of the samples is  $n$ . Two values of HR1(0.25) and HR1(0.5) are used here, which means the percentage of their estimation errors are less than 0.25% and 0.5%, respectively.

Compared with these related algorithms, the proposed MGCP-R-D method achieves the best performance and is the only method with a hit rate (HR1) exceeding 80%. The statistical results of different methods are illustrated in Table 5. STD stands for the standard deviation. From the table, our proposed MGCP-R-D has the best performance compared to other methods. It has the lowest MAE value and the largest HR(0.25) and HR(0.5). The values of them are 0.204, 81.2%, and 96.9%, respectively. MGCP has the worst values and failed to realize the imputation. The worst performance of MGCP may be caused by the failure of parameter estimation according to the analysis in numerical study. Compared to other methods, both GCP and EM use only the current output but ignore the effects of the auxiliary outputs, which greatly reduces the ability to obtain richer

information. The performance of LSTM for data imputation is poor. This may be due to the heavy reliance on the quality and quantity of the data. Moreover, the box plot of the top seven algorithms with the best performance is shown in Figure 11, which is applied to present the distribution characteristics of these methods.

The results of MGCP-R-D and MGCP-R are shown in Figure 12. We can observe that the shape of the MGCP-R-D is more gentle. Compared to MGCP, the trend of MGCP-R-D tends to fit the data globally. Because both MGCP-R and MGCP-R have the same covariance structure. MGCP-R regularize the relation among the current and auxiliary outputs. Based on MGCP-R, MGCP-R-D further consider the penalty of shape distortion. To sum up all the above analysis, we consider that MGCP-R-D have the best performance due to its borrowing strength from other outputs and the consideration of distortion loss. Especially, while the missing ratio is larger, the penalty of shape loss can enhance the capability for global fitting.

As shown in Figure 13, the proposed MGCP-R-D method has been applied to an actual running sinter plant.

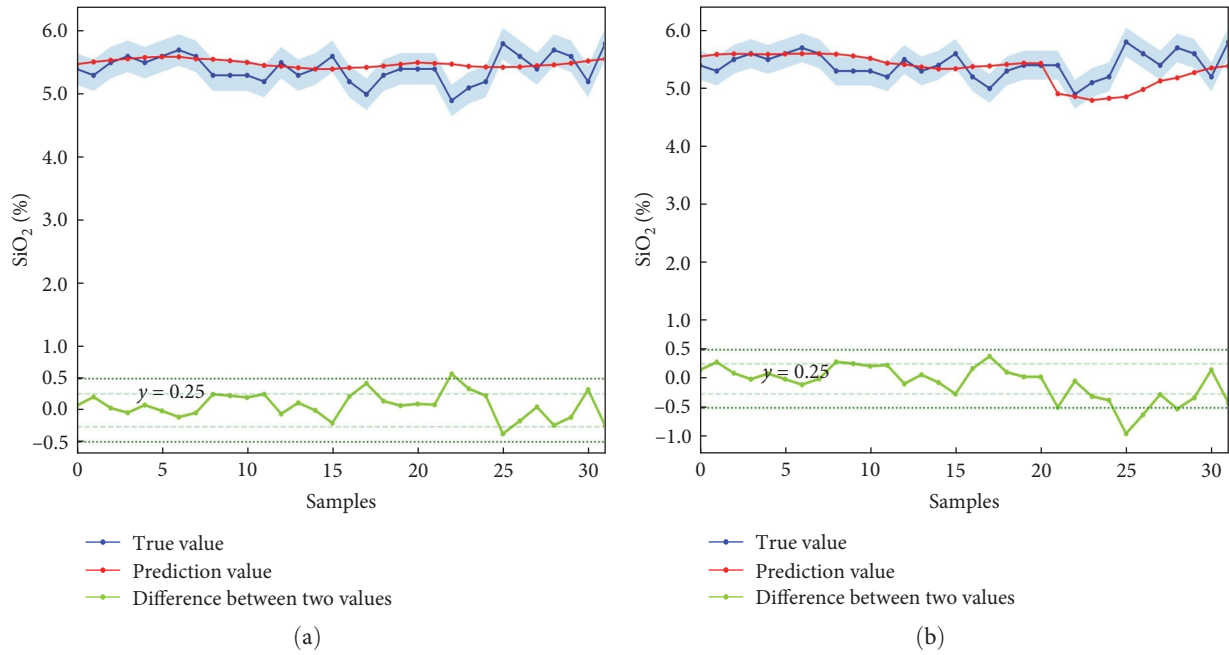


FIGURE 12: Results of proposed models. (a) MGCP-R-D and (b) MGCP-R.

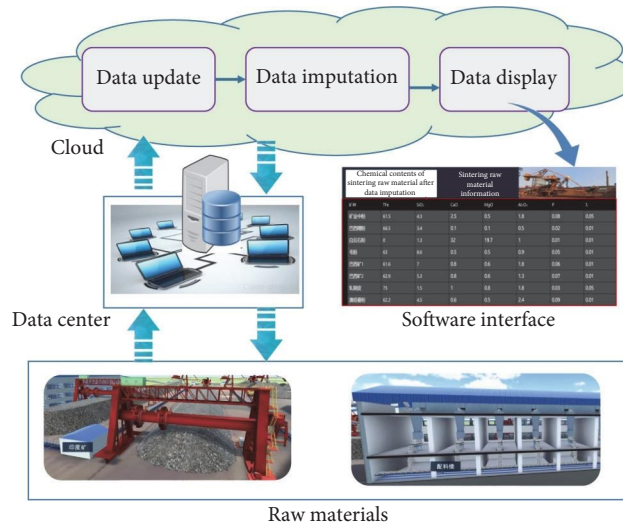


FIGURE 13: On-site software.

The chemical content information of the raw materials collected from the actual plant is transmitted to the data center. Data exchange between the cloud and data center is carried out to update the data timely. These missing values of these data are imputed by our proposed algorithm. Finally, these data are displayed on our designed software interface. This method achieves good performance in actual running.

## 5. Conclusion

The chemical contents of sinter ore material are of great significance to the monitoring and modeling of sintering process. In this paper, we proposed a novel MGCP-based framework to

deal with the data imputation of these chemical contents with a high missing ratio. The strength of the current output is enhanced by the knowledge transferred from the related outputs through the common latent function. Meanwhile, the negative effects of pseudo relationships and the missing situations over a long period are considered by introducing regularization item and dynamic time distortion, respectively. Compared with other related methods, the proposed method achieved the best estimation accuracy, which verify its effectiveness. Our approaches can also be used for a variety of industrial applications with large quantities of missing values.

According our experiments, the parameter selection and the outliers are the main challenges for our proposed

method. Especially, the inappropriate selection of parameters  $\lambda_1$  and  $\lambda_2$  will greatly decrease the precision. Moreover, The time interval between two missing points is an important factor to consider. Nowadays, some researches about the varying length of the missing interval have been reported. In the future, the effects of time interval will be further investigated based on the analysis of the characteristics of sintering process.

## Appendix

### Derivation of $\Theta_{jc}^y(\mathbf{x}, \mathbf{x}')$

$$\begin{aligned}
\Theta_{jc}^y(\mathbf{x}, \mathbf{x}') &= \text{cov}(y_j, y_c) \\
&= \mathbb{E}(y_j \cdot y_c) - \mathbb{E}(y_j) \cdot \mathbb{E}(y_c) \\
&= \mathbb{E}\left(\left(Q_{jj}(\mathbf{x}) \star R_j(\mathbf{x}) + \epsilon_j(\mathbf{x})\right) \times \left(\sum_{l=1}^q Q_{lc}(\mathbf{x}) \star R_l(\mathbf{x}) + \epsilon_c(\mathbf{x})\right)\right) \\
&= \mathbb{E}\left(\sum_{l=1}^q \int_{-\infty}^{\infty} Q_{lc}(\mathbf{u}) R_l(\mathbf{x} - \mathbf{u}) d\mathbf{u} \times \int_{-\infty}^{\infty} Q_{jj}(\mathbf{u}') R_j(\mathbf{x}' - \mathbf{u}') d\mathbf{u}'\right) \\
&\quad + \Sigma\tau_{jc} \\
&= \sum_{l=1}^q \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q_{lc}(\mathbf{u}) Q_{jj}(\mathbf{u}') \times \mathbb{E}(R_l(\mathbf{x} - \mathbf{u}) R_j(\mathbf{x}' - \mathbf{u}')) d\mathbf{u} d\mathbf{u}' \\
&\quad + \Sigma\tau_{jc} \\
&= \sum_{l=1}^q \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q_{lc}(\mathbf{u}) Q_{jj}(\mathbf{u}') \times \{\text{cov}(R_l(\mathbf{x} - \mathbf{u}), R_j(\mathbf{x}' - \mathbf{u}')) \\
&\quad - \mathbb{E}(R_l(\mathbf{x} - \mathbf{u})) \mathbb{E}(R_j(\mathbf{x}' - \mathbf{u}'))\} d\mathbf{u} d\mathbf{u}' + \Sigma\tau_{jc} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q_{jc}(\mathbf{u}) Q_{jj}(\mathbf{u}') \times \text{cov}(R_j(\mathbf{x} - \mathbf{u}), R_j(\mathbf{x}' - \mathbf{u}')) d\mathbf{u} d\mathbf{u}' \\
&\quad + \Sigma\tau_{jc} \\
&= \int_{-\infty}^{\infty} Q_{jc}(\mathbf{u} - \mathbf{t}) Q_{jj}(\mathbf{u}) d\mathbf{u} + \Sigma\tau_{jc},
\end{aligned} \tag{27}$$

where the second step is derived by the definition of covariance, the third step is based on the fact that the mean of the output ( $\mathbb{E}(y_j)$ ) is equal to zero and Equation (6), the fourth step is obtained by the definition of convolution and the independence between the function and measurement variance, the fifth step is the variables exchange, the seventh step is based on the mean of Gaussian white noise process  $R(\mathbf{x})$  is zero, and the eighth step is derived according to the  $\text{cov}(R_j(\mathbf{x}), R_j(\mathbf{x}')) = \delta(\mathbf{x} - \mathbf{x}')$  and  $\delta$  is the Dirac delta function.

Combined with Equation (10):

$$\begin{aligned}
Q_{kj}(x) &= \alpha_{kj} \left(4\pi\mu_{kj}^2\right)^{-\frac{1}{4}} \mathcal{N}\left(t|0, \mu_{kj}^2\right) \\
&= \alpha_{kj} \pi^{-\frac{1}{4}} |\mu_{kj}|^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2\mu_{kj}^2}\right),
\end{aligned} \tag{28}$$

we can derive the final expression of  $\Theta_{jc}^y(\mathbf{x}, \mathbf{x}')$ :

$$\Theta_{jc}^y(\mathbf{x}, \mathbf{x}') = \alpha_{jj} \alpha_{jc} \frac{2^{\frac{1}{2}} |\mu_{jj} \mu_{jc}|^{\frac{1}{2}}}{|\mu_{jj}^2 + \mu_{jc}^2|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{x} - \mathbf{x}')^2}{2(\mu_{jj}^2 + \mu_{jc}^2)}\right] + \Sigma\tau_{jc}. \tag{29}$$

The derivation processes of  $\Theta_{jj}^y(\mathbf{x}, \mathbf{x}')$  and  $\Theta_{jj}^y(\mathbf{x}, \mathbf{x}')$  are similar to that of  $\Theta_{jc}^y(\mathbf{x}, \mathbf{x}')$ .

### Data Availability

The data related to this study are available from the corresponding author with reasonable request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest.

### References

- [1] D. Fernández-González, I. Ruiz-Bustanza, J. Mochón, C. González-Gasca, and L. F. Verdeja, "Iron ore sintering: process," *Mineral Processing and Extractive Metallurgy Review*, vol. 38, no. 4, pp. 215–227, 2017.
- [2] Y. Jiang, N. Yang, Q. Yao, Z. Wu, and W. Jin, "Real-time moisture control in sintering process using offline-online NARX neural networks," *Neurocomputing*, vol. 396, pp. 209–215, 2020.
- [3] N. Chen, F. Hu, J. Chen, K. Wang, C. Yang, and W. Gui, "A monitoring method based on FDALM and its application in the sintering process of ternary cathode material," *Sensors*, vol. 22, no. 19, Article ID 7203, 2022.
- [4] J. Park, S. Lee, and J. Y. Park, "Review of computational fluid dynamics modeling of iron sintering process," *Journal of Mechanical Science and Technology*, vol. 36, pp. 4501–4508, 2022.
- [5] Y. Chen, Y. Huang, and J. Song, "Robust sparse time-frequency analysis for data missing scenarios," *IET Signal Processing*, vol. 17, no. 1, Article ID e12184, 2023.
- [6] H. Lin and S. Sun, "Distributed fusion estimator for multi-sensor asynchronous sampling systems with missing measurements," *IET Signal Processing*, vol. 10, no. 7, pp. 724–731, 2016.
- [7] L. Stanković, M. Daković, and S. Vujović, "Adaptive variable step algorithm for missing samples recovery in sparse signals," *IET Signal Processing*, vol. 8, no. 3, pp. 246–256, 2014.
- [8] J. Li, C. Hua, Y. Yang, L. Zhang, and X. Guan, "Output space transfer based multi-input multi-output takagi-sugeno fuzzy modeling for estimation of molten iron quality in blast furnace," *Knowledge-Based Systems*, vol. 219, Article ID 106906, 2021.
- [9] N. Mourad, "Robust smoothing of one-dimensional data with missing and/or outlier values," *IET Signal Processing*, vol. 15, no. 5, pp. 323–336, 2021.
- [10] R. Little and D. B. Rubin, "Statistical analysis with missing data," *Technometrics*, vol. 45, no. 4, pp. 364–365, 2002.

- [11] G. Zhang and R. Little, "Extensions of the penalized spline of propensity prediction method of imputation," *Biometrics*, vol. 65, no. 3, pp. 911–918, 2009.
- [12] H. Demirtas, "Multiple imputation under bayesianly smoothed pattern-mixture models for non-ignorable drop-out," *Statistics in Medicine*, vol. 24, no. 15, pp. 2345–2363, 2005.
- [13] M. Stockmann, R. Haber, and U. Schmitz, "Source identification of plant-wide faults based on  $k$  nearest neighbor time delay estimation," *Journal of Process Control*, vol. 22, no. 3, pp. 583–598, 2012.
- [14] P. Zhang, P. Ren, Y. Liu, and H. Sun, "Autoregressive matrix factorization for imputation and forecasting of spatiotemporal structural monitoring time series," *Mechanical Systems and Signal Processing*, vol. 169, Article ID 108718, 2022.
- [15] 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 Transactions on Cybernetics*, vol. 53, no. 2, pp. 695–706, 2023.
- [16] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific Reports*, vol. 8, Article ID 6085, 2018.
- [17] M. Wu, X. Chen, W. Cao, J. She, and C. Wang, "An intelligent integrated optimization system for the proportioning of iron ore in a sintering process," *Journal of Process Control*, vol. 24, no. 1, pp. 182–202, 2014.
- [18] J. Hu, M. Wu, X. Chen et al., "A multilevel prediction model of carbon efficiency based on the differential evolution algorithm for the iron ore sintering process," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 11, pp. 8778–8787, 2018.
- [19] W. E. Noori and A. S. Albahri, "Towards trustworthy myopia detection: integration methodology of deep learning approach, XAI visualization, and user interface system," *Applied Data Science and Analysis*, vol. 2023, pp. 1–15, 2023.
- [20] M. Alajjabi, D. Malerba, and H. Liu, "Distributed reduced convolution neural networks," *Mesopotamian Journal of Big Data*, vol. 2021, pp. 25–28, 2021.
- [21] A. H. Ali, H. Kumar, and P. J. Soh, "Big data sentiment analysis of twitter data," *Mesopotamian Journal of Big Data*, vol. 2021, pp. 1–5, 2021.
- [22] J. Li, C. Hua, and Y. Yang, "Output space transfer-based MIMO RVFLNs modeling for estimation of blast furnace molten iron quality with missing indexes," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.
- [23] J. Li, C. Hua, Y. Yang, and X. Guan, "A novel MIMO T-S fuzzy modeling for prediction of blast furnace molten iron quality with missing outputs," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 6, pp. 1654–1666, 2021.
- [24] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, 2003.
- [25] J. Quinonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate gaussian process regression," *Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, 2005.
- [26] H. Liu, J. Cai, and Y.-S. Ong, "Remarks on multi-output gaussian process regression," *Knowledge-Based Systems*, vol. 144, pp. 102–121, 2018.
- [27] A. A. Chehade and A. A. Hussein, "A multioutput convolved Gaussian process for capacity forecasting of Li-ion battery cells," *IEEE Transactions on Power Electronics*, vol. 37, no. 1, pp. 896–909, 2022.
- [28] F. Rodrigues, K. Henrickson, and F. C. Pereira, "Multi-output gaussian processes for crowdsourced traffic data imputation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 594–603, 2019.
- [29] W. Jiang, N. Zheng, and I. Kim, "Missing data imputation for transfer passenger flow identified from in-station WiFi systems," *Transportmetrica B: Transport Dynamics*, vol. 11, no. 1, pp. 325–342, 2023.
- [30] T. Hori, D. Montcho, C. Agbangla, K. Ebana, K. Futakuchi, and H. Iwata, "Multi-task Gaussian process for imputing missing data in multi-trait and multi-environment trials," *Theoretical and Applied Genetics*, vol. 129, pp. 2101–2115, 2016.
- [31] R. Kontar, S. Zhou, C. Sankavaram, X. Du, and Y. Zhang, "Nonparametric modeling and prognosis of condition monitoring signals using multivariate Gaussian convolution processes," *Technometrics*, vol. 60, no. 4, pp. 484–496, 2018.
- [32] X. Emery, "Iterative algorithms for fitting a linear model of coregionalization," *Computers & Geosciences*, vol. 36, no. 9, pp. 1150–1160, 2010.
- [33] O. Babak and C. V. Deutsch, "An intrinsic model of coregionalization that solves variance inflation in collocated cokriging," *Computers & Geosciences*, vol. 35, no. 3, pp. 603–614, 2009.
- [34] X. Wang, C. Wang, X. Song, L. Kirby, and J. Wu, "Regularized multi-output gaussian convolution process with domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6142–6156, 2023.
- [35] P. Kasarla, C. Wang, T. L. Brown, and D. McGehee, "Modeling and prediction of driving performance measures based on multi-output convolutional Gaussian process," *Accident Analysis & Prevention*, vol. 161, Article ID 106360, 2021.
- [36] M. Ç. Pınar and W. M. Hartmann, "Huber approximation for the non-linear  $\ell_1$  problem," *European Journal of Operational Research*, vol. 169, no. 3, pp. 1096–1107, 2006.
- [37] M. Cuturi and M. Blondel, "Soft-DTW: a differentiable loss function for time-series," in *Proceedings of the 34th International Conference on Machine Learning*, pp. 894–903, PMLR, 2017.
- [38] A. G. D. G. Matthews, M. van der Wilk, T. Nickson et al., "Gpflow: a Gaussian process library using tensorflow," *Journal of Machine Learning Research*, vol. 18, no. 40, pp. 1–6, 2017.
- [39] Kdd Cup, 2018, <http://www.kdd.org/kdd2018/>.
- [40] X. Jing, J. Luo, J. Wang, G. Zuo, and N. Wei, "A multi-imputation method to deal with hydro-meteorological missing values by integrating chain equations and random forest," *Water Resources Management*, vol. 36, pp. 1159–1173, 2022.
- [41] F. Sağlam, T. Şanlı, M. A. Cengiz, and Y. Terzi, "Alternative expectation approaches for expectation-maximization missing data imputations in cox regression," *Communications in Statistics-Simulation and Computation*, 2022.
- [42] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [43] I. R. White, P. Royston, and A. M. Wood, "Multiple imputation using chained equations: issues and guidance for practice," *Statistics in Medicine*, vol. 30, no. 4, pp. 377–399, 2011.
- [44] A. Rubinsteyn and S. Feldman, "Fancyimpute: an imputation library for python," [Online]. Available: <https://github.com/iskandr/fancyimpute>.



- [45] Y. Luo, X. Cai, Y. Zhang, J. Xu, and Y. Xiaojie, "Multivariate time series imputation with generative adversarial networks," in *Advances in Neural Information Processing Systems*, vol. 31, pp. 1603–1614, Curran Associates, Inc., 2018.
- [46] Z. Guo, Y. Wan, and H. Ye, "A data imputation method for multivariate time series based on generative adversarial network," *Neurocomputing*, vol. 360, pp. 185–197, 2019.