



Article

HSIC-DIMFMC: A Multi-View Functional Matrix Completion Method with Dual-Information Graph Regularization for Meteorological Data Imputation

Haiyan Gao ^{1,2,*} and Youdi Bian ¹

¹ School of Statistics and Data Science, Lanzhou University of Finance and Economics, Lanzhou 730020, China; byd@lzufe.edu.cn

² Key Laboratory of Digital Economy and Social Computing Science of Gansu, Lanzhou 730020, China

* Correspondence: gaohy@lzufe.edu.cn

Abstract

Continuous and complete meteorological observations are essential for reliable climate analysis and environmental assessment. However, missing values caused by sensor malfunctions and transmission failures can introduce systematic biases and increase uncertainty in downstream applications. Meteorological variables can be modeled as functional data and typically exhibit nonlinear inter-variable dependencies alongside temporal smoothness; these properties provide valuable prior information for missing data recovery. To address this issue, we propose HSIC-DIMFMC, a multi-view functional matrix completion method for meteorological data imputation that integrates the Hilbert–Schmidt Independence Criterion (HSIC) and dual-information graph regularization. Within a unified framework of functional data analysis and multi-view learning, HSIC is utilized to capture nonlinear dependencies across multiple views, while dual-information graph regularization preserves local structural relationships and temporal smoothness. This joint modeling strategy significantly improves latent representation learning and enhances imputation performance. Experiments on real meteorological datasets demonstrate that the proposed method consistently outperforms several state-of-the-art baselines, especially for strongly correlated variable pairs such as temperature–dew point and wind speed–maximum wind speed. Compared with seven representative approaches—ranging from traditional spatial interpolation to advanced functional matrix completion models—HSIC-DIMFMC achieves average reductions of 43.97–73.59% in RMSE. The results indicate that HSIC-DIMFMC effectively exploits nonlinear cross-view dependencies and structural information, providing a robust solution for collaborative meteorological data imputation.

Keywords: functional data analysis; multi-view learning; HSIC; dual-information graph regularization; matrix completion; meteorological data imputation

Academic Editor: Giuseppe Ciaburro

Received: 10 April 2026

Revised: 2 July 2026

Accepted: 4 July 2026

Published: 8 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Meteorological data serve as the foundation for monitoring climate change, forecasting weather, and supporting meteorological research, playing a critical role in environmental protection and sustainable development. However, data incompleteness or inaccuracies may arise due to sensor malfunctions, data transmission disruption, severe

weather conditions impairing device functionality, or challenging geographical locations and topographical factors [1]. These issues compromise data quality and reliability, thereby adversely affecting climate analysis and assessment. Consequently, it is essential to develop efficient data imputation techniques to fill the missing portions in meteorological records.

The application objectives of data, missing patterns, and the availability of multi-source observational data all influence the selection of imputation methods. Missing data imputation can be performed using simple techniques such as mean or median values, or by leveraging correlational information among data points to enhance imputation accuracy. For instance, employing the Grouped Method of Data Handling (GMDH) [2] for imputing missing meteorological data reduced imputation errors [3], but it may become trapped in local optima and suffer from overfitting issues. Meanwhile, the level of information utilization affects imputation precision; relying solely on data from a single station cannot fully represent regional climate patterns [4]. Although some studies have considered inter-station information through multi-station joint modeling of the same variables, their frameworks fail to extend to multiple meteorological indicators, thereby neglecting the inherent correlations among different elements [5]. Spatial imputation methods based on spatial correlation [6], which assume similarity between missing data and surrounding known data, perform well for short-term missing data but are unsuitable for scenarios with missing data spanning several consecutive months. Recent studies indicate that homogenization methods [7] perform well in imputing air temperature series. For missing daily temperature data at regional stations, a comparative study of Inverse Distance Weighting (IDW), Ordinary Kriging (OK), and Multiple Linear Regression (MLR) demonstrated that MLR yielded the best imputation results [8]. Furthermore, the latest research has explored integrating high-resolution numerical simulations with traditional Kriging to enhance temperature interpolation accuracy in complex terrains [9]. Least squares regression [10] is commonly applied to reconstruct missing daily meteorological records, but while it enhances predictive capability for normal meteorological data, it may compromise estimation accuracy for extreme meteorological data. Furthermore, factors such as monitoring station density [11] and weather classification [12] can influence the accuracy of missing value estimates, and the manner in which information is mined and processed also affects data imputation outcomes. For instance, regression imputation methods utilizing inter-station information yield smaller errors compared to those utilizing intrastation information [13].

Meteorological data, by nature, constitutes time-varying sequences, with sampled data at adjacent time points typically exhibiting significant correlations. These temporal and autocorrelational characteristics provide a foundation for the application of time series analysis methods. For instance, in wind speed forecasting for wind farms, commonly employed methods include neural network methods [14] and time series techniques [15]. Neural networks, owing to their strengths in non-linear modeling and feature extraction, have found extensive application in time series analysis. For example, the Residual Inference by Short Paths-Long Short-Term Memory (RIMP-LSTM) method [16] achieves effective imputation of time series data by simulating latent correlations between missing items and historical observations. However, RIMP-LSTM may experience a notable decline in performance when handling complex or extreme missing scenarios, such as large-scale consecutive missing data or random block missing data. Furthermore, although recurrent neural networks (RNNs) [17] offer a nonparametric approach for time series missing value imputation, their ability to model long-term dependencies remains limited. To address this limitation, the Gated Recurrent Unit with Decay (GRU-D) [18], the recurrent neural network model for sparse multivariate time series, incorporates information on temporal dependencies and missing patterns, enabling it to capture long-term temporal

relationships and achieve favorable results in prediction and imputation tasks. Nevertheless, the performance of the GRU-D may also be constrained when dealing with lengthy missing segments in time series data.

Due to the inherent difficulties in time series data imputation, deep learning methods have become widely adopted for handling missing data tasks. For instance, a Sequence-to-Sequence Imputation Model (SSIM) [19] demonstrates smaller error compared to methods such as ARIMA [20] and Seasonal ARIMA, and is capable of handling various missing patterns. However, the model's efficacy relies heavily on both the volume and the quality of the available training data, and its imputation effectiveness may decline for longer consecutive missing segments due to insufficient contextual information. Although LSTM-based deep learning methods generally outperform traditional ARIMA models in temperature data imputation, their imputation performance remains highly dependent on specific model architectures [21]. Most of these imputation methods for meteorological data are limited to processing single meteorological indicators or data from single meteorological stations, resulting in low information utilizations. They primarily focus on exploiting shared information among data while neglecting the differential information between them. Concurrently, deep learning methods require substantial data for model training and are prone to risks of local optima and overfitting. Furthermore, when dealing with large-scale or continuous blocks of missing data, their imputation effectiveness diminishes.

Recently, deep learning has made significant progress in time-series imputation. For instance, BRITS and its variants capture long-range dependencies via bidirectional recurrent structures [17]; SAITS achieves efficient imputation using self-attention mechanisms and diagonal masking strategies; and GAIN introduces a generative adversarial framework to enhance imputation realism. The performance differences among these methods have been systematically evaluated in air quality prediction tasks [22]. Regarding Transformer architectures, ImputeFormer [23] improves the generalization of spatiotemporal imputation through a low-rank-induced mechanism, while SERT [24] provides a dedicated missing-value-handling scheme for environmental monitoring sensors. Additionally, Graph Neural Networks (GNNs) have shown immense potential in meteorological sensor imputation due to their advantage in explicitly modeling spatial correlations [25]. Although these methods offer novel perspectives for reconstructing complex meteorological data, they generally face challenges such as inadequate adaptability to extreme continuous missing scenarios and insufficient exploitation of nonlinear synergies among multiple variables.

Meteorological data exhibits temporal characteristics, continuously evolving over time. These discrete observations can be represented as continuous functions through curve fitting, enabling the exploration of meteorological data via functional data analysis (FDA) methods. Given the real-time collection of meteorological data, substantial volumes are generated. However, data transmission interruptions, data entry errors, and other factors result in contiguous blocks of missing data. To effectively reconstruct such missing data, matrix completion has been widely adopted as an imputation technique for large-scale data gaps in practical applications [26]. Moreover, studies have shown that functional data imputation problems are equivalent to rank-constrained matrix completion [27]. Moreover, meteorological data can be measured through indicators from multiple views, such as temperature, dew point temperature, and wind speed, and other indicators, with each view typically providing complementary information [28]. Since multi-view learning generally achieves superior performance compared to single-view approaches, integrating information from multiple perspectives allows for more precise extraction of data features than depending on a single view [29]. Consequently, effectively harnessing the complementary effects of multi-view data can substantially improve the

accuracy and reliability of meteorological feature extraction. Xue et al. [30] also attempted to apply the Multi-View Functional Matrix Completion model (MVNFMC) to meteorological data processing, demonstrating its favorable data imputation effects. Considering the complementary information among samples and their high-order neighborhood information, Gao and Ma [31] proposed another multi-view functional matrix completion method (GRMFMC) and applied it to air quality data imputation. Compared to typical imputation methods, this model exhibits a marked advantage in data imputation accuracy.

1.1. Motivation and Method Construction

Although existing multi-view functional matrix completion methods have shown effectiveness in missing data recovery, they still face important limitations in meteorological data imputation. Traditional graph regularization methods are usually based on the local manifold assumption and preserve local geometric structure by constructing k -nearest-neighbor graphs. Such methods are useful for characterizing spatial relationships among meteorological stations and temporal similarity in observational sequences. However, they mainly focus on local similarity and short-range correlations, while often overlooking globally discriminative information across different climatic regions [32,33]. This limitation may lead to unstable imputation results for anomalous samples, such as extreme weather observations, and reduces the ability of the model to exploit inter-sample heterogeneity [34].

Moreover, conventional graph regularization typically relies on a single neighborhood construction strategy, which makes it difficult to capture the complex nonlinear structure of meteorological data. To address this issue, this study adopts a dual-information graph regularization strategy that combines local manifold information and discriminative manifold information. This design preserves local consistency among adjacent stations or neighboring time periods while enhancing the discriminability of learned representations, thereby enabling a more comprehensive description of multi-scale nonlinear characteristics in meteorological data [35]. In addition, to enhance the plausibility and generalization capability of the imputation results, we introduce the Hilbert–Schmidt Independence Criterion (HSIC). Compared with traditional linear dependence measures, such as the Pearson correlation coefficient, HSIC can effectively characterize complex nonlinear dependencies among meteorological variables and extract complementary information across different views [36].

The imputation of meteorological data faces a threefold complexity, rendering traditional baseline methods—such as mean imputation, linear regression, and single spatial interpolation—inadequate: (1) Complex nonlinear coupling relationships exist among meteorological variables like temperature, dew point, and wind speed, which linear models (e.g., MLR) or distance-weighting methods (e.g., IDW) fail to capture effectively; (2) Meteorological data inherently couple temporal dynamics (e.g., diurnal and seasonal variations) with spatial heterogeneity (e.g., topographical and elevational impacts). Traditional time-series methods often ignore spatial topologies, while spatial interpolation methods struggle to capture temporal evolution; (3) Real-world meteorological observations frequently suffer from continuous block missingness across time or multi-station missingness. Under such a severe lack of local contextual information, the performance of simple methods degrades drastically. Consequently, these challenges necessitate an imputation model equipped with capabilities for nonlinear modeling, spatiotemporal synergy, and multi-source joint inference—requirements that no single conventional method can simultaneously fulfill.

Motivated by the fact that meteorological states are jointly characterized by multiple interacting indicators, we propose HSIC-DIMFMC, a novel multi-view functional matrix

completion method for meteorological data imputation under the unified framework of functional data analysis and multi-view learning. The proposed method integrates HSIC and dual-information graph regularization to jointly model nonlinear dependence, structural information, and temporal smoothness. The overall architecture of HSIC-DIMFMC is shown in Figure 1.

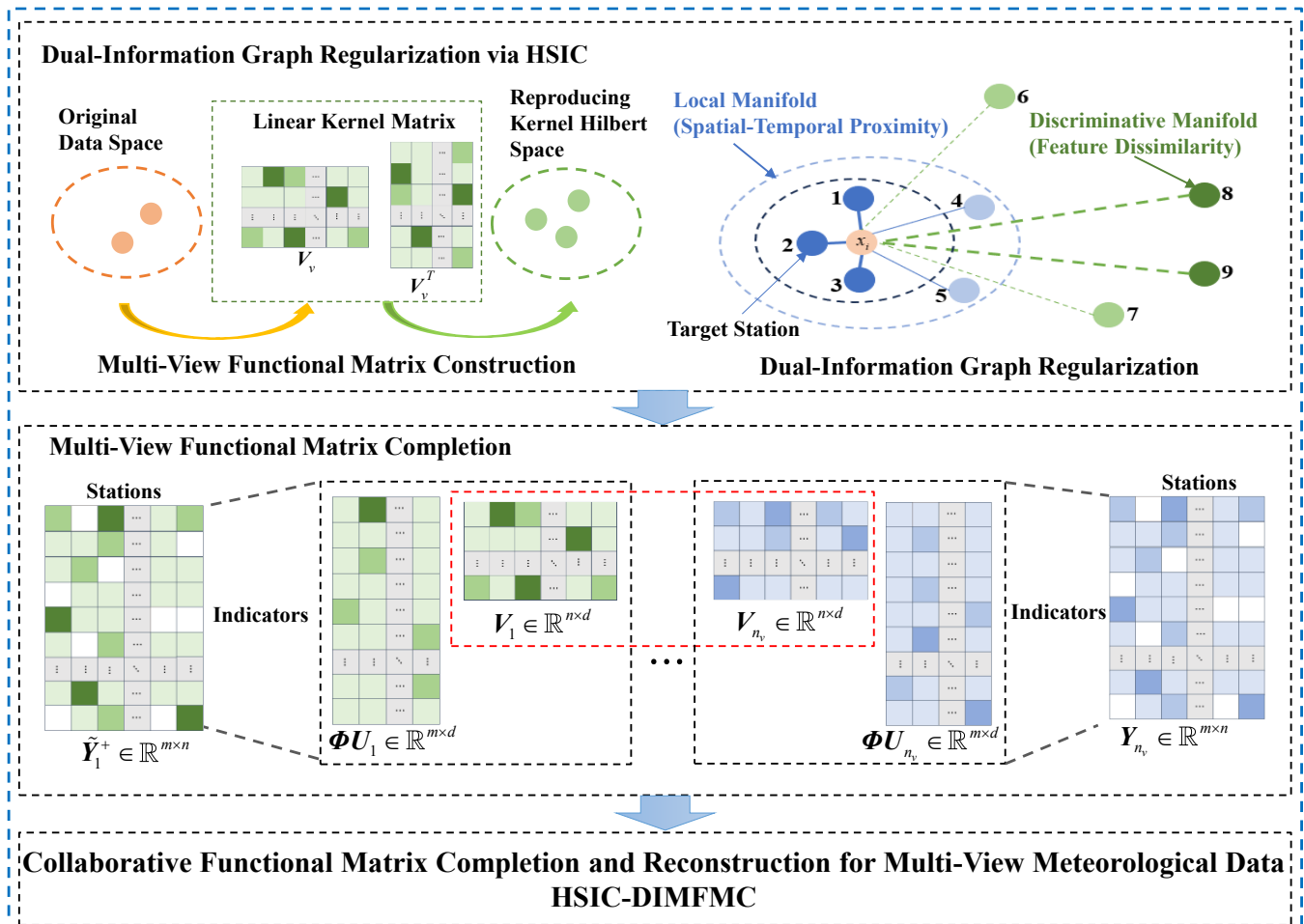


Figure 1. The overall architecture of the proposed HSIC-DIMFMC model.

Unlike traditional approaches that treat HSIC or graph regularization merely as independent optimization tools, the core contribution of this paper lies in revealing and deeply exploiting the synergistic complementarity among the HSIC criterion, the local manifold graph, and the discriminative manifold graph. Specifically, while functional matrix completion and multi-view learning establish the foundational inference framework, they cannot effectively characterize nonlinear coupling and global discriminative structures. To overcome this, our method employs the local manifold graph to preserve spatiotemporal continuity, utilizes the discriminative manifold graph to enhance categorical distinctiveness, and embeds the HSIC criterion to specifically capture deep nonlinear dependencies among variables. The integration of these three components is not a mere stacking of techniques; rather, it is a targeted theoretical design tailored to address the threefold dilemma of meteorological data: nonlinear coupling, spatiotemporal heterogeneity, and a lack of categorical discriminative information.

The main contributions of this work are summarized as follows:

(1) We propose a novel multi-view functional matrix completion method for meteorological data imputation by integrating HSIC and dual-information graph regularization into a unified framework of functional data analysis and multi-view learning.

(2) The proposed method captures complex nonlinear dependencies among meteorological indicators, preserves local manifold structure and temporal consistency, and enhances the discriminability of latent representations for collaborative imputation.

(3) Extensive experiments under simulated missing data settings and on real meteorological datasets demonstrate that HSIC-DIMFMC consistently outperforms several representative baseline methods in terms of RMSE, MAE, and NRMSE, while maintaining robust performance across different missing ratios and meteorological variables.

1.2. Structure of This Paper

The remainder of this paper is organized as follows. Section 2 introduces the characteristics of the meteorological data used in this study. Section 3 reviews the relevant background on curve fitting and functional matrix completion and then presents the construction of the proposed HSIC-DIMFMC method, including the iterative optimization procedure, convergence analysis, and computational complexity analysis. Section 4 evaluates the effectiveness of the proposed method through experiments under simulated missing data settings and real-world meteorological applications. Finally, Section 5 concludes the paper and discusses future research directions.

2. Meteorological Data Description

With continuous advancements in detection technology and ever-increasing computational processing capabilities, the scope of meteorological data observation and forecasting has expanded significantly, achieving markedly higher levels of refinement. This not only broadens the geographical coverage and enhances the spatiotemporal density of data but also introduces richer observation types and more diverse data presentation formats. Common meteorological indicators include temperature, visibility, humidity, air pressure, and wind speed. This paper includes temperature, dew point temperature, wind speed, and maximum wind speed for analysis as the indicators for study. The rationale for this selection is based on the following considerations: Firstly, temperature and dew point temperature share a physical correlation in meteorology, jointly characterizing the thermal and humidity state of the atmosphere, exhibiting strong synergistic variation patterns and non-linear correlations. Secondly, wind speed and maximum wind speed both describe atmospheric motion characteristics, demonstrating high temporal consistency and complementarity. Secondly, significant physical and statistical correlations exist between TEMP and DEWP, as well as between WDSP and MXSPD. These correlations facilitate the construction of an effective shared latent representation. The Higher-Order Statistical Correlation (HSIC) metric can be employed to measure the higher-order statistical dependencies between indicators, while Dual-information Graph Regularization enables more effective fusion of inter-sample information. Finally, the selection of these four meteorological indicators sufficiently validates the efficacy of the HSIC-DIMFMC method in meteorological data imputation.

The meteorological data presented in this study are openly available in the National Climatic Data Center (NCDC) at <https://www.ncei.noaa.gov/data/global-summary-of-the-day/archive> (accessed on 1 June 2026). We selected three independent batches of meteorological station datasets covering regions in China, which are specifically dedicated to the visual analysis of missing patterns, the quantitative evaluation of imputation accuracy, and the validation of the model's generalization capability. These datasets comprise surface daily observation records, primarily encompassing core meteorological variables such as temperature (TEMP), dew point temperature (DEWP), wind speed (WDSP), and maximum wind speed (MXSPD). Data pertaining to four meteorological variables-TEMP, DEWP, WDSP, and MXSPD-were collected from 30 stations in western China, covering the period from 1 January 2024 to 13 September 2024. A visualization of the data distribu-

tion is presented in Figure 2. Upon examining Figure 2, it is clear that these four meteorological indicators exhibit varying degrees of missing values, including both randomly distributed missing data and contiguous striped or block-like missing patterns. Therefore, the issue of missing data in meteorological datasets represents a prevalent quality concern.

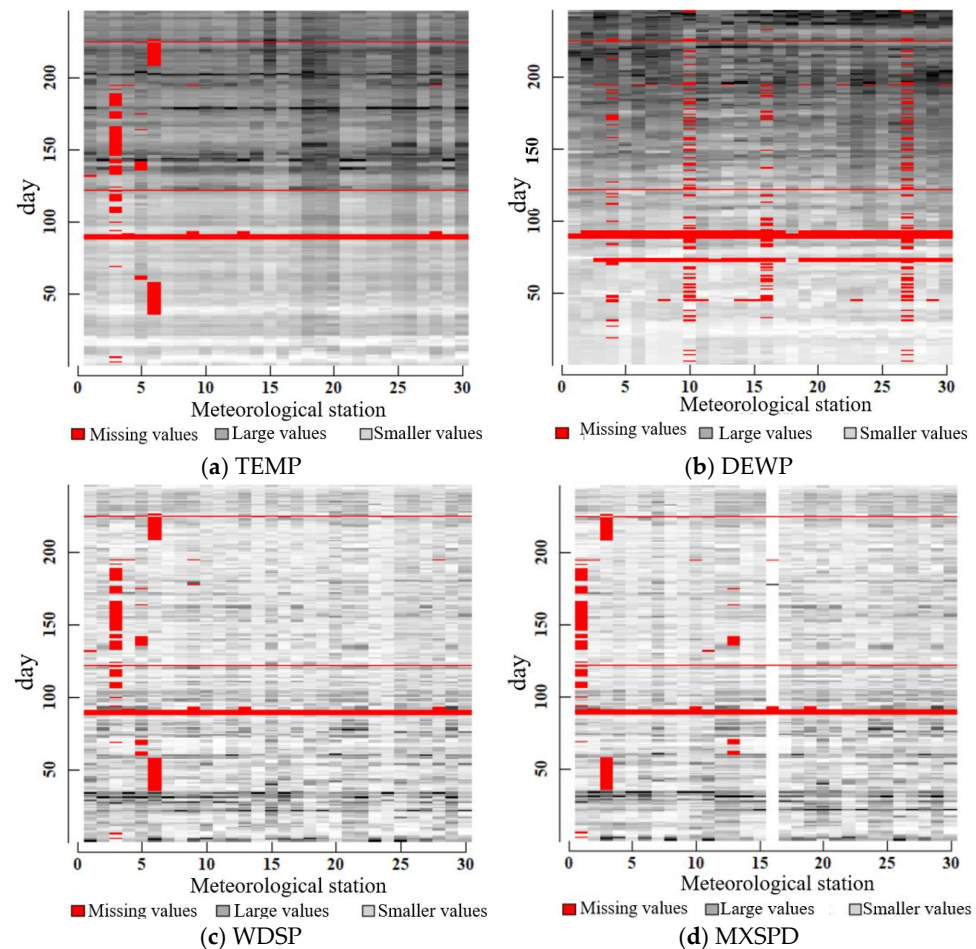


Figure 2. Visualization of the data absence situation of meteorological indicators.

2.1. Datasets I

Complete meteorological data for summer Temperature (TEMP), Dew Point Temperature (DEWP), Wind Speed (WDSP), and Maximum Wind Speed (MXSPD) were chosen from 20 stations in Northeast China. The precise locations of these stations are indicated as A to T in Figure 3. This map was generated using the standard map with approval number GS(2024)0650 from the National Geographic Information Public Service Platform (Tianditu), with the base map content unchanged.

This paper employs four key indicators—TEMP, DEWP, WDSP, and MXSPD—to represent different ways of characterizing meteorological conditions. Figure 4 displays the continuous trajectories of these indicators over time, with their smooth curve patterns providing intuitive confirmation of the inherent functional characteristics within the meteorological data.

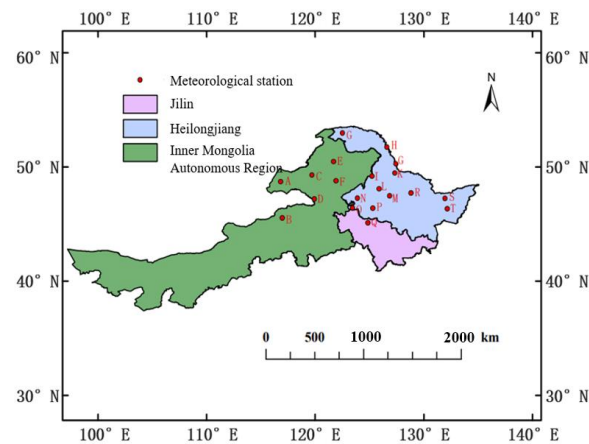


Figure 3. Distribution map of the 20 meteorological stations included in Dataset I, with annotated letters (A, B, C, ...) marking the corresponding stations on the plot.

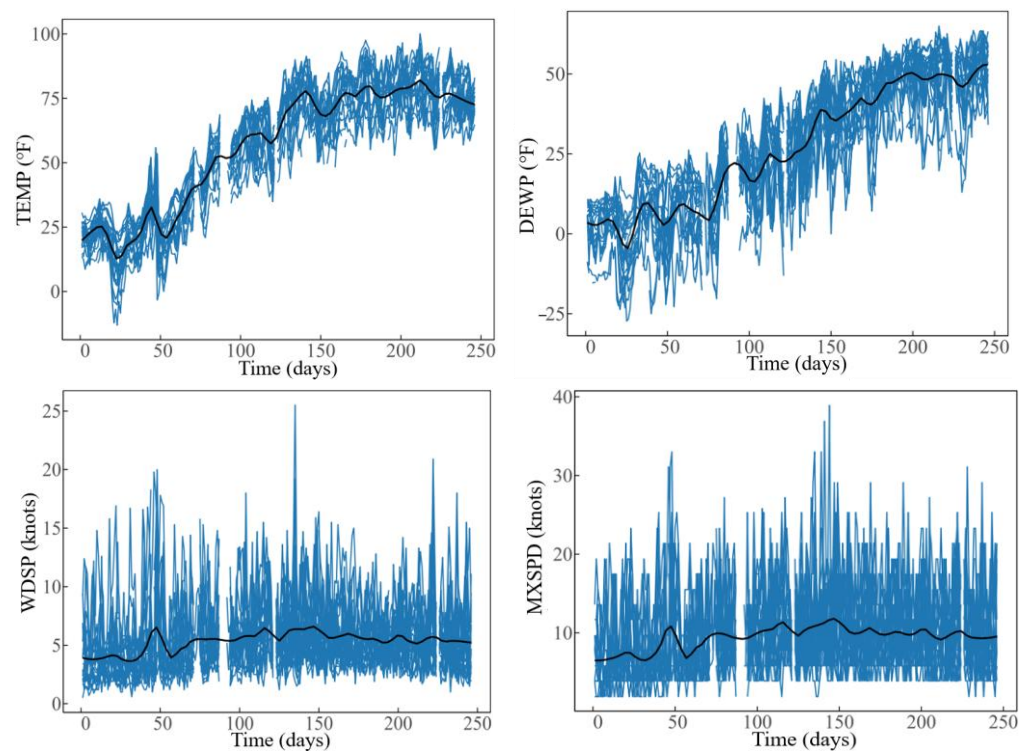


Figure 4. Trend of numerical changes of meteorological indicators. The blue lines show the observed values, and the black line indicates their mean.

To verify the underlying correlations among different meteorological variables, Figure 5 presents the Pearson correlation coefficient heatmaps between TEMP and DEWP, and between WDSP and MXSPD. The results reveal a correlation coefficient of 0.8 between TEMP and DEWP, while WDSP exhibits a remarkably high correlation coefficient of 0.95 with MXSPD. This indicates that both sets of indicators demonstrate strong positive correlations internally. Such robust correlations provide a theoretical foundation for multi-view learning, ensuring that information from different views possesses sufficient complementarity. Consequently, this supports the rationality and effectiveness of missing data imputation based on inter-view dependencies.

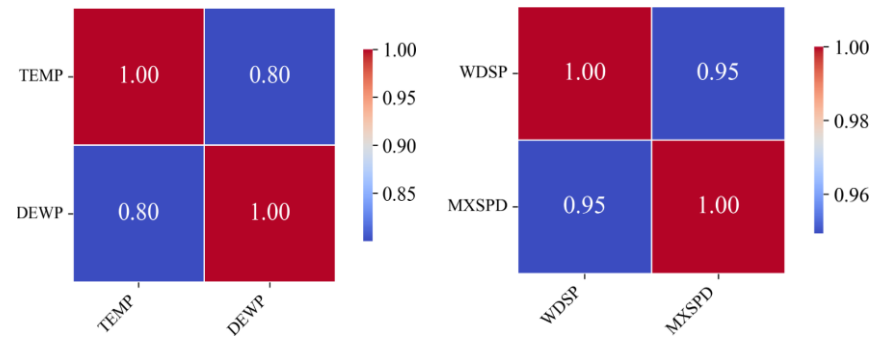


Figure 5. Heatmap of Pearson correlation coefficients of meteorological indicators. The **left** figure shows the heat map of the correlation coefficient between TEMP and DEWP, while the **right** figure shows the heat map of the correlation coefficient between WDSP and MXSPD.

As shown in Figure 6, plotting paired meteorological indicator relationship diagrams further illustrates the correlation between indicator pairs. The diagonal lines represent histograms of indicator distributions. Observation reveals that TEMP and DEWP exhibit left-skewed distributions, while WDSP and MXSPD show right-skewed distributions. In skewed distributions, the mean may be elevated or depressed by extreme values, failing to represent typical conditions. As shown in Figure 6, The lower-left and upper-right corners display scatter plots of paired indicators. The scatter plots for TEMP versus DEWP and WDSP versus MXSPD both display a clear positive linear correlation, indicating a linear positive correlation between TEMP and DEWP, and between WDSP and MXSPD. The statistical correlation among meteorological indicators provides the theoretical basis for designing a missing data imputation method based on complementarity in this study.

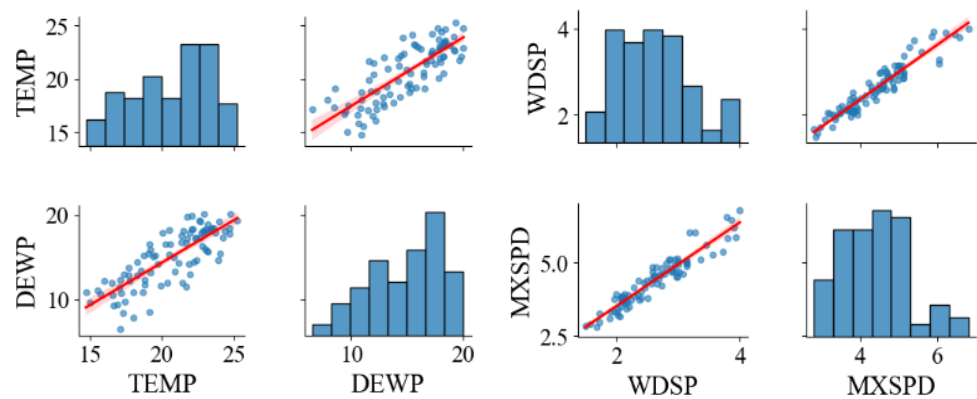


Figure 6. Pairwise relationship diagram of meteorological indicators in Dataset I.

2.2. Datasets II

Complete data for four meteorological indicators from 1 June to 31 August 2024, were collected from 20 meteorological stations in Sichuan Province, Guizhou Province, Chongqing Municipality, Hunan Province, Hubei Province, and Jiangxi Province. Table 1 presents the basic information for 20 stations. The statistical description of the data is shown in Table 2.

Table 1. Basic information for 20 stations.

Station Number	Station Name	Latitude and Longitude Co-ordinates	Station Number	Station Name	Latitude and Longitude Co-ordinates
57776099999	NANYUE	(27.3 N, 112.7 E)	57655099999	YUANLING	(28.4666666 N, 110.4 E)
57476099999	JIANGLING	(30.3333333 N, 112.1833333 E)	57662099999	CHANGDE	(29.05 N, 111.6833333 E)
57494099999	TIANHE	(30.783758 N, 114.2081 E)	57687099999	CHANGSHA	(28.1166666 N, 112.7833333 E)

57503099999	NEIJIAN	(29.6166666 N, 105.1166666 E)	57707099999	BIJIE	(27.3 N, 105.2333333 E)
57516099999	JIANGBEI	(29.719217 N, 106.641678 E)	57712099999	HUICHUAN	(27.7333333 N, 106.95 E)
57554099999	SANGZHI	(29.4 N, 110.1666666 E)	57731099999	SINAN	(27.95 N, 108.25 E)
57584099999	YUEYANG	(29.3833333 N, 113.0833333 E)	57745099999	ZHIJIANG	(27.45 N, 109.6833333 E)
57598099999	XIUSHUI	(29.0333333 N, 114.5833333 E)	57766099999	SHAOYANG	(27.1833333 N, 111.45 E)
57604099999	NAXI	(28.7833333 N, 105.3833333 E)	57793099999	YICHUN	(27.8 N, 114.3833333 E)
57633099999	YOUYANG	(28.8333333 N, 108.7666666 E)	57799099999	JIAN	(27.1166666 N, 114.9666666 E)

The spatial distribution of stations in Dataset II differs from that in Dataset I, thereby validating the applicability of the proposed HSIC-DIMFMC method to meteorological data from diverse regions. Secondly, employing the complete dataset as a benchmark facilitates the construction of a controlled experimental environment for missing data. By artificially setting varying proportions, locations, and types of missing data, the performance differences among imputation methods can be systematically evaluated under a unified comparative standard. Furthermore, since the observations are known, imputation errors can be precisely quantified, thereby providing a reliable basis for objective comparisons between methods.

Table 2. Statistical description of meteorological data.

Variable	Number	Max	Min	Mean	SD	Q1	Median	Q3
TEMP (°C)	1840	36.11	13.11	27.61	4.15	24.78	28.00	31.00
DEWP (°C)	1840	29.06	2.61	22.61	2.73	20.97	23.22	24.61
WDSP (m/s)	1840	15.18	0.31	2.15	1.55	1.34	1.75	2.42
MXSPD (m/s)	1840	18.98	0.98	3.64	2.20	2.01	2.98	4.01

3. The Proposed Method

3.1. Curve Fitting of Meteorological Data

This study utilizes functional data analysis methods to investigate the stochastic characteristics of meteorological data. Monitoring data for meteorological indicators may be regarded as realizations of a random function, thereby enabling the imputation of missing values within the meteorological dataset. One dataset comprises four views: TEMP, DEWP, WDSP, and MXSPD. For a dataset $\mathbf{X}(t)$ comprising n monitoring stations, let $\{X_1(t), X_2(t), \dots, X_v(t), \dots, X_n(t)\}$ denote n_v viewpoints, where the data for the v -th view is $X_v(t) = \{\mathbf{x}_1^v(t), \mathbf{x}_2^v(t), \dots, \mathbf{x}_n^v(t)\}$ with a sample observation dimension of m_v , comprising n curves in total. The j -th observation $\tilde{y}_{ij}^v$ of curve i -th in view v -th is produced by the following general regression model (1):

$$\tilde{y}_{ij}^v = x_i^v(t_{ij}) + \varepsilon_{ij}^v \quad (1)$$

where ε_{ij}^v represents the random error term, while $j = 1, 2, \dots, m_v$ corresponds to the observation time point at interval T . Furthermore, in finite-dimensional space, $x_i^v(t)$ can be approximated as

$$x_i^v(t) \approx \sum_{l=1}^r \alpha_{il}^v \varphi_l^v(t) = \boldsymbol{\varphi}_i^v(t)^T \boldsymbol{\alpha}_i^v \quad (2)$$

Here, $\boldsymbol{\varphi}_i^v(t) = (\varphi_{i1}^v(t), \varphi_{i2}^v(t), \dots, \varphi_{ir}^v(t))^T$ denotes a set of basis vectors in a given space, while $\boldsymbol{\alpha}_i^v = (\alpha_{i1}^v, \alpha_{i2}^v, \dots, \alpha_{ir}^v)^T$ denotes the vector of coefficients requiring estimation. Expressing Equations (1) and (2), respectively, in matrix form gives

$$\tilde{\mathbf{X}}_v^+ \approx \boldsymbol{\Phi}_v \mathbf{A}_v \quad (3)$$

$$\tilde{\mathbf{Y}}_v^+ = \Phi_v \mathbf{A}_v + \mathbf{E}_v \quad (4)$$

where $\Phi_v = (\phi_1^v(t), \phi_2^v(t), \dots, \phi_{m_v}^v(t))^T$, $\mathbf{A}_v = (a_1^v, a_2^v, \dots, a_n^v)^T$, $\mathbf{E}_v = (\epsilon_1^v, \epsilon_2^v, \dots, \epsilon_n^v)$, $\epsilon_i^v = (\epsilon_{i1}^v, \epsilon_{i2}^v, \dots, \epsilon_{im_v}^v)$.

The core objective function for multi-view functional matrix completion [30] may be expressed as follows, where $\mathbf{Y} = \{\tilde{\mathbf{Y}}_1^+, \dots, \tilde{\mathbf{Y}}_{m_v}^+\}$ denotes the given multi-view data matrix:

$$\min_{U_v, V_v \geq 0} \left\{ \sum_{v=1}^{n_v} \theta_v \left\| \mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T) \right\|_F^2 \right\} \quad (5)$$

Here, θ_v denotes the weighting parameter for different viewpoints, $\odot$ represents the Hadamard matrix product, $\Phi \in \mathbb{R}^{m \times r}$ and $\mathbf{O} \in \mathbb{R}^{m \times n}$ are projection matrices of the same type as $\tilde{\mathbf{Y}}_v^+$. That is, if the entry in $\tilde{\mathbf{Y}}_v^+$ is observable, then $O_{ij} = 1$; otherwise $O_{ij} = 0$. $\mathbf{U}_v \in \mathbb{R}^{r \times d}$ and $\mathbf{V}_v \in \mathbb{R}^{n \times d}$ denote the basis matrix and coefficient matrix, respectively, for the v -th viewpoint.

3.2. HSIC-DIMFMC Method

A key advantage of the HSIC criterion lies in its ability to effectively capture both linear and nonlinear statistical dependencies among variables. From a computational perspective, HSIC can be formulated as a matrix trace operation, which enables its seamless incorporation as a regularization term into optimization frameworks and thereby facilitates the solution process [31].

Given two separable regenerative Hilbert spaces $\mathcal{F}$ and $\mathcal{G}$, for random variables x and y , there exist mapping functions $\phi: x \rightarrow \mathcal{F}$ and $\varphi: y \rightarrow \mathcal{G}$. For point $x_i, x_j \in \phi$, the inner product is defined as $k_1(x_i, x_j)_F = \langle \phi(x_i), \phi(x_j) \rangle$; for point $y_i, y_j \in \varphi$, the inner product is $k_2(y_i, y_j)_F = \langle \varphi(y_i), \varphi(y_j) \rangle$. Where k_1 and k_2 are the kernels of random variables x and y , respectively, the covariance between the two random variables can be expressed as

$$C_{xy} = E_{xy}[(\phi(x) - u_x) \otimes (\varphi(y) - u_y)] \quad (6)$$

Here, $u_x = E(\phi(x))$ and $u_y = E(\varphi(y))$ represent the expectations of x and y , respectively, while $\otimes$ denotes the tensor product. HSIC is defined as

$$\text{HSIC}(p_{xy}, \phi, \varphi) = \|C_{xy}\|_{HS}^2 \quad (7)$$

Given that p_{xy} denotes the joint measure of x and y , and HSIC represents the square of the Hilbert norm on the mutual covariance matrix C_{xy} , then

$$\text{HSIC}(p_{xy}, \phi, \varphi) = \|C_{xy}\|_{HS}^2 = \|E((\phi(x) - \mu_x) \otimes (\varphi(y) - \mu_y))\|_{HS}^2 \quad (8)$$

In multi-view data, the empirical HSIC between $\mathbf{V}_v$ and $\mathbf{V}_s$ is defined as

$$\text{HSIC}(\mathbf{V}_v, \mathbf{V}_s) = (N-1)^{-2} \text{tr}(\mathbf{K}_{V_v} \mathbf{H} \mathbf{K}_{V_s} \mathbf{H}) \quad (9)$$

Here, $\mathbf{K}_{V_v}$ and $\mathbf{K}_{V_s}$ represent the nuclear matrices of $\mathbf{V}_v$ and $\mathbf{V}_s$, respectively, while $\mathbf{H} = \mathbf{I} - (1/N)\mathbf{N}$ and $\mathbf{I}$ denote the identity matrix, with all elements of $\mathbf{N} \in \mathbb{R}^{N \times N}$ being 1.

Traditional single-graph regularization considers only the local manifold information of samples, preserving similarity among neighbors. It lacks discriminative power and tends to incorporate outliers into the graph regularization. To integrate local and discriminative manifold information, a novel regularization term was proposed by leverag-

ing multiple scales of the KNN graph [35]. Each sample's edge is defined as $e_i = \text{knn}(k, x_i) \cup x_i$. Unlike conventional approaches, the weight distribution within this regularization term is based on the order of the sample's nearest neighbors.

Define the local manifold edge weight function as

$$Hf_i(Y_{ij}) = \begin{cases} 1, & 0 \leq Y_{ij} < (2/5)k_1 \\ (1/7)(2^{(5/k_1)(-Y_{ij}+k_1)} - 1), & (2/5)k_1 \leq Y_{ij} < k_1 \end{cases} \quad (10)$$

The design objective of the edge weight function for the local manifold graph is to assign higher weights to closer neighbors while ensuring a smooth transition of these weights. Specifically, for a given sample i , the K-Nearest Neighbors (KNN) algorithm is first employed to identify its k_1 nearest neighbors. These neighbors are then sorted in ascending order based on their distances, yielding a proximity rank index $Y_{ij} \in [0, k_1)$. Here, $Y_{ij} = 0$, represents the nearest neighbor, and a larger value of Y_{ij} indicates a greater distance between sample j and sample i . A piecewise local weighting function $Hf_i(\cdot)$ is then applied, with the purpose of assigning a weight of 1 to the nearest neighbor $Y_{ij} \leq (2/5)k_1$ and exponentially decaying weights to more distant samples $Y_{ij} > (2/5)k_1$, thus constraining significant correlations to within the local neighborhood.

Define the edge weight function of the discriminant manifold as follows:

$$Hg_i(Y_{ij}) = \begin{cases} (1/3)(2^{(k_2/3)(Y_{ij}-N+k_2)} - 1), & N - k_2 \leq Y_{ij} < N - (1/3)k_2 \\ 1, & N - (1/3)k_2 \leq Y_{ij} < N \end{cases} \quad (11)$$

The design objective of the edge weight function for the discriminative manifold graph is complementary to that of the local manifold: it aims to assign larger discriminative weights to distant samples, thereby maximizing the distinction between samples with dissimilar features. For a given sample i , all N samples are sorted in ascending order based on their distances to i , resulting in a rank index $Y_{ij} \in [0, N)$. A larger Y_{ij} implies that sample j is further away from sample i . Let k_2 be the hyperparameter controlling the range of distant neighbors considered in the discriminative manifold. The function $Hg_i(\cdot)$ is designed to allocate an exponentially increasing penalty weight to further points, encouraging the effective separation of inliers and outliers. Specifically, when $N - (1/3)k_2 \leq Y_{ij} < N$, the weight becomes a constant 1, ensuring adequate discrimination among distant samples. It should be noted that the aforementioned distances are computed based on the differences in the meteorological indicator values between samples.

After defining these edge weight functions, two different manifold regularization terms are established:

$$\begin{aligned} \text{tr}(\mathbf{V}^T \mathbf{L}_{H+} \mathbf{V}) &= \frac{1}{2} \sum_{i,j=1}^N w_{+ij} \|v_i - v_j\|^2 \\ \text{tr}(\mathbf{V}^T \mathbf{L}_{H-} \mathbf{V}) &= \frac{1}{2} \sum_{i,j=1}^N w_{-ij} \|v_i - v_j\|^2 \end{aligned} \quad (12)$$

Here, $w_{+ij} = Hf_i(Y_{ij}) + Hf_j(Y_{ji})$, $w_{-ij} = 2 - Hg_i(Y_{ij}) - Hg_j(Y_{ji})$, and $\mathbf{L}_{H+}$ denote the Laplacian matrices of the local manifold graphs, while $\mathbf{L}_{H-}$ represents the Laplacian matrix of the discriminative manifold graph. $\text{tr}(\mathbf{V}^T \mathbf{L}_{H+} \mathbf{V})$ promotes clustering of neighboring samples, whilst $\text{tr}(\mathbf{V}^T \mathbf{L}_{H-} \mathbf{V})$ enhances discriminative power between different categories or manifolds. This is achieved by distancing representations in the feature space that were originally dissimilar in the original space, thereby better preserving the global dis-

criminative structure of the data. By combining local and discriminative structures, this framework captures relationships more comprehensively compared to single-scale KNN-based approaches. Integrating both regularization terms yields the final manifold regularization expression

$$\text{tr}(\mathbf{V}^T \mathbf{L}_H \mathbf{V}) = \gamma \text{tr}(\mathbf{V}^T \mathbf{L}_{H+} \mathbf{V}) + (1 - \gamma) \text{tr}(\mathbf{V}^T \mathbf{L}_{H-} \mathbf{V}) \quad (13)$$

By integrating the HSIC criterion with dual-information graph regularization, a novel method termed HSIC-DIMFMC is proposed for multi-view functional matrix completion:

$$\begin{aligned} \min \{ & \sum_{v=1}^{n_v} \theta_v \left\| \mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T) \right\|_F^2 + \beta \sum_{v \neq s} \theta_v \text{HSIC}(\mathbf{V}_v, \mathbf{V}_s) \\ & + \sum_{v=1}^{n_v} \gamma \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H+} \mathbf{V}_v) + \sum_{v=1}^{n_v} (1 - \gamma) \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H-} \mathbf{V}_v) \} \\ \text{s.t. } & \mathbf{U}_v \geq 0, \mathbf{V}_v \geq 0 \end{aligned} \quad (14)$$

Here, $v = 1, 2, \dots, n_v$, γ , and β are parameters that modulate the weighting relationships among the components. The first term formulates the core optimization objective for completing multi-view functional matrices; the second term utilizes HSIC to extract complementary information embedded across different meteorological indicators; the third and fourth terms constitute dual-information graph regularization. The third term implements local manifold regularization, promoting clustering of neighboring samples, while the fourth term employs discriminative manifold regularization, enhancing the discriminative power between different categories or manifolds.

This paper selects $\mathbf{K}_{V_v} = \mathbf{V}_v \mathbf{V}_v^T$ as the kernel function for HSIC. To simplify computations, the constant term is omitted without altering the essence of the constraint, thereby deriving an equivalent form of the diversity constraint:

$$\begin{aligned} \text{HSIC}(\mathbf{V}_v, \mathbf{V}_s) &= \sum_{1 \leq s < n_v, v \neq s} \text{tr}(\mathbf{K}_{V_v} \mathbf{H} \mathbf{K}_{V_s} \mathbf{H}) \\ &= \sum_{1 \leq s < n_v, v \neq s} \text{tr}(\mathbf{V}_v^T \mathbf{H} \mathbf{K}_{V_s} \mathbf{H} \mathbf{V}_v) \\ &= \sum_{1 \leq s < n_v, v \neq s} \text{tr}(\mathbf{V}_v^T \mathbf{K}_{V_s} \mathbf{V}_v) \end{aligned} \quad (15)$$

where $\mathbf{K} = \sum_{1 \leq s < n_v, v \neq s} \mathbf{H} \mathbf{K}_{V_s} \mathbf{H}$, the objective function represented by Equation (14), can be expressed as

$$\begin{aligned} \min \{ & \sum_{v=1}^{n_v} \theta_v \left\| \mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T) \right\|_F^2 + \beta \sum_{v=1}^{n_v} \theta_v \text{tr}(\mathbf{V}_v^T \mathbf{K} \mathbf{V}_v) \\ & + \sum_{v=1}^{n_v} \gamma \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H+} \mathbf{V}_v) + \sum_{v=1}^{n_v} (1 - \gamma) \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H-} \mathbf{V}_v) \} \\ \text{s.t. } & \mathbf{U}_v \geq 0, \mathbf{V}_v \geq 0 \end{aligned} \quad (16)$$

3.3. Optimization Solution

The objective function in Equation (16) exhibits non-convexity with respect to both $\mathbf{U}$ and $\mathbf{V}$, which poses challenges in finding the globally optimal solution. Therefore, the alternating iterative update technique is employed to optimize and solve the objective function in Equation (16), updating one variable at a time while fixing the others. The augmented Lagrangian function for Equation (16) is formulated as follows:

$$\begin{aligned} \mathcal{L} = & \sum_{v=1}^{n_v} \theta_v \left\| \mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T) \right\|_F^2 + \beta \sum_{v=1}^{n_v} \theta_v \text{tr}(\mathbf{V}_v^T \mathbf{K} \mathbf{V}_v) \\ & + \sum_{v=1}^{n_v} \gamma \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H+} \mathbf{V}_v) + \sum_{v=1}^{n_v} (1-\gamma) \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H-} \mathbf{V}_v) \\ & - \text{tr}(\Psi_v^T \mathbf{U}_v) - \text{tr}(\Gamma_v^T \mathbf{V}_v) \end{aligned} \quad (17)$$

(1) Fix $\mathbf{V}_v$ and update $\mathbf{U}_v$. The part of Equation (17) with respect to $\mathbf{U}_v$ is simplified as follows:

$$\mathcal{L}_1 = \theta_v \text{tr}[\mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T)] [\mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T)]^T - \text{tr}(\Psi_v^T \mathbf{U}_v) \quad (18)$$

Differentiating Equation (18) with respect to $\mathbf{U}_v$ and setting the derivative equal to zero gives

$$-2\theta_v \Phi^T (\mathbf{O} \odot \tilde{\mathbf{Y}}_v^+) \mathbf{V}_v + 2\theta_v \Phi^T (\mathbf{O} \odot \Phi \mathbf{U}_v \mathbf{V}_v^T) \mathbf{V}_v = \Psi_v \quad (19)$$

According to the KKT conditions, the solution must satisfy $\Psi_v \odot \mathbf{U}_v = 0$, leading to

$$[-2\theta_v \Phi^T (\mathbf{O} \odot \tilde{\mathbf{Y}}_v^+) \mathbf{V}_v + 2\theta_v \Phi^T (\mathbf{O} \odot \Phi \mathbf{U}_v \mathbf{V}_v^T) \mathbf{V}_v]_{ij} \odot \mathbf{U}_{vij} = 0 \quad (20)$$

Consequently, the updating rule for $\mathbf{U}_v$ can be derived as

$$\mathbf{U}_{vij} \leftarrow \mathbf{U}_{vij} \sqrt{\frac{(\Phi^T (\mathbf{O} \odot \tilde{\mathbf{Y}}_v^+) \mathbf{V}_v)_{ij}}{(\Phi^T (\mathbf{O} \odot \Phi \mathbf{U}_v \mathbf{V}_v^T) \mathbf{V}_v)_{ij}}} \quad (21)$$

(2) With $\mathbf{U}_v$ fixed, we update $\mathbf{V}_v$. The portion of Equation (17) pertaining to $\mathbf{V}_v$ can be simplified to

$$\begin{aligned} \mathcal{L}_2 = & \theta_v \text{tr}[\mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T)] [\mathbf{O} \odot (\tilde{\mathbf{Y}}_v^+ - \Phi \mathbf{U}_v \mathbf{V}_v^T)]^T \\ & + \beta \theta_v \text{tr}(\mathbf{V}_v^T \mathbf{K} \mathbf{V}_v) + \gamma \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H+} \mathbf{V}_v) \\ & + (1-\gamma) \text{tr}(\mathbf{V}_v^T \mathbf{L}_{H-} \mathbf{V}_v) - \text{tr}(\Gamma_v^T \mathbf{V}_v) \end{aligned} \quad (22)$$

By differentiating Equation (22) with respect to $\mathbf{V}_v$, setting the derivative to zero, and incorporating the KKT conditions for $\Gamma_v \odot \mathbf{V}_v = 0$, we obtain

$$\begin{aligned} & [-2\theta_v (\mathbf{O}^T \odot \tilde{\mathbf{Y}}_v^{+T}) \Phi \mathbf{U}_v + 2\theta_v (\mathbf{O}^T \odot \mathbf{V}_v \mathbf{U}_v^T \Phi^T) \Phi \mathbf{U}_v \\ & + [2\beta \theta_v \mathbf{K} \mathbf{V}_v + 2\gamma (\mathbf{D}_{H+} - \mathbf{W}_{H+}) \mathbf{V}_v \\ & + 2(1-\gamma) (\mathbf{D}_{H-} - \mathbf{W}_{H-}) \mathbf{V}_v]_{ij} \odot \mathbf{V}_{vij} \\ & = 0 \end{aligned} \quad (23)$$

The parameter $\mathbf{V}_v$ is updated using the following expression:

$$\mathbf{V}_{vij} \leftarrow \mathbf{V}_{vij} \sqrt{\frac{(\theta_v (\mathbf{O}^T \odot \tilde{\mathbf{Y}}_v^{+T}) \Phi \mathbf{U}_v + \gamma \mathbf{W}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{W}_{H-} \mathbf{V}_v)_{ij}}{(\theta_v (\mathbf{O}^T \odot \mathbf{V}_v \mathbf{U}_v^T \Phi^T) \Phi \mathbf{U}_v + \beta \theta_v \mathbf{K} \mathbf{V}_v + \gamma \mathbf{D}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{D}_{H-} \mathbf{V}_v)_{ij}}} \quad (24)$$

In summary, the detailed implementation procedure of the HSIC-DIMFMC method is outlined as follows: The pseudo-code of the proposed algorithm is presented in Algorithm 1.

Algorithm 1 HSIC-DIMFMC

Input: Given multi-view data matrix $\mathbf{Y}$, parameter $\theta_v, \beta, \gamma, \varepsilon_0 = 10^{-7}$, and iteration count t

Output: $\{\mathbf{U}_1^{t+1}, \mathbf{U}_2^{t+1}, \dots, \mathbf{U}_{n_v}^{t+1}\}$ and $\{\mathbf{V}_1^{t+1}, \mathbf{V}_2^{t+1}, \dots, \mathbf{V}_{n_v}^{t+1}\}$.

1. Randomly generate $\mathbf{U}_v^0$ and $\mathbf{V}_v^0$;

-
2. Keep V_v fixed and update U_v^{t+1} according to Equation (21);
 3. Keep U_v fixed and update V_v^{t+1} according to Equation (24);
 4. Repeat steps 2 and 3 until the convergence condition $\|U_v^{t+1}V_v^{t+1T} - U_v^tV_v^{tT}\|^2 / \|U_v^tV_v^{tT}\|^2 \leq \varepsilon_0$ is met.
-

3.4. Convergence and Time Complexity Analysis

A convergence analysis of the update rules for U_v and V_v (Equations (21) and (24)) is presented below.

Theorem 1. According to the update rules in Equations (21) and (24), Equation (16) defines an objective function characterized by a monotonic decrease.

Proof. By constructing auxiliary functions, the convergence of update Equations (21) and (24) can be verified separately. At this point, the terms in Equation (16) unrelated to U_v are eliminated, retaining only the portions associated with U_v .

$$\mathcal{L}(U_v) = \text{tr}(U_v^T \Phi^T (\mathcal{O} \odot \Phi U_v V_v^T) V_v - 2U_v^T \Phi^T (\mathcal{O} \odot \tilde{Y}_v^+) V_v) \quad (25)$$

The auxiliary function, denoted as $G(U_v, U_v^t)$, is defined for with $\mathcal{L}(U_v)$ as follows:

$$G(U_v, U_v^t) = \sum_{i,j} \frac{(\Phi^T (\mathcal{O} \odot \Phi U_v^t V_v^T) V_v)_{ij} U_{vij}^2}{U_{vij}^t} - 2 \sum_{i,j} (\Phi^T (\mathcal{O} \odot \tilde{Y}_v^+) V_v)_{ij} U_{vij}^t (1 + \log(U_{vij} / U_{vij}^t)) \quad (26)$$

If with $G(U_v, U_v) = \mathcal{L}(U_v)$, $G(U_v, U_v^t) \geq \mathcal{L}(U_v)$ holds, then by selecting U_v^{t+1} , we can ensure that

$$U_v^{t+1} = \arg \min_t G(U_v, U_v^t) \quad (27)$$

Establishing that $\mathcal{L}(U_v)$ is clearly monotonically decreasing, we can thus obtain

$$G(U_v^{t+1}, U_v^{t+1}) \leq G(U_v^{t+1}, U_v^t) \leq G(U_v^t, U_v^t) \quad (28)$$

To minimize the auxiliary function (26) under the condition (27), solve for U_v^{t+1} . The update rule for U_{vij} is derived by first taking the partial derivative of Equation (26) with respect to U_{vij} and setting it to zero. Subsequently, substituting U_{vij} with U_{vij}^{t+1} yields

$$\frac{(\Phi^T (\mathcal{O} \odot \Phi U_v^t V_v^T) V_v)_{ij} U_{vij}^{t+1}}{U_{vij}^t} = (\Phi^T (\mathcal{O} \odot \tilde{Y}_v^+) V_v)_{ij} \frac{U_{vij}^t}{U_{vij}^{t+1}} \quad (29)$$

Thus, the following is obtained:

$$U_{vij}^{t+1} = U_{vij}^t \sqrt{\frac{(\Phi^T (\mathcal{O} \odot \tilde{Y}_v^+) V_v)_{ij}}{(\Phi^T (\mathcal{O} \odot \Phi U_v^t V_v^T) V_v)_{ij}}} \quad (30)$$

Equation (30) is the update rule Equation (21) corresponding to U_v .

Next, to prove the convergence of the update rule for V_v in Equation (24), we simplify Equation (17) by retaining only the terms involving V_v .

$$\begin{aligned} \mathcal{L}(V_v) = & \text{tr}(2\theta_v V_v^T (\mathcal{O}^T \odot \tilde{Y}_v^{+T}) \Phi U_v - 2\theta_v (\mathcal{O}^T \odot V_v U_v^T \Phi^T) \Phi U_v) \\ & + \beta \theta_v \text{tr}(V_v^T K V_v) + \gamma \text{tr}(V_v^T L_{H+} V_v) + (1 - \gamma) \text{tr}(V_v^T L_{H-} V_v) \end{aligned} \quad (31)$$

Construct auxiliary function $G(V_v, V_v^t)$ of $\mathcal{L}(V_v)$, defined as

$$G(\mathbf{V}_v, \mathbf{V}_v^t) = \sum_{i,j} \frac{(\theta_v(\mathbf{O}^T \odot \mathbf{V}_v \mathbf{U}_v^T \Phi^T) \Phi \mathbf{U}_v + \beta \theta_v \mathbf{K} \mathbf{V}_v + \gamma \mathbf{D}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{D}_{H-} \mathbf{V}_v)_{ij} \mathbf{V}_{vij}^2}{\mathbf{V}_{vij}^t} - 2\theta_v \sum_{i,j} ((\mathbf{O}^T \odot \tilde{\mathbf{Y}}_v^{+T}) \Phi \mathbf{U}_v + \gamma \mathbf{W}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{W}_{H-} \mathbf{V}_v)_{ij} \mathbf{V}_{vij}^t (1 + \log(\mathbf{V}_{vij} / \mathbf{V}_{vij}^t)) \quad (32)$$

If condition AA $G(\mathbf{V}_v, \mathbf{V}_v) = \mathcal{L}(\mathbf{V}_v)$ and $G(\mathbf{V}_v, \mathbf{V}_v^t) \geq \mathcal{L}(\mathbf{V}_v)$ holds, then $\mathbf{V}_v^{t+1}$ can be chosen such that

$$\mathbf{V}_v^{t+1} = \arg \min G(\mathbf{V}_v, \mathbf{V}_v^t) \quad (33)$$

Therefore, $\mathcal{L}(\mathbf{V}_v)$ exhibits a monotonic decrease, with

$$G(\mathbf{V}_v^{t+1}, \mathbf{V}_v^{t+1}) \leq G(\mathbf{V}_v^{t+1}, \mathbf{V}_v^t) \leq G(\mathbf{V}_v^t, \mathbf{V}_v^t) \quad (34)$$

The solution for $\mathbf{V}_v^{t+1}$ is obtained by minimizing the auxiliary function (32) under the constraint of Equation (33). This is followed by setting the partial derivative of (32) with respect to $\mathbf{V}_{vij}$ to zero, resulting in

$$\begin{aligned} & (\theta_v(\mathbf{O}^T \odot \tilde{\mathbf{Y}}_v^{+T}) \Phi \mathbf{U}_v + \gamma \mathbf{W}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{W}_{H-} \mathbf{V}_v)_{ij} \frac{\mathbf{V}_{vij}^t}{\mathbf{V}_{vij}^{t+1}} \\ &= \frac{(\theta_v(\mathbf{O}^T \odot \mathbf{V}_v \mathbf{U}_v^T \Phi^T) \Phi \mathbf{U}_v + \beta \theta_v \mathbf{K} \mathbf{V}_v + \gamma \mathbf{D}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{D}_{H-} \mathbf{V}_v)_{ij}}{\mathbf{V}_{vij}^t} \mathbf{V}_{vij}^{t+1} \end{aligned} \quad (35)$$

Thus, there is

$$\mathbf{V}_{vij}^{t+1} \leftarrow \mathbf{V}_{vij}^t \sqrt{\frac{(\theta_v(\mathbf{O}^T \odot \tilde{\mathbf{Y}}_v^{+T}) \Phi \mathbf{U}_v + \gamma \mathbf{W}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{W}_{H-} \mathbf{V}_v)_{ij}}{(\theta_v(\mathbf{O}^T \odot \mathbf{V}_v \mathbf{U}_v^T \Phi^T) \Phi \mathbf{U}_v + \beta \theta_v \mathbf{K} \mathbf{V}_v + \gamma \mathbf{D}_{H+} \mathbf{V}_v + (1-\gamma) \mathbf{D}_{H-} \mathbf{V}_v)_{ij}}} \quad (36)$$

The above Equation is the update rule for $\mathbf{V}_v$ in Equation (24). $\square$

3.5. Time Complexity Analysis

The optimization model of the HSIC-DIMFMC algorithm comprises four components. Here, m denotes the number of observation time points; n denotes the number of meteorological stations; and d represents the rank, i.e., the dimension of latent variables used for low-dimensional representation and imputation. The algorithm has a time complexity of $O(mnd + mn)$ for matrix multiplication and Frobenius norm calculation and a time complexity of $O(nd^2 + m^2)$ for trace operation and kernel matrix $\mathbf{K}$ computation. Similarly, it has a time complexity of $O(nd^2 + n^2)$ for local manifold regularization and a time complexity of $O(nd^2 + n^2)$ for discriminative manifold regularization. The primary computational overhead stems from the calculation of the kernel matrix $\mathbf{K}$ and the graph Laplacian matrices $\mathbf{L}_{H+}$ and $\mathbf{L}_{H-}$. Therefore, when iterating t times across n_v viewpoints, the time complexity is $O(t(mnd + mn + m^2 + n^2 + 2n_v nd^2))$.

4. Experiments and Analysis

4.1. Experiments Settings

To verify the effectiveness of the HSIC-DIMFMC imputation method, its performance was first tested on Dataset I, followed by a practical application on Dataset II.

4.1.1. Evaluation Metrics

We evaluated the imputation performance using three metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Normalized Root Mean Square Error (NRMSE).

4.1.2. Implementation

Experiments were conducted in an R 4.3.0 environment on hardware configured with an Intel(R) Core (TM) i5-8250U CPU @ 1.60 GHz (boost up to 1.80 GHz), 8 GB of memory, and a Windows 10 64-bit operating system.

4.1.3. Compared Methods

To evaluate the performance of the proposed HSIC-DIMFMC method, seven representative missing data imputation methods were selected for comparison. These methods cover different technical paradigms, including deterministic interpolation, classical statistical methods, traditional machine learning, functional matrix completion, and multi-view functional matrix completion.

- IDW [37] was adopted as a deterministic baseline method, which estimates missing values through a weighted average of neighboring observations.
- OKFD [38] represents a classical statistical interpolation method for functional data, which performs imputation by learning the optimal kernel function from the observed functional curves.
- KNN [39] is a traditional machine learning method that imputes missing values based on the similarity among observed samples, typically measured by distance-based metrics. Specifically, it selects the K nearest neighbors based on Euclidean distance and utilizes their weighted average to fill in the missing values.
- SFI [40] is a functional matrix completion method for recovering missing functional observations.
- HFI [40] is its variant with a modified regularization strategy, specifically adopting a hard-thresholding regularization approach.
- MVNFMC [30] extends functional matrix completion to a multi-view learning framework to jointly exploit information from different sources.
- GRMFMC [31] further incorporates the HSIC criterion and graph regularization to better capture higher-order statistical dependencies and local geometric structures across multiple views.

4.1.4. Datasets and Experimental Settings

This study employs the Missing Completely at Random (MCAR) mechanism to simulate missing datasets [41]. Specifically, under a given fixed missing ratio, observations are independently, uniformly, and randomly masked as missing (i.e., Point Missing, PM mode). The occurrence of each missing position is mutually independent and equiprobable, aiming to simulate random transmission errors or transient communication noise in real-world meteorological monitoring. The MCAR mechanism effectively isolates the evaluation of imputation performance from the confounding effects of the missingness mechanism itself. By providing a controlled and interpretable environment to assess the intrinsic capabilities of imputation algorithms, it has been widely adopted in benchmarking studies. While we acknowledge that real-world meteorological data may exhibit more complex missing patterns, this paper serves as a proof-of-concept study focusing on methodology development and baseline comparison. Therefore, our experimental evaluation is first conducted under the standard MCAR baseline setting, and extending the proposed framework to more complex scenarios will be an important direction for future work.

The missing ratios were set to 10%, 20%, 30%, and 40%. The proposed HSIC-DIMFMC method was implemented in a two-view learning framework, and experiments were conducted separately on two strongly correlated meteorological variable pairs: TEMP-DEWP and WDSP-MXSPD.

4.2. Ablation Study

We analyzed the independent contributions of the HSIC criterion and the dual-information graph regularization module to the overall performance of the HSIC-DIMFMC method through ablation experiments.

(1) HSIC-DIMFMC1: The dual-information graph regularization term was removed from the full model (16) to assess its role in preserving the geometric structure of the data.

(2) HSIC-DIMFMC2: The HSIC criterion term was removed from the complete model (16) to assess its contribution to the statistical dependence between mining perspectives.

Table 3 presents the RMSE and MAE performance of each method at 10% to 40% missingness rates (bold indicates optimal results). Analysis reveals that the complete HSIC-DIMFMC method consistently achieves significantly lower RMSE and MAE values than both HSIC-DIMFMC1 and HSIC-DIMFMC2 across all settings. Specifically, its RMSE was reduced by an average of 53.09% to 74.88% and 52.23% to 74.84% compared to the latter two methods, respectively. This result confirms that both the HSIC criterion and dual-information graph regularization play crucial and irreplaceable roles in improving imputation accuracy.

Table 3. RMSE and MAE results of four meteorological indicators under different missing rates (mean of 50 results).

Missing Rate (%)	Imputation Method	Evaluation Metric	TEMP	DEWP	WDSP	MXSPD
10	HSIC-DIMFMC	RMSE	1.0996 ± 0.0020	0.6930 ± 0.0030	0.4293 ± 0.0002	0.7452 ± 0.0005
		MAE	0.5962 ± 0.0010	0.3504 ± 0.0035	0.2100 ± 0.0002	0.3946 ± 0.0003
	HSIC-DIMFMC1	RMSE	3.0931 ± 0.0701	2.3657 ± 0.0454	0.7133 ± 0.0178	1.1848 ± 0.0166
		MAE	2.5369 ± 0.0536	1.9430 ± 0.0321	0.5018 ± 0.0116	0.8965 ± 0.0118
	HSIC-DIMFMC2	RMSE	3.0767 ± 0.0613	2.3636 ± 0.0568	0.7121 ± 0.0117	1.1899 ± 0.0235
		MAE	2.5187 ± 0.0446	1.9420 ± 0.0395	0.5057 ± 0.0085	0.8971 ± 0.0135
20	HSIC-DIMFMC	RMSE	0.7896 ± 0.0033	0.4913 ± 0.0007	0.3138 ± 0.0027	0.5400 ± 0.0009
		MAE	0.3024 ± 0.0023	0.1758 ± 0.0004	0.1059 ± 0.0003	0.1996 ± 0.0002
	HSIC-DIMFMC1	RMSE	3.0769 ± 0.0636	2.3559 ± 0.0658	0.7198 ± 0.0164	1.2052 ± 0.0220
		MAE	2.5199 ± 0.0506	1.9277 ± 0.0459	0.5057 ± 0.0107	0.9094 ± 0.0127
	HSIC-DIMFMC2	RMSE	3.0840 ± 0.0655	2.3759 ± 0.0520	0.7251 ± 0.0202	1.1972 ± 0.0198
		MAE	2.5258 ± 0.0481	1.9446 ± 0.0373	0.5116 ± 0.0110	0.9006 ± 0.0103
30	HSIC-DIMFMC	RMSE	0.6492 ± 0.0037	0.4123 ± 0.0099	0.2580 ± 0.0005	0.5115 ± 0.0082
		MAE	0.2036 ± 0.0009	0.1208 ± 0.0024	0.0719 ± 0.0001	0.1382 ± 0.0002
	HSIC-DIMFMC1	RMSE	3.1042 ± 0.0767	2.4021 ± 0.0584	0.7269 ± 0.0181	1.2009 ± 0.0283
		MAE	2.5474 ± 0.0596	1.9563 ± 0.0430	0.5102 ± 0.0099	0.8982 ± 0.0156
	HSIC-DIMFMC2	RMSE	3.0850 ± 0.0601	2.3640 ± 0.0542	0.7425 ± 0.0218	1.2087 ± 0.0196
		MAE	2.5223 ± 0.0446	1.9295 ± 0.0374	0.5128 ± 0.0099	0.9071 ± 0.0121
40	HSIC-DIMFMC	RMSE	0.5739 ± 0.0019	0.3592 ± 0.0018	0.2374 ± 0.0014	0.4325 ± 0.0038
		MAE	0.1557 ± 0.0007	0.0909 ± 0.0006	0.0554 ± 0.0003	0.1044 ± 0.0004
	HSIC-DIMFMC1	RMSE	3.0814 ± 0.0652	2.4321 ± 0.0643	0.7328 ± 0.0185	1.2469 ± 0.0265
		MAE	2.5253 ± 0.0490	1.9893 ± 0.0447	0.5120 ± 0.0104	0.9320 ± 0.0138
	HSIC-DIMFMC2	RMSE	3.1059 ± 0.0622	2.3689 ± 0.0664	0.7389 ± 0.0193	1.2403 ± 0.0324
		MAE	2.5351 ± 0.0501	1.9356 ± 0.0452	0.5073 ± 0.0094	0.9252 ± 0.0129

4.3. Parameter Analysis

To validate the efficacy of the HSIC-DIMFMC method under varying levels of data missingness and analyze its sensitivity to hyperparameters γ and β , Figures 7–10 present the variations in RMSE across different parameter settings.

(1) The HSIC criterion and dual-information graph regularization collectively form the HSIC-DIMFMC method proposed in this paper. Parameters β and γ respectively reveal

the effects of the HSIC criterion and dual-information graph regularization. Set the value range for β to $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$; set the value range for γ to $\{0.2, 0.4, 0.6, 0.8, 1.0\}$.

(2) As shown in Figures 8–11, lighter bar colors indicate higher RMSE values, while darker colors correspond to lower RMSE values. It is observed that the HSIC-DIMFMC method exhibits low sensitivity to variations in parameters β and γ , suggesting a high degree of robustness and excellent generalization capability. Based on Figures 7–10, the HSIC-DIMFMC method in this work adopts the parameter values $\beta = 0.1$, $\gamma = 0.6$.

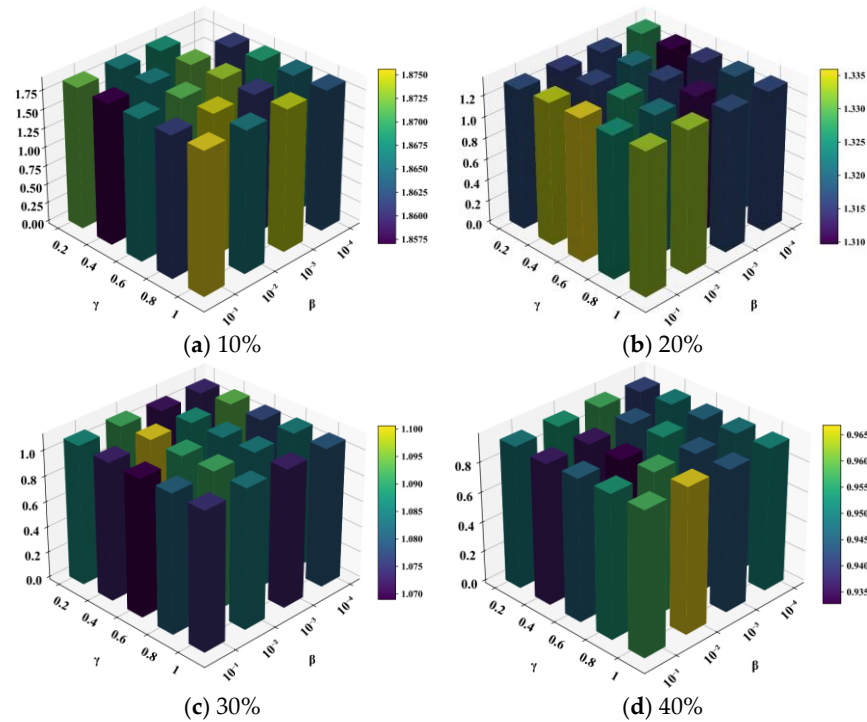


Figure 7. RMSE of γ and β in the TEMP view of the HSIC-DIMFMC method under different missing rates.

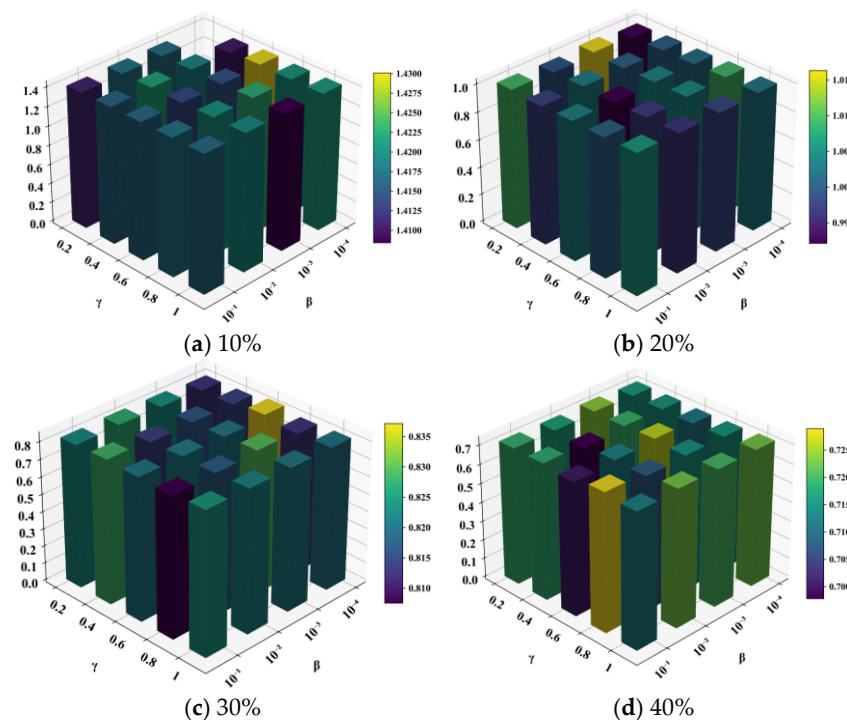


Figure 8. RMSE of γ and β in the DEWP view of the HSIC-DIMFMC method under different missing rates.

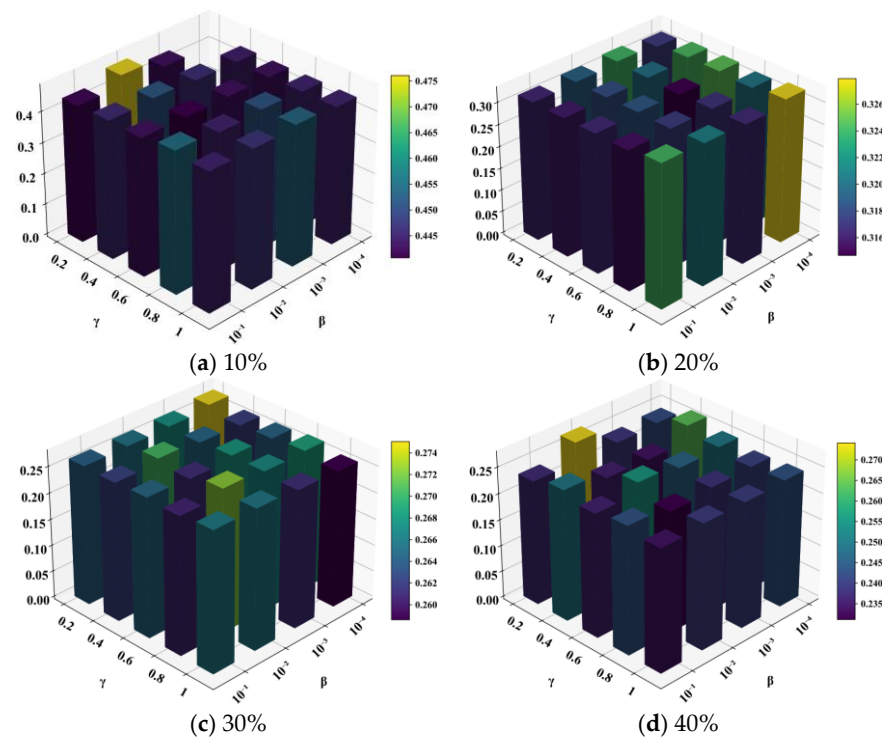


Figure 9. RMSE of γ and β in the WDSP view of the HSIC-DIMFMC method under different missing rates.

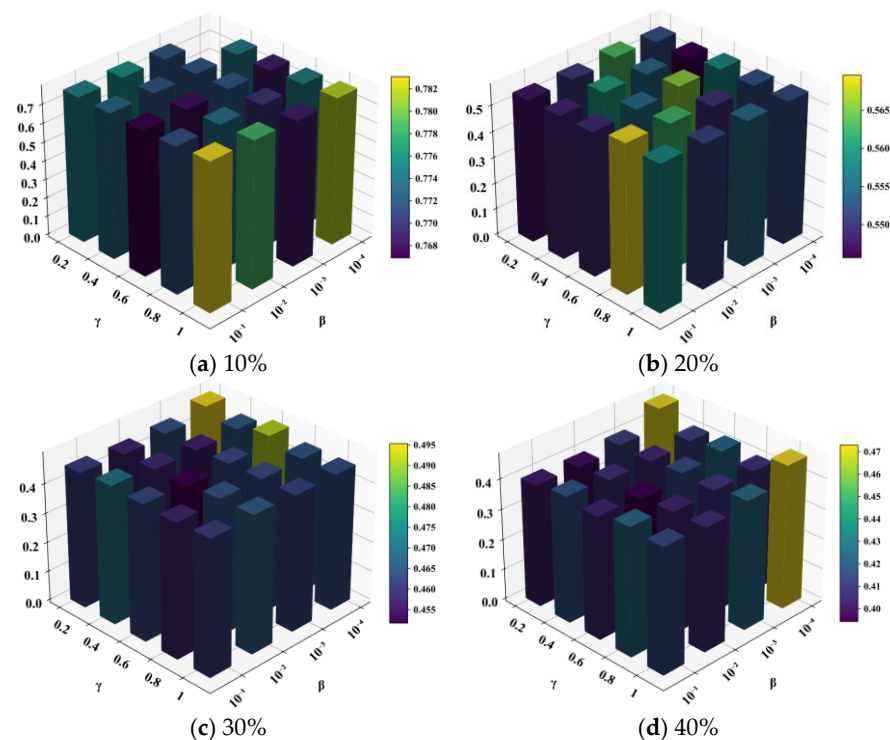


Figure 10. RMSE of γ and β in the MXSPD view of the HSIC-DIMFMC method under different missing rates.

4.4. Completion Results

Based on the aforementioned experimental design, the IDW, OKFD, KNN, HFI, SFI, MVNFMC, GRMFMC, and HSIC-DIMFMC methods were selected to perform imputation and repair on the four meteorological indicators in Dataset I. Their imputation performance was systematically compared. Tables 4–6 present the RMSE, MAE, and NRMSE results for each method under 10% to 40% missing data rates, with the optimal values highlighted in bold.

Table 4. RMSE results of 8 imputation methods under different missing rates (mean of 50 results).

Missing Rate (%)	Baseline Method	TEMP	DEWP	WDSP	MXSPD
10	IDW	1.9601 ± 0.0484	2.0821 ± 0.1107	0.7664 ± 0.0532	1.4356 ± 0.0685
	OKFD	1.7507 ± 0.1010	1.6264 ± 0.3684	1.4941 ± 0.3491	2.2944 ± 0.3969
	KNN	1.7654 ± 0.1319	1.1820 ± 0.2802	1.0806 ± 0.2530	1.5985 ± 0.2275
	HFI	2.1250 ± 0.0000	1.2549 ± 0.0000	0.7614 ± 0.0000	1.2286 ± 0.0000
	SFI	2.1154 ± 0.0005	1.2421 ± 0.0012	0.7656 ± 0.0035	1.2419 ± 0.0014
	MVNFMC	5.1528 ± 0.0227	4.1577 ± 0.0180	0.6441 ± 0.0097	1.0731 ± 0.0148
	GRMFMC	7.3083 ± 0.6989	6.6767 ± 0.4804	1.2336 ± 0.0914	1.7464 ± 0.0892
	HSIC-DIMFMC	1.0996 ± 0.0020	0.6930 ± 0.0030	0.4293 ± 0.0002	0.7452 ± 0.0005
20	IDW	1.9837 ± 0.0566	2.0921 ± 0.1027	0.8137 ± 0.0194	1.5037 ± 0.0487
	OKFD	1.8311 ± 0.1011	1.5917 ± 0.1956	1.6179 ± 0.1697	2.2419 ± 0.2126
	KNN	2.0479 ± 0.1070	1.3772 ± 0.1670	1.1185 ± 0.1243	1.7121 ± 0.1380
	HFI	2.1353 ± 0.0000	1.2578 ± 0.0000	0.7685 ± 0.0000	1.2401 ± 0.0000
	SFI	2.1313 ± 0.0021	1.2531 ± 0.0010	0.7673 ± 0.0040	1.2379 ± 0.0023
	MVNFMC	3.6565 ± 0.0182	2.9466 ± 0.0141	0.4669 ± 0.0081	0.7768 ± 0.0071
	GRMFMC	5.4348 ± 0.6415	4.3718 ± 0.3688	0.8595 ± 0.0214	1.2050 ± 0.0342
	HSIC-DIMFMC	0.7896 ± 0.0033	0.4913 ± 0.0007	0.3138 ± 0.0027	0.5400 ± 0.0009
30	IDW	2.0687 ± 0.0548	2.1742 ± 0.0554	0.8323 ± 0.0265	1.5220 ± 0.0309
	OKFD	1.8923 ± 0.0709	1.6047 ± 0.1061	1.6409 ± 0.1272	2.3111 ± 0.1551
	KNN	2.3491 ± 0.1403	1.5835 ± 0.1464	1.1576 ± 0.1117	1.7448 ± 0.1091
	HFI	2.1699 ± 0.0000	1.2972 ± 0.0000	0.7678 ± 0.0000	1.2438 ± 0.0000
	SFI	2.1579 ± 0.0079	1.2546 ± 0.0017	0.7758 ± 0.0023	1.2476 ± 0.0031
	MVNFMC	2.9863 ± 0.0140	2.3942 ± 0.0106	0.3936 ± 0.0048	0.6256 ± 0.0087
	GRMFMC	4.7282 ± 0.5382	3.5093 ± 0.1470	0.7447 ± 0.0490	1.0590 ± 0.1219
	HSIC-DIMFMC	0.6492 ± 0.0037	0.4123 ± 0.0099	0.2580 ± 0.0005	0.5115 ± 0.0082
40	IDW	2.1532 ± 0.0604	2.2619 ± 0.0772	0.8518 ± 0.0196	1.5303 ± 0.0235
	OKFD	2.0113 ± 0.0848	1.6489 ± 0.0743	1.6217 ± 0.0695	2.3202 ± 0.1578
	KNN	2.6899 ± 0.1330	1.7859 ± 0.1143	1.1939 ± 0.0881	1.7807 ± 0.1049
	HFI	2.2166 ± 0.0000	1.2826 ± 0.0000	0.7751 ± 0.0000	1.2513 ± 0.0000
	SFI	2.1755 ± 0.0109	1.2805 ± 0.0045	0.7770 ± 0.0047	1.2636 ± 0.0034
	MVNFMC	2.5772 ± 0.0164	2.0807 ± 0.0124	0.3216 ± 0.0041	0.5400 ± 0.0073
	GRMFMC	4.0230 ± 0.3754	3.2953 ± 0.3107	0.6556 ± 0.0393	0.9231 ± 0.0468
	HSIC-DIMFMC	0.5739 ± 0.0019	0.3592 ± 0.0018	0.2374 ± 0.0014	0.4325 ± 0.0038

Table 5. MAE results of 8 imputation methods under different missing rates (mean of 50 results).

Missing Rate (%)	Baseline Method	TEMP	DEWP	WDSP	MXSPD
10	IDW	1.5486 ± 0.0624	1.5743 ± 0.0687	0.5935 ± 0.0356	1.1271 ± 0.0385
	OKFD	1.3499 ± 0.0781	1.1001 ± 0.1228	0.6922 ± 0.0947	1.2275 ± 0.1434
	KNN	1.3097 ± 0.0906	0.8283 ± 0.0644	0.5931 ± 0.0742	1.0210 ± 0.0940
	HFI	1.6984 ± 0.0000	0.9085 ± 0.0000	0.4859 ± 0.0000	0.8739 ± 0.0000
	SFI	1.6964 ± 0.0019	0.9061 ± 0.0029	0.4882 ± 0.0038	0.8743 ± 0.0018
	MVNFMC	3.4173 ± 0.0081	2.7921 ± 0.0071	0.3340 ± 0.0040	0.5870 ± 0.0054

	GRMFMC	4.3027 ± 0.4886	4.0027 ± 0.3029	0.6538 ± 0.0463	0.9373 ± 0.0430
	HSIC-DIMFMC	0.5962 ± 0.0010	0.3504 ± 0.0035	0.2100 ± 0.0002	0.3946 ± 0.0003
20	IDW	1.5675 ± 0.0487	1.5783 ± 0.0821	0.6231 ± 0.0169	1.1628 ± 0.0362
	OKFD	1.4159 ± 0.0789	1.1234 ± 0.0630	0.7340 ± 0.0629	1.2071 ± 0.0750
	KNN	1.5228 ± 0.0811	0.9756 ± 0.0529	0.6246 ± 0.0401	1.0901 ± 0.0648
	HFI	1.7104 ± 0.0000	0.9062 ± 0.0000	0.4868 ± 0.0000	0.8826 ± 0.0000
	SFI	1.7138 ± 0.0012	0.9109 ± 0.0022	0.4946 ± 0.0043	0.8850 ± 0.0022
	MVNFMC	1.7135 ± 0.0059	1.3984 ± 0.0046	0.1675 ± 0.0024	0.2995 ± 0.0019
	GRMFMC	2.2686 ± 0.2733	1.8309 ± 0.1792	0.3170 ± 0.0071	0.4501 ± 0.0207
	HSIC-DIMFMC	0.3024 ± 0.0023	0.1758 ± 0.0004	0.1059 ± 0.0003	0.1996 ± 0.0002
30	IDW	1.6290 ± 0.0401	1.6427 ± 0.0576	0.6404 ± 0.0167	1.1715 ± 0.0290
	OKFD	1.4607 ± 0.0628	1.1536 ± 0.0438	0.7519 ± 0.0428	1.2523 ± 0.0531
	KNN	1.7436 ± 0.0940	1.1047 ± 0.0533	0.6585 ± 0.0398	1.1162 ± 0.0503
	HFI	1.7243 ± 0.0000	0.9258 ± 0.0000	0.4904 ± 0.0000	0.8817 ± 0.0000
	SFI	1.7133 ± 0.0020	0.9159 ± 0.0029	0.4934 ± 0.0038	0.8916 ± 0.0046
	MVNFMC	1.1417 ± 0.0048	0.9274 ± 0.0032	0.1140 ± 0.0014	0.1957 ± 0.0020
	GRMFMC	1.6270 ± 0.1984	1.1875 ± 0.0471	0.2286 ± 0.0188	0.3249 ± 0.0408
	HSIC-DIMFMC	0.2036 ± 0.0009	0.1208 ± 0.0024	0.0719 ± 0.0001	0.1382 ± 0.0002
40	IDW	1.6891 ± 0.0458	1.7101 ± 0.0618	0.6521 ± 0.0085	1.1801 ± 0.0181
	OKFD	1.5450 ± 0.0595	1.1897 ± 0.0294	0.7527 ± 0.0161	1.2720 ± 0.0515
	KNN	1.9916 ± 0.091	1.2444 ± 0.0548	0.6814 ± 0.031	1.1471 ± 0.05
	HFI	1.7519 ± 0.0000	0.9311 ± 0.0000	0.5013 ± 0.0000	0.8923 ± 0.0000
	SFI	1.7404 ± 0.0040	0.9269 ± 0.0023	0.5005 ± 0.0061	0.9020 ± 0.0026
	MVNFMC	0.8509 ± 0.0049	0.6970 ± 0.0039	0.0818 ± 0.0009	0.1464 ± 0.0014
	GRMFMC	1.2000 ± 0.1186	0.9875 ± 0.0929	0.1726 ± 0.0080	0.2464 ± 0.0122
	HSIC-DIMFMC	0.1557 ± 0.0007	0.0909 ± 0.0006	0.0554 ± 0.0003	0.1044 ± 0.0004

Table 6. NRMSE results of 8 imputation methods under different missing rates (mean of 50 results).

Missing Rate (%)	Baseline Method	TEMP	DEWP	WDSP	MXSPD
10	IDW	0.1087 ± 0.0086	0.0903 ± 0.0115	0.1359 ± 0.0162	0.1554 ± 0.0234
	OKFD	0.0927 ± 0.0061	0.1014 ± 0.0135	0.1289 ± 0.0192	0.1574 ± 0.0175
	KNN	0.4258 ± 0.0318	0.4327 ± 0.1026	0.6989 ± 0.1637	0.7261 ± 0.1034
	HFI	0.0924 ± 0.0000	0.0475 ± 0.0000	0.0512 ± 0.0000	0.0682 ± 0.0000
	SFI	0.0920 ± 0.0000	0.0470 ± 0.0000	0.0515 ± 0.0002	0.0690 ± 0.0001
	MVNFMC	0.2240 ± 0.0010	0.1572 ± 0.0007	0.0433 ± 0.0007	0.0596 ± 0.0008
	GRMFMC	0.3178 ± 0.0304	0.2525 ± 0.0182	0.0830 ± 0.0061	0.0970 ± 0.0050
	HSIC-DIMFMC	0.0478 ± 0.0001	0.0262 ± 0.0001	0.0289 ± 0.0000	0.0414 ± 0.0000
20	IDW	0.1046 ± 0.0079	0.0806 ± 0.0072	0.1217 ± 0.0128	0.1482 ± 0.0085
	OKFD	0.0884 ± 0.0046	0.0945 ± 0.0132	0.1269 ± 0.0149	0.1466 ± 0.0174
	KNN	0.4939 ± 0.0258	0.5042 ± 0.0612	0.7234 ± 0.0804	0.7778 ± 0.0627
	HFI	0.0928 ± 0.0000	0.0476 ± 0.0000	0.0517 ± 0.0000	0.0689 ± 0.0000
	SFI	0.0927 ± 0.0001	0.0474 ± 0.0000	0.0516 ± 0.0003	0.0687 ± 0.0001
	MVNFMC	0.1590 ± 0.0008	0.1114 ± 0.0005	0.0314 ± 0.0005	0.0431 ± 0.0004
	GRMFMC	0.2363 ± 0.0279	0.1653 ± 0.0139	0.0578 ± 0.0014	0.0669 ± 0.0019
	HSIC-DIMFMC	0.0343 ± 0.0001	0.0186 ± 0.0000	0.0211 ± 0.0002	0.0300 ± 0.0001
30	IDW	0.1057 ± 0.0068	0.0822 ± 0.0036	0.1227 ± 0.0129	0.1429 ± 0.0052
	OKFD	0.0867 ± 0.0069	0.0950 ± 0.0160	0.1204 ± 0.0115	0.1433 ± 0.0124
	KNN	0.5665 ± 0.0338	0.5798 ± 0.0536	0.7488 ± 0.0722	0.7926 ± 0.0495
	HFI	0.0943 ± 0.0000	0.0491 ± 0.0000	0.0516 ± 0.0000	0.0691 ± 0.0000
	SFI	0.0938 ± 0.0003	0.0474 ± 0.0001	0.0522 ± 0.0002	0.0693 ± 0.0002
	MVNFMC	0.1298 ± 0.0006	0.0905 ± 0.0004	0.0265 ± 0.0003	0.0347 ± 0.0005
	GRMFMC	0.2056 ± 0.0234	0.1327 ± 0.0056	0.0501 ± 0.0033	0.0588 ± 0.0068

	HSIC-DIMFMC	0.0282 ± 0.0002	0.0156 ± 0.0004	0.0174 ± 0.0000	0.0284 ± 0.0005
	IDW	0.1082 ± 0.0074	0.0840 ± 0.0036	0.1236 ± 0.0110	0.1405 ± 0.0024
	OKFD	0.0915 ± 0.0075	0.0930 ± 0.0193	0.1157 ± 0.0051	0.1394 ± 0.0099
	KNN	0.6487 ± 0.0321	0.6539 ± 0.0418	0.7722 ± 0.057	0.8089 ± 0.0477
	HFI	0.0964 ± 0.0000	0.0485 ± 0.0000	0.0521 ± 0.0000	0.0695 ± 0.0000
40	SFI	0.0946 ± 0.0005	0.0484 ± 0.0002	0.0523 ± 0.0003	0.0702 ± 0.0002
	MVNFMC	0.1121 ± 0.0007	0.0787 ± 0.0005	0.0216 ± 0.0003	0.0300 ± 0.0004
	GRMFMC	0.1749 ± 0.0163	0.1246 ± 0.0117	0.0441 ± 0.0026	0.0513 ± 0.0026
	HSIC-DIMFMC	0.0250 ± 0.0001	0.0136 ± 0.0001	0.0160 ± 0.0001	0.0240 ± 0.0002

As can be seen from Tables 4–6:

(1) As the missing rate increases, the interpolation RMSE, MAE, and NRMSE values of the HSIC-DIMFMC method are all lower than those of IDW, OKFD, KNN, HFI, SFI, MVNFMC, and GRMFMC. Notably, the imputation error of the proposed method exhibits a decreasing trend as the missing rate increases, which sharply contrasts with traditional methods (e.g., IDW and KNN), whose errors typically escalate with higher missing rates.

The underlying mechanism for this counterintuitive phenomenon is that, under the stringent constraints of high missing rates, the low-rank representation of functional matrix completion is forced to filter out redundant noise and learn the most essential latent manifold structures, thereby achieving more robust reconstruction in severely incomplete scenarios [42,43]. Specifically, when observations become highly sparse, the low-rank constraints, the HSIC criterion, and the dual-information graph regularization take on dominant roles. These global regularizations function equivalently to applying a strong smoothing filter to the imputation process, effectively eliminating local random perturbations present in the original data. Traditional methods (such as IDW and KNN) rely heavily on local neighborhood observations; as the missing rate rises and local information diminishes, their performance inevitably degrades. In contrast, HSIC-DIMFMC depends on global low-rank topological structures and robust cross-view dependencies. Even under severe data loss, as long as the global structural integrity is preserved, it can still achieve robust trajectory reconstruction. Therefore, this behavior is not an experimental anomaly, but rather an inherent denoising property of the proposed model under extreme missing conditions.

(2) Under experimental conditions with data-missing rates ranging from 10% to 40%, the HSIC-DIMFMC method demonstrated significant performance advantages in interpolation tasks for multiple meteorological variables. Specifically, for TEMP and DEWP data, the method achieved maximum RMSE reductions of 86.27% and 89.62%, respectively, compared to seven mainstream benchmark methods (IDW, OKFD, KNN, HFI, SFI, MVNFMC, and GRMFMC), despite slight fluctuations in certain low missing data scenarios. When processing WDSP and MXSPD data, the method demonstrated more robust superiority, achieving comprehensive RMSE reductions. The maximum reduction reached 85.36% for WDSP and 81.36% for MXSPD, significantly outperforming comparison models such as MVNFMC and GRMFMC. Although GRMFMC also incorporates the HSIC criterion, its graph regularization relies solely on the local manifold assumption, making it difficult to capture global discriminative feature information. Consequently, it exhibits unstable performance on variables such as TEMP and DEWP. In contrast, the proposed HSIC-DIMFMC deeply integrates local spatiotemporal consistency with global discriminative manifold information via dual-information graph regularization, effectively overcoming the limitations of a single manifold assumption.

(3) The imputation errors of the three multi-view methods (MVNFMC, GRMFMC, and HSIC-DIMFMC) are significantly lower than those of the single-view methods (IDW,

OKFD, and KNN), compellingly demonstrating that multi-view information fusion is a critical pathway for enhancing meteorological imputation accuracy. Among the multi-view models, HSIC-DIMFMC consistently outperforms MVNFMC and GRMFMC. This fully validates the necessity of jointly introducing the HSIC criterion and dual-information graph regularization: the former extracts nonlinear complementary features across the view dimension, while the latter preserves multi-scale topological structures across the sample dimension. The deep synergy between these two components yields performance gains that far exceed those of individual modular improvements.

To evaluate the stability of the imputation results, we calculated the 95% confidence intervals (CIs) for the RMSE of each method at 10% and 40% missing rates across 50 independent runs. The results are summarized in Table 7. (Bold values indicate the best results.)

Table 7. The 95% confidence intervals of RMSE for the eight imputation methods at 10% and 40% missing rates (over 50 independent runs).

Missing Rate (%)	Baseline Method	TEMP	DEWP	WDSP	MXSPD
10	IDW	[1.8985, 1.9648]	[1.9784, 2.0479]	[0.7709, 0.7987]	[1.4246, 1.4677]
	OKFD	[1.7210, 1.7926]	[1.8222, 1.9000]	[0.8581, 0.8946]	[1.5953, 1.6631]
	KNN	[1.7279, 1.8029]	[1.1024, 1.2616]	[1.0087, 1.1525]	[1.5338, 1.6631]
	HFI	[2.1244, 2.1244]	[1.2453, 1.2453]	[0.7604, 0.7604]	[1.2299, 1.2299]
	SFI	[2.1220, 2.1223]	[1.2416, 1.2427]	[0.7631, 0.7651]	[1.2303, 1.2313]
	MVNFMC	[5.1391, 5.1501]	[4.1468, 4.1549]	[0.6515, 0.6562]	[1.0746, 1.0810]
	GRMFMC	[7.7598, 8.1646]	[6.0194, 6.3226]	[1.2460, 1.2852]	[1.7567, 1.8148]
	HSIC-DIMFMC	[1.0950, 1.0965]	[0.6896, 0.6905]	[0.4266, 0.4268]	[0.7497, 0.7502]
40	IDW	[2.1174, 2.1608]	[2.2058, 2.2464]	[0.8439, 0.8581]	[1.5168, 1.5423]
	OKFD	[2.0027, 2.0495]	[2.1089, 2.1594]	[0.9353, 0.9625]	[1.6628, 1.7007]
	KNN	[2.6521, 2.7277]	[1.7534, 1.8184]	[1.1689, 1.2190]	[1.7509, 1.8105]
	HFI	[2.1924, 2.1924]	[1.2879, 1.2879]	[0.8011, 0.8011]	[1.2557, 1.2557]
	SFI	[2.1807, 2.1903]	[1.2647, 1.2660]	[0.7708, 0.7730]	[1.2676, 1.2688]
	MVNFMC	[2.5699, 2.5787]	[2.0661, 2.0744]	[0.3266, 0.3288]	[0.5446, 0.5483]
	GRMFMC	[3.9194, 4.0881]	[3.0439, 3.1859]	[0.6271, 0.6458]	[0.8839, 0.9103]
	HSIC-DIMFMC	[0.5729, 0.5748]	[0.3571, 0.3584]	[0.2290, 0.2291]	[0.5017, 0.5145]

A narrower CI indicates the superior repeatability and robustness of a method. As clearly observed from the table, HSIC-DIMFMC achieves the narrowest CIs across all views and at both missing rates. This demonstrates that our proposed method not only yields the highest imputation accuracy (i.e., the lowest mean RMSE) but also exhibits remarkable stability under varying random initializations. In contrast, baseline methods such as GRMFMC and MVNFMC exhibit much wider CI ranges, indicating that their performance is more sensitive to initial conditions or data missing patterns.

To verify whether the performance improvements of HSIC-DIMFMC over the other methods are statistically significant, paired sample *t*-tests were conducted to compare the RMSE of HSIC-DIMFMC against the seven baseline methods at 10% and 40% missing rates. Table 8 summarizes the mean differences in RMSE, their corresponding 95% CIs, *t*-values, and *p*-values.

Table 8. Results of paired sample *t*-tests between HSIC-DIMFMC and baseline methods at 10% and 40% missing rates.

Missing Rate (%)	View	Baseline Method	Mean Difference	95% CI	<i>t</i> -Value	<i>p</i> -Value
10	TEMP	IDW	0.8359	[0.8028, 0.8690]	50.757	<0.001
		OKFD	0.6610	[0.6253, 0.6967]	37.245	<0.001
		KNN	0.6696	[0.6321, 0.7072]	35.846	<0.001
		HFI	1.0286	[1.0279, 1.0294]	2825.572	<0.001
		SFI	1.0263	[1.0256, 1.0270]	2904.346	<0.001
		MVNFMC	4.0488	[4.0433, 4.0544]	1473.804	<0.001
		GRMFMC	6.8664	[6.6640, 7.0689]	68.158	<0.001
	DEWP	IDW	1.3231	[1.2883, 1.3579]	76.413	<0.001
		OKFD	1.1711	[1.1322, 1.2099]	60.537	<0.001
		KNN	0.4919	[0.4124, 0.5715]	12.428	<0.001
		HFI	0.5553	[0.5548, 0.5558]	2274.391	<0.001
		SFI	0.5521	[0.5514, 0.5528]	1496.924	<0.001
		MVNFMC	3.4608	[3.4567, 3.4649]	1692.150	<0.001
		GRMFMC	5.4810	[5.3293, 5.6326]	72.617	<0.001
	WDSP	IDW	0.3580	[0.3441, 0.3720]	51.681	<0.001
		OKFD	0.4496	[0.4313, 0.4679]	49.293	<0.001
		KNN	0.6538	[0.5820, 0.7257]	18.277	<0.001
		HFI	0.3336	[0.3335, 0.3337]	6066.952	<0.001
		SFI	0.3374	[0.3363, 0.3384]	661.629	<0.001
		MVNFMC	0.2272	[0.2248, 0.2295]	193.960	<0.001
		GRMFMC	0.8389	[0.8193, 0.8584]	86.207	<0.001
	MXSPD	IDW	0.6962	[0.6746, 0.7177]	64.963	<0.001
		OKFD	0.8793	[0.8454, 0.9131]	52.173	<0.001
		KNN	0.8485	[0.7838, 0.9133]	26.343	<0.001
		HFI	0.4800	[0.4798, 0.4802]	4339.037	<0.001
		SFI	0.4808	[0.4803, 0.4814]	1802.942	<0.001
		MVNFMC	0.3279	[0.3247, 0.3310]	207.313	<0.001
		GRMFMC	1.0358	[1.0068, 1.0649]	71.696	<0.001
40	TEMP	IDW	1.5653	[1.5436, 1.5869]	145.336	<0.001
		OKFD	1.4523	[1.4289, 1.4756]	124.835	<0.001
		KNN	2.1160	[2.0783, 2.1538]	112.589	<0.001
		HFI	1.6185	[1.6176, 1.6195]	3418.193	<0.001
		SFI	1.6116	[1.6067, 1.6165]	664.411	<0.001
		MVNFMC	2.0005	[1.9962, 2.0047]	948.332	<0.001
		GRMFMC	3.4299	[3.3457, 3.5141]	81.861	<0.001
	DEWP	IDW	1.8684	[1.8479, 1.8888]	183.823	<0.001
		OKFD	1.7765	[1.7512, 1.8017]	141.445	<0.001
		KNN	1.4282	[1.3957, 1.4606]	88.380	<0.001
		HFI	0.9302	[0.9295, 0.9308]	2928.102	<0.001
		SFI	0.9076	[0.9068, 0.9085]	2140.213	<0.001
		MVNFMC	1.7125	[1.7082, 1.7168]	800.185	<0.001
		GRMFMC	2.7571	[2.6862, 2.8281]	78.132	<0.001
	WDSP	IDW	0.6220	[0.6149, 0.6290]	177.167	<0.001
		OKFD	0.7198	[0.7062, 0.7334]	106.534	<0.001
		KNN	0.9649	[0.9398, 0.9899]	77.447	<0.001
		HFI	0.5721	[0.5720, 0.5721]	18,767.206	<0.001
		SFI	0.5429	[0.5418, 0.5439]	1024.814	<0.001
		MVNFMC	0.0987	[0.0976, 0.0998]	181.500	<0.001
		GRMFMC	0.4074	[0.3980, 0.4168]	87.429	<0.001

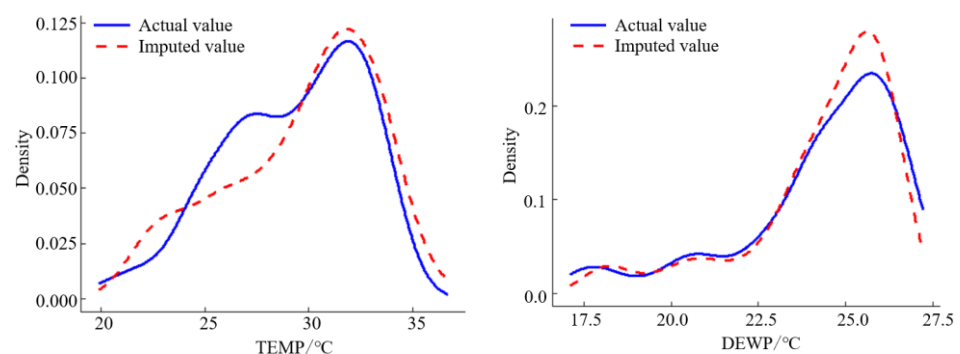
MXSPD	IDW	1.0214	[1.0074, 1.0355]	146.070	<0.001
	OKFD	1.1736	[1.1544, 1.1928]	122.840	<0.001
	KNN	1.2726	[1.2425, 1.3027]	84.972	<0.001
	HFI	0.7476	[0.7412, 0.7540]	234.555	<0.001
	SFI	0.7601	[0.7536, 0.7666]	235.588	<0.001
	MVNFMC	0.0383	[0.0318, 0.0448]	11.839	<0.001
	GRMFMC	0.3890	[0.3741, 0.4039]	52.470	<0.001

The results indicate that across all four views, the mean differences between the baseline methods and HSIC-DIMFMC are consistently positive, and the lower bounds of the 95% CIs are strictly greater than zero. This confirms that HSIC-DIMFMC achieves lower RMSE values in all scenarios. Furthermore, all p -values are less than 0.001, indicating that this performance advantage is strongly statistically significant.

In summary, HSIC-DIMFMC not only outperforms all baseline methods in terms of the average RMSE but also exhibits the minimal performance fluctuation (the narrowest 95% CIs) across the 50 independent runs. The paired t -tests further substantiate that these performance gains are statistically significant across all views, thereby thoroughly demonstrating the effectiveness and superiority of the proposed HSIC-DIMFMC framework.

4.5. Application Outcomes

To further validate the interpolation performance of the HSIC-DIMFMC method on meteorological data, experiments were conducted on dataset II, which features complete meteorological indicators. To comprehensively evaluate the method's performance, the mean error metrics were calculated for all 20 stations at a 40% missing rate (see Table 9). Additionally, due to space constraints and to visually illustrate interpolation details, stations JIANGILING and JIAN—selected for their median RMSE values—were chosen for visualization analysis based on overall statistical analysis of all 20 stations. This reflects the model's average performance. Both stations exhibit meteorological characteristics and model performance close to the dataset average, making them highly representative. Figures 11 and 12 present the probability density distributions comparing the original meteorological observations at stations JIANGILING and JIAN with the imputed values from the HSIC-DIMFMC method under a 40% missing rate. It can be observed that the distributions of observed and predicted values are similar, with consistent peak positions and distribution ranges. This indicates that the method accurately captures the distribution characteristics of the original data, and the interpolation results exhibit high fidelity.



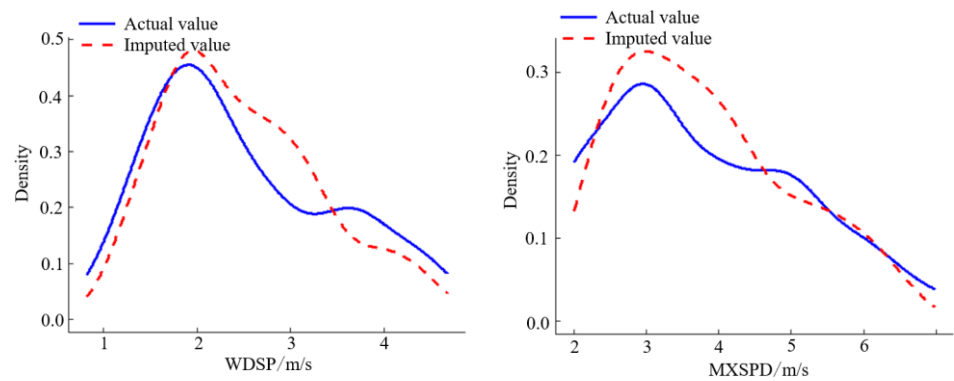


Figure 11. Density plot of observations and imputed values at station JIANGLING under a 40% missing rate.

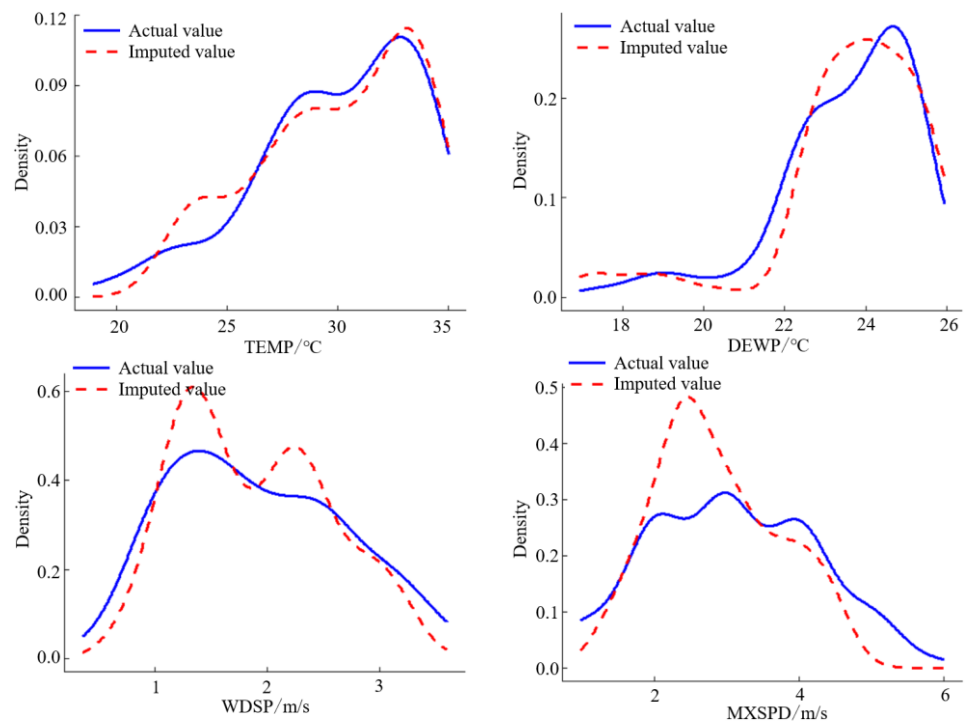


Figure 12. Density plot of observations and imputed values at station JIAN under a 40% missing rate.

Table 9. RMSE of 20 stations with 40% missing rate.

Station	Meteorological Indicators			
	TEMP	DEWP	WDSP	MXSPD
NANYUE	1.12	0.77	2.88	2.11
JIANGLING	1.58	0.93	0.61	0.98
TIANHE	1.40	0.96	0.65	1.20
NEIJIAN	2.05	1.07	0.46	0.89
JIANGBEI	2.00	0.95	0.49	1.37
SANGZHI	1.53	0.92	0.39	1.24
YUEYANG	1.47	1.30	0.69	1.16
XIUSHUI	1.62	0.89	0.45	0.90
NAXI	2.00	0.92	0.43	0.96
YOUYANG	1.44	0.78	0.41	1.02
YUANLING	1.65	0.78	0.32	0.78
CHANGDE	1.65	2.25	0.61	1.15

CHANGSHA	1.55	0.84	0.78	1.09
BIJIE	1.64	1.11	0.34	0.64
HUICHUAN	1.36	0.68	0.43	0.90
SINAN	1.73	0.84	0.53	1.11
ZHIJIANG	1.41	0.78	0.40	1.09
SHAOYANG	1.56	0.72	0.42	0.91
YICHUN	1.63	0.85	0.47	0.89
JIAN	1.79	0.88	0.55	0.89

For stations JIANGLING and JIAN, Taylor plots were employed to systematically evaluate the overall performance of eight interpolation methods in preserving data statistical characteristics. These plots presented three complementary statistical metrics: the ratio of standard deviation, root mean square error (RMSE), and correlation coefficient between interpolated results and observations at a 40% missing rate. Taylor plots for observations versus predicted values at 40% missing data rates for stations JIANGLING and JIAN are shown in Figures 13 and 14. The radial dashed lines (radiating from the origin) represent the correlation coefficient; the closer to the right, the higher the correlation with the observed data. The solid concentric arcs represent the root mean square error (RMSE); the smaller the arc (closer to the origin), the smaller the error and the more accurate the model. The hollow circles (typically on the X-axis) represent the standard deviation of the observed data.

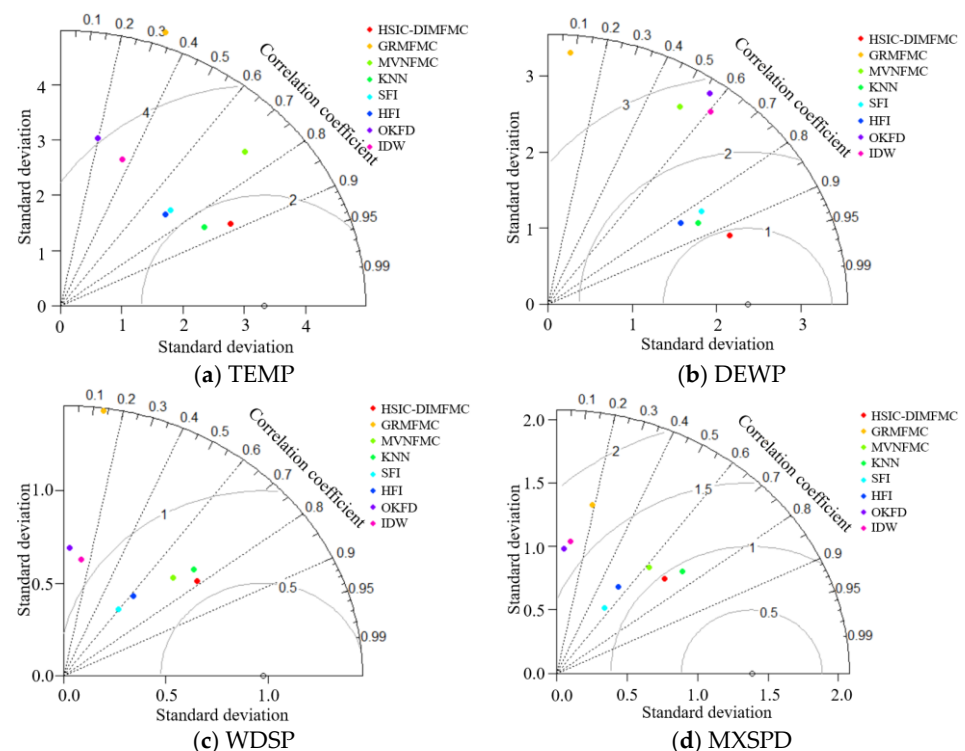


Figure 13. Taylor diagram of observations and imputed values at station JIANGLING with 40% missing rate. The numbers on the dashed radial lines and solid arcs denote correlation coefficients and normalized RMSE, respectively; the blank circle on the x-axis indicates the standard deviation of the observed data.

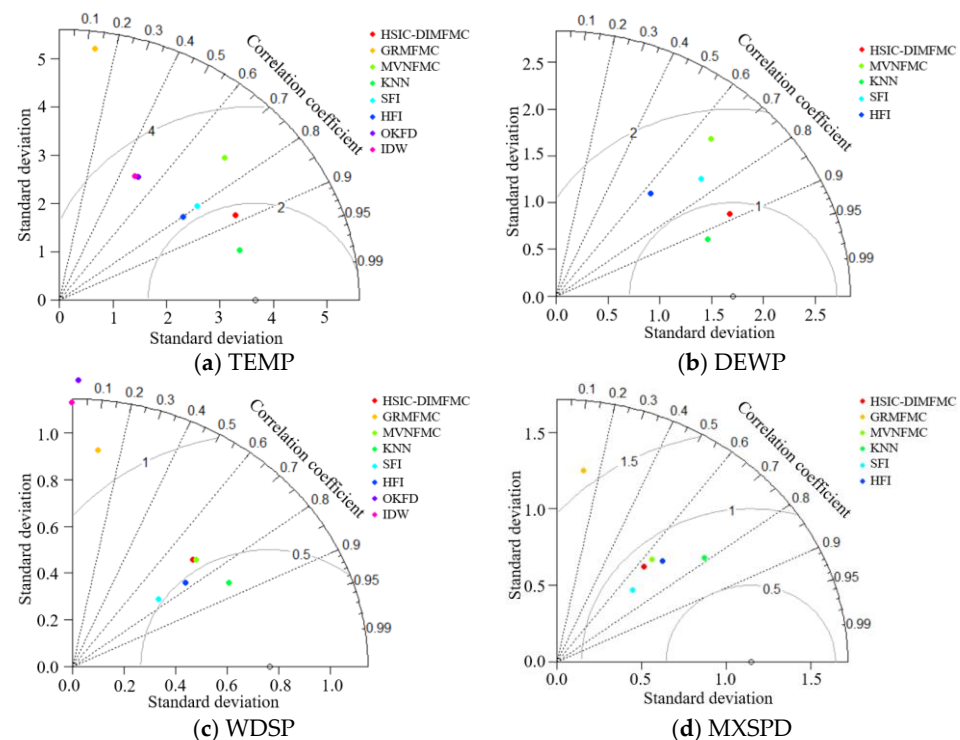


Figure 14. Taylor diagram of observations and imputed values at station JIAN with 40% missing rate. The numbers on the dashed radial lines and solid arcs denote correlation coefficients and normalized RMSE, respectively; the blank circle on the x-axis indicates the standard deviation of the observed data.

As shown in Figures 13 and 14, in the Taylor plot evaluation of four key meteorological indicators (TEMP, DEWP, WDSP, MXSPD) at stations JIANGILING and JIAN, the HSIC-DIMFMC method (red dots) demonstrates a significant and consistent overall advantage: it maintains correlation coefficients exceeding 0.85 in both Figure 13a,b and Figure 14a,b, substantially outperforming interpolation methods such as SFI, HFI, MVNFMC, GRMFMC, OKFD, and IDW. While its standard deviation is not the absolute minimum across all subplots, it consistently remains within a reasonable range. This effectively avoids the biases caused by methods like GRMFMC (which over-amplify variability) or IDW (which over-smooth data). The model demonstrates excellent stability and robustness at both representative stations and across variable types. In summary, the HSIC-DIMFMC method significantly enhances the accuracy of missing value imputation and the preservation of statistical properties, demonstrating clear performance advantages in meteorological data interpolation tasks.

To further demonstrate the effectiveness of the HSIC-DIMFMC method, missing data were estimated using this approach. Scatter plots comparing imputed values with observations were plotted for four perspectives at 10% and 40% missing data rates, with results shown in Figures 15–18.

Figures 15–18 jointly validate the high-precision interpolation capability and robustness of the HSIC-DIMFMC method for meteorological indicators through both time series and scatter plot representations. As shown in time series plots (a) and (b), imputed values fit observed data well at both 10% and 40% missing rates while preserving the original data fluctuations. This finding is corroborated in scatter plots (c) and (d), where imputed values and observations cluster closely on either side of the 45-degree line. Crucially, neither the fit quality in the time series nor the clustering in the scatter plots shows significant deterioration when the missing rate increases from 10% to 40%. This demonstrates the

method's stable performance, unaffected by increased missingness, and its practical value for handling large-scale incomplete datasets.

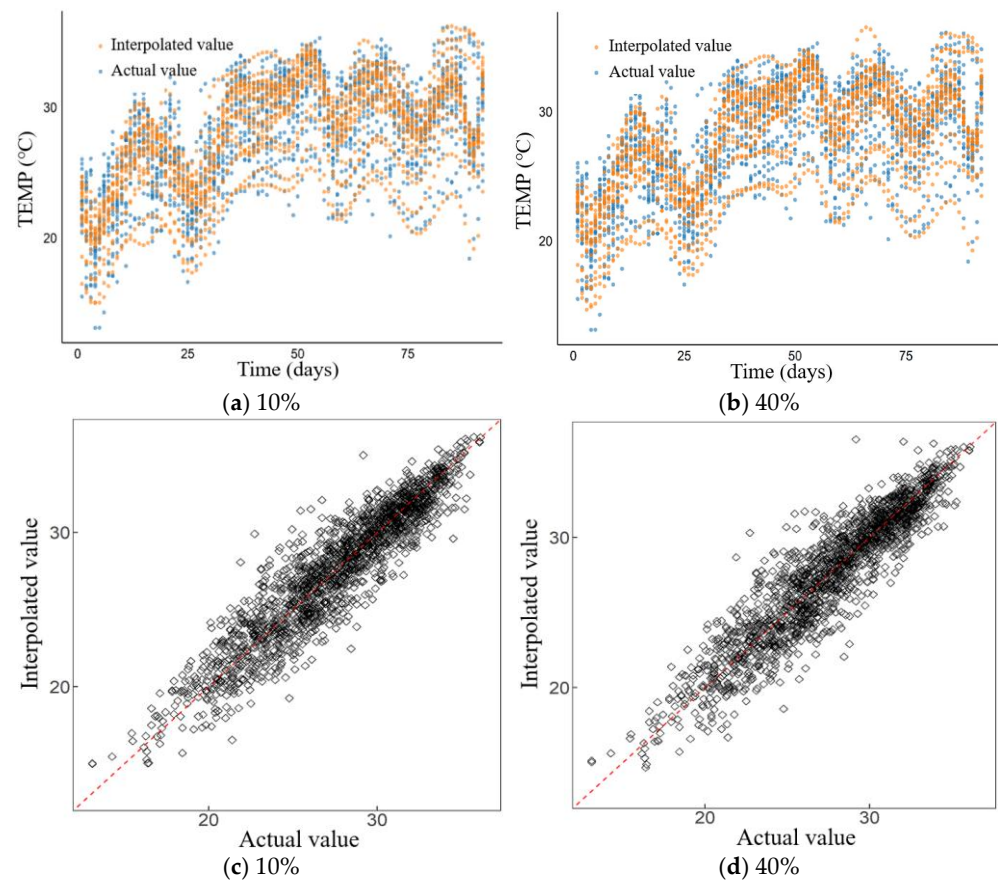
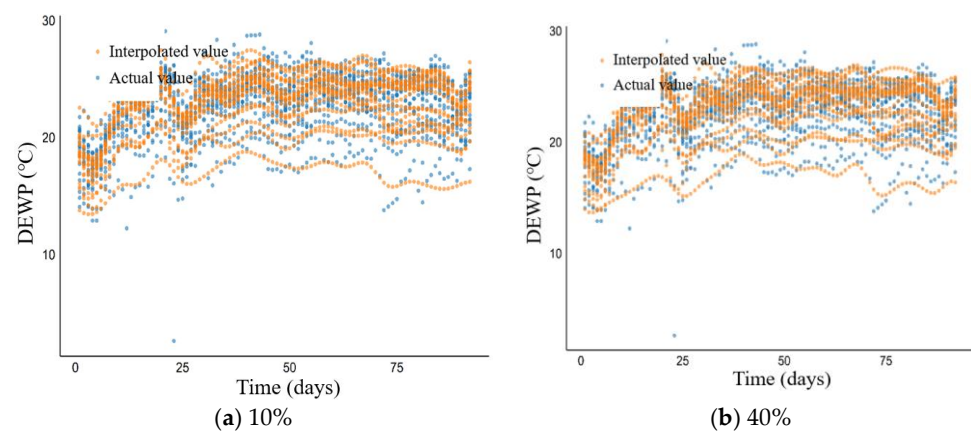


Figure 15. Scatter plot of the true values and imputed values of TEMP under 10% and 40% missing rates. The red line indicates perfect agreement ($y = x$), and the black squares represent individual samples.



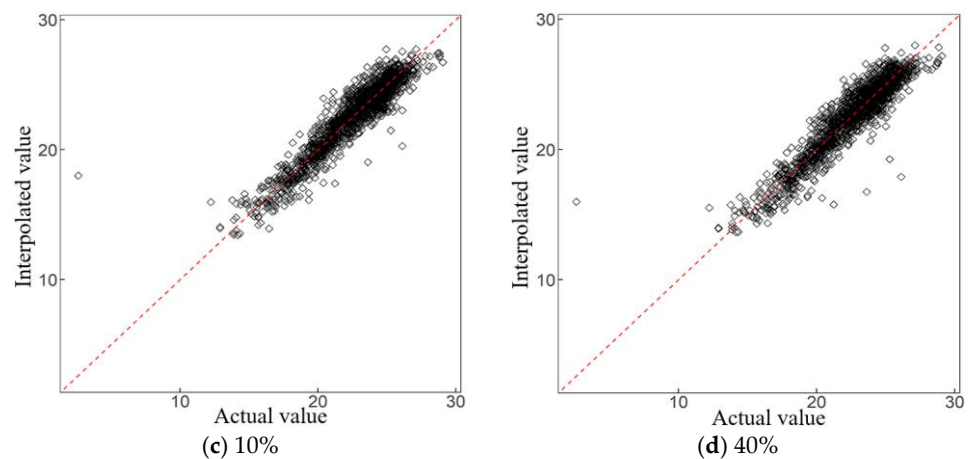


Figure 16. Scatter plot of the true values and imputed values of DEWP under 10% and 40% missing rates. The red line indicates perfect agreement ($y = x$), and the black squares represent individual samples.

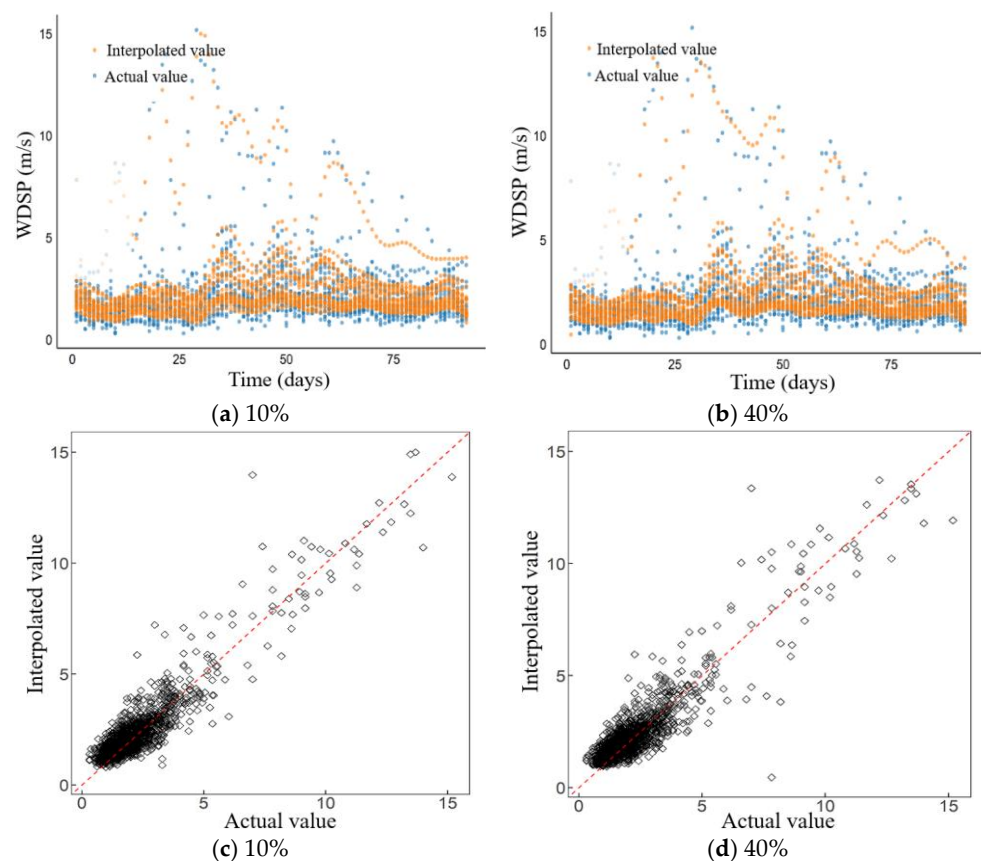


Figure 17. Scatter plot of the true values and imputed values of WDSP under 10% and 40% missing rates. The red line indicates perfect agreement ($y = x$), and the black squares represent individual samples.

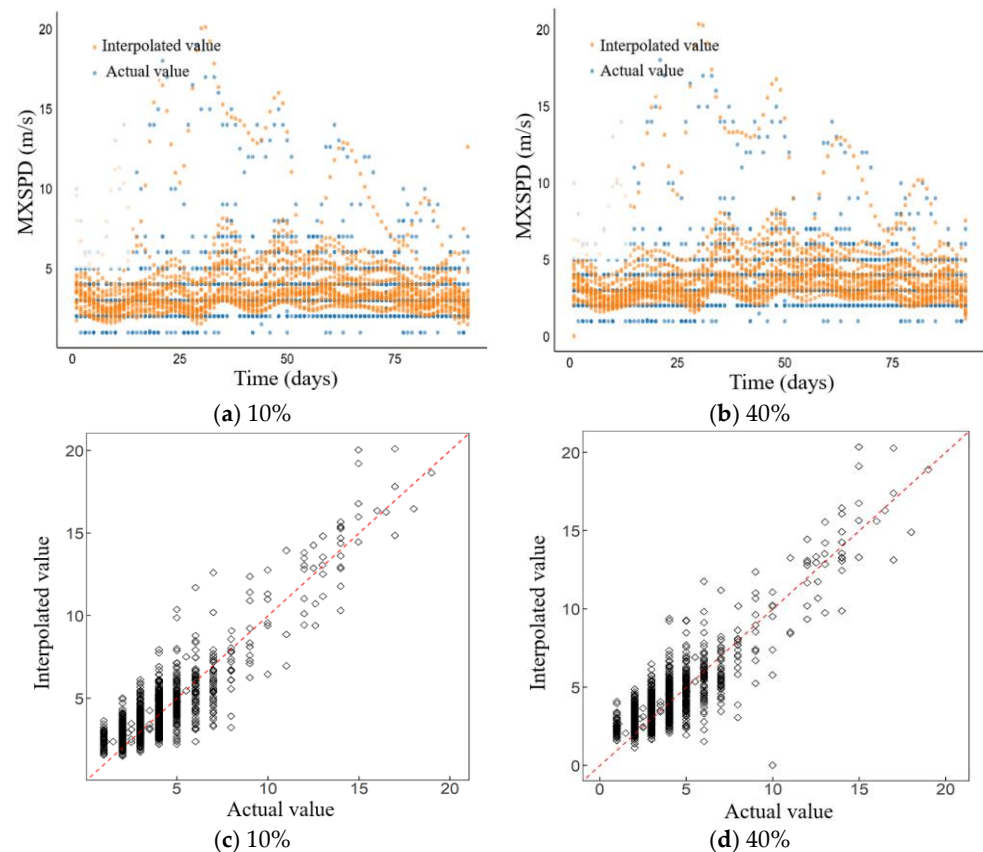


Figure 18. Scatter plot of the true values and imputed values of MXSPD under 10% and 40% missing rates. The red line indicates perfect agreement ($y = x$), and the black squares represent individual samples.

5. Conclusions

This paper proposed HSIC-DIMFMC, a multi-view functional matrix completion method with the HSIC criterion and dual-information graph regularization, to address the problem of missing meteorological data. Within the framework of functional data analysis, the proposed model integrates the complex interdependencies among meteorological indicators and effectively captures both local structural information and nonlinear cross-view dependencies. The dual-information graph regularization preserves local manifold and discriminative structures, while the HSIC criterion enhances statistical dependence between highly correlated views, thereby improving collaborative imputation accuracy.

Experimental results on meteorological datasets demonstrated that HSIC-DIMFMC consistently outperformed seven representative baseline methods (IDW, OKFD, KNN, HFI, SFI, MVNFMC, and GRMFMC) across various missing ratios. Quantitatively, the proposed method achieved significant improvements in RMSE, MAE, and NRMSE, with maximum RMSE reductions reaching 85.31%, effectively proving its capability in recovering incomplete observations. Furthermore, the method maintained highly stable imputation performance across different meteorological indicators. Most notably, it exhibited highly precise collaborative reconstruction capabilities when processing strongly correlated physical pairs, such as temperature (TEMP) and dew point (DEWP), as well as wind speed (WDSP) and maximum wind speed (MXSPD). Consequently, under the evaluated data conditions, the proposed approach demonstrates excellent adaptability and significant practical value, offering a robust technical pathway for meteorological data quality control.

From the perspective of practical operational deployment, real-world meteorological observation networks often face harsher challenges than ideal experimental setups, such as sudden sensor data jumps, prolonged equipment disconnections, drastic fluctuations driven by extreme weather, and concurrent multi-station failures caused by regional heavy precipitation. Although this study did not design isolated experiments for all extreme scenarios, the theoretical architecture of HSIC-DIMFMC inherently provides technical advantages to address these complex conditions: First, by utilizing temporal continuity, the model acts as a buffer to smooth out reconstruction deviations caused by sudden single-point anomalies. Second, for prolonged data gaps, rather than relying solely on scarce adjacent observations, the model extracts inherent diurnal or seasonal periodic patterns to assist reconstruction. Third, its nonlinear mapping architecture is far more adaptable to massive data swings during extreme weather events than traditional linear extrapolation. Fourth, in scenarios of concurrent multi-station failures where spatial information is severely depleted, the spatiotemporal joint modeling mechanism can still leverage temporal evolution trends to maintain physically reasonable results. Objectively, these advantages are currently grounded in theoretical structural deduction; their empirical validity in complex real-world meteorological operations requires dedicated verification through extensive authentic fault cases in future work.

Despite its superior imputation performance, HSIC-DIMFMC possesses certain limitations: (1) Sensitivity to outliers: The dual-information graph regularization heavily relies on KNN to construct neighborhood topologies. Significant outliers hidden in the raw data can distort the manifold structure, thereby compromising imputation accuracy; (2) High computational complexity: The joint optimization of matrix completion and graph regularization incurs a considerable computational overhead when applied to large-scale spatial station datasets.

Future work will focus on incorporating physical constraints and deep learning components to further improve the physical consistency and predictive utility of the reconstruction results. Furthermore, the evaluation framework will be expanded to encompass MAR and MNAR missing mechanisms. By designing tailored simulation experiments, we aim to verify the applicability and robustness of the proposed method under complex missing-data scenarios. Ultimately, the framework will be extended to related domains, including air quality monitoring, hydrological time series, and ocean observation data, in order to assess its broader applicability and cross-domain transferability.

Author Contributions: Conceptualization, H.G. and Y.B.; methodology, H.G. and Y.B.; software, Y.B.; validation, H.G. and Y.B.; formal analysis, H.G. and Y.B.; investigation, H.G. and Y.B.; resources, H.G. and Y.B.; data curation, Y.B.; writing—original draft preparation, Y.B.; writing—review and editing, H.G. and Y.B.; visualization, Y.B.; supervision, H.G.; project administration, H.G. and Y.B.; funding acquisition, H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Fund of China (NO. 19XTJ002); the National Key Statistical Science Research Project (NO. 2025LZ007); the Gansu Provincial Natural Science Foundation (NO. 23JRRA1186); and the Gansu Provincial Universities' Young Doctor Support Program (NO. 2025QB-058).

Data Availability Statement: The data presented in this study are available in the National Climatic Data Center (NCDC) of the United States at <https://www.ncei.noaa.gov/data/global-summary-of-the-day/archive>. (accessed on 1 June 2026)

Acknowledgments: The authors sincerely appreciate the editors and reviewers for their valuable comments and professional suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Reek, T.; Doty, S.R.; Owen, T.W. A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network. *Bull. Am. Meteorol. Soc.* **1992**, *73*, 753–765. [https://doi.org/10.1175/1520-0477\(1992\)073<0753:ADATTV>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<0753:ADATTV>2.0.CO;2).
2. Farlow, S.J. The GMDH algorithm of Ivakhnenko. *Am. Stat.* **1981**, *35*, 210–215. <https://doi.org/10.1080/00031305.1981.10479358>.
3. Acock, M.C.; Pachepsky, Y.A. Estimating missing weather data for agricultural simulations using group method of data handling. *J. Appl. Meteorol.* **2000**, *39*, 1176–1184. [https://doi.org/10.1175/1520-0450\(2000\)039<1176:EMW DFA>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1176:EMW DFA>2.0.CO;2).
4. Pielke, R.A., Sr.; Stohlgren, T.; Schell, L.; Parton, W.; Doesken, N.; Redmond, K.; Moeny, J.; McKee, T.; Kittel, T.G.F. Problems in evaluating regional and local trends in temperature: An example from eastern Colorado, USA. *Int. J. Climatol.* **2002**, *22*, 421–434. <https://doi.org/10.1002/joc.706>.
5. Chapon, A.; Ouarda, T.B.M.J.; Hamdi, Y. Imputation of missing values in environmental time series by D-vine copulas. *Weather Clim. Extrem.* **2023**, *41*, 100591. <https://doi.org/10.1016/j.wace.2023.100591>.
6. Eischeid, J.K.; Pasteris, P.A.; Diaz, H.F.; Plantico, M.S.; Lott, N.J. Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteorol.* **2000**, *39*, 1580–1591. [https://doi.org/10.1175/1520-0450\(2000\)039<1580:CASCND>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1580:CASCND>2.0.CO;2).
7. Chen, Y.; Li, Q.; Jiao, B. Homogenization of Air Temperature in Xinxiang, China, Using RHtestsV4 and Implications for Regional Climate Change Assessment. *Atmosphere* **2026**, *17*, 485. <https://doi.org/10.3390/atmos17050485>.
8. Sun, Y.; Wang, H.; Zhou, Y.; Yan, J.; Liu, Y. Applicability of three interpolation methods in estimating daily temperature with missing data from regional automatic weather station. *Torrential Rain Disasters* **2023**, *42*, 97–104. <https://doi.org/10.12406/byzh.2022-183>.
9. Varentsov, M.; Esau, I.; Wolf, T. High-resolution temperature mapping by geostatistical kriging with external drift from large-eddy simulations. *Mon. Weather Rev.* **2020**, *148*, 1029–1048. <https://doi.org/10.1175/MWR-D-19-0196.1>.
10. Allen, R.J.; DeGaetano, A.T. Estimating missing daily temperature extremes using an optimized regression approach. *Int. J. Climatol.* **2001**, *21*, 1305–1319. <https://doi.org/10.1002/joc.679>.
11. DeGaetano, A.T.; Eggleston, K.L.; Knapp, W.W. A method to estimate missing daily maximum and minimum temperature observations. *J. Appl. Meteorol. Climatol.* **1995**, *34*, 371–380. <https://doi.org/10.1175/1520-0450-34.2.371>.
12. Huth, R.; Nemešová, I. Estimation of missing daily temperatures: Can a weather categorization improve its accuracy? *J. Clim.* **1995**, *8*, 1901–1916. [https://doi.org/10.1175/1520-0442\(1995\)008<1901:EOMDTC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1901:EOMDTC>2.0.CO;2).
13. Kemp, W.P.; Burnell, D.G.; Everson, D.O.; Thomson, A.J. Estimating missing daily maximum and minimum temperatures. *J. Clim. Appl. Meteorol.* **1983**, *22*, 1587–1593. [https://doi.org/10.1175/1520-0450\(1983\)022<1587:EMDMAM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1587:EMDMAM>2.0.CO;2).
14. Shao, H.; Wei, H.; Deng, X.; Xing, S. Short-term wind speed forecasting using wavelet transformation and AdaBoosting neural networks in Yunnan wind farm. *IET Renew. Power Gener.* **2017**, *11*, 374–381. <https://doi.org/10.1049/iet-rpg.2016.0118>.
15. Li, X.; Li, K.; Shen, S.; Tian, Y. Exploring time series models for wind speed forecasting: A comparative analysis. *Energies* **2023**, *16*, 7785. <https://doi.org/10.3390/en16237785>.
16. Shen, L.; Ma, Q.; Li, S. End-to-end time series imputation via residual short paths. In *Asian Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2018; pp. 248–263. Available online: <https://proceedings.mlr.press/v95/shen18a.html> (accessed on 3 July 2026).
17. Cao, W.; Wang, D.; Li, J.; Zhou, H.; Li, L.; Li, Y. Brits: Bidirectional recurrent imputation for time series. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6775–6785. <https://doi.org/10.48550/arXiv.1805.10572>.
18. Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent neural networks for multivariate time series with missing values. *Sci. Rep.* **2018**, *8*, 6085. <https://doi.org/10.1038/s41598-018-24271-9>.
19. Zhang, Y.F.; Thorburn, P.J.; Xiang, W.; Fitch, P. SSIM—A deep learning approach for recovering missing time series sensor data. *IEEE Internet Things J.* **2019**, *6*, 6618–6628. <https://doi.org/10.1109/JIOT.2019.2909038>.
20. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0).
21. Elgohari, H.; Asim, M. Effects of missing at random imputation methods on time series data analysis. *Statistics* **2026**, *60*, 835–858. <https://doi.org/10.1080/02331888.2025.2584622>.
22. Hua, V.; Nguyen, T.; Dao, M.S.; Nguyen, H.D.; Nguyen, B.T. The impact of data imputation on air quality prediction problem. *PLoS ONE* **2024**, *19*, e0306303. <https://doi.org/10.1371/journal.pone.0306303>.

23. Nie, T.; Qin, G.; Ma, W.; Mei, Y.; Sun, J. ImputeFormer: Low rankness-induced transformers for generalizable spatiotemporal imputation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery (ACM): New York, NY, USA, 2024; pp. 2260–2271. <https://doi.org/10.1145/3637528.3671751>.
24. Nejad, A.S.; Alaiz-Rodríguez, R.; McCarthy, G.D.; Kelleher, B.; Grey, A.; Parnell, A. SERT: A transformer based model for multivariate temporal sensor data with missing values for environmental monitoring. *Comput. Geosci.* **2024**, *188*, 105601. <https://doi.org/10.1016/j.cageo.2024.105601>.
25. Dimitri, G.M.; Cappelli, I.; Scarselli, F.; Fort, A.; Gori, M. Graph neural networks for missing data imputation in time series from meteorological sensors. In *Proceedings of the 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRaine)*; IEEE: New York, NY, USA, 2024; pp. 1242–1247. <https://doi.org/10.1109/MetroXRaine62247.2024.10796616>.
26. Yao, X.H. Research on Several Multivariate Function-Based Clustering Methods Based on Multi-Perspective Learning. Ph.D. Dissertation, Lanzhou University of Finance and Economics, Lanzhou, China, 2022.
27. Descary, M.H.; Panaretos, V.M. Functional data analysis by matrix completion. *Ann. Stat.* **2019**, *47*, 1–38. <https://doi.org/10.1214/17-AOS1590>.
28. Kong, S.; Wang, X.; Wang, D.; Wu, F. Multiple feature fusion for face recognition. In *Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*; IEEE: New York, NY, USA 2013; pp. 1–7. <https://doi.org/10.1109/FG.2013.6553718>.
29. Xu, H.; Zhang, X.; Xia, W.; Gao, Q.; Gao, X. Low-rank tensor constrained co-regularized multi-view spectral clustering. *Neural Netw.* **2020**, *132*, 245–252. <https://doi.org/10.1016/j.neunet.2020.08.019>.
30. Xue, J.; Fu, D.Y.; Han, H.B.; Gao, H.Y. Matrix completion algorithm for non-negative functions based on multi-perspective learning. *Stat. Decis. Mak.* **2022**, *38*, 5–11. <https://doi.org/10.13546/j.cnki.tjyjc.2022.07.001>.
31. Gao, H.Y.; Ma, W.J. Air Quality Data Repair Based on Graph Regularization and Multi-view Functional Matrix Completion. *Chin. J. Environ. Sci.* **2024**, *44*, 5357–5370. <https://doi.org/10.19674/j.cnki.issn1000-6923.20240708.003>.
32. Khan, G.A.; Hu, J.; Li, T.; Diallo, B.; Wang, H. Multi-view data clustering via non-negative matrix factorization with manifold regularization. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 677–689. <https://doi.org/10.1007/s13042-021-01307-7>.
33. Zong, L.; Zhang, X.; Zhao, L.; Yu, H.; Zhao, Q. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. *Neural Netw.* **2017**, *88*, 74–89. <https://doi.org/10.1016/j.neunet.2017.02.003>.
34. Chen, F.; Li, G.; Wang, S.; Pan, Z. Multiview clustering via robust neighboring constraint nonnegative matrix factorization. *Math. Probl. Eng.* **2019**, *2019*, 6084382. <https://doi.org/10.1155/2019/6084382>.
35. Li, S.; Yang, C.; Guo, H. Auto-adjustable dual-information graph regularized NMF for multiview data clustering. *Pattern Recognit.* **2025**, *166*, 111679. <https://doi.org/10.1016/j.patcog.2025.111679>.
36. Gretton, A.; Bousquet, O.; Smola, A.; Schölkopf, B. Measuring statistical dependence with Hilbert-Schmidt norms. In *International Conference on Algorithmic Learning Theory*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 63–77. https://doi.org/10.1007/11564089_7.
37. Chen, F.W.; Liu, C.W. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy Water Environ.* **2012**, *10*, 209–222. <https://doi.org/10.1007/s10333-012-0319-1>.
38. Giraldo, R.; Delicado, P.; Mateu, J. Ordinary kriging for function-valued spatial data. *Environ. Ecol. Stat.* **2011**, *18*, 411–426. <https://doi.org/10.1007/s10651-010-0143-y>.
39. Mucherino, A.; Papajorgji, P.J.; Pardalos, P.M. K-nearest neighbor classification. In *Data Mining in Agriculture*; Springer: New York, NY, USA, 2009; pp. 83–106. https://doi.org/10.1007/978-0-387-88615-2_4.
40. Kidziński, Ł.; Hastie, T. Longitudinal data analysis using matrix completion. *Stat* **2018**, *1050*, 24. <https://doi.org/10.48550/arXiv.1809.08771>.
41. Jamshidian, M.; Jalal, S.; Jansen, C. MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *J. Stat. Softw.* **2014**, *56*, 1–31. <https://doi.org/10.18637/jss.v056.i06>.

42. Bai, Z.; Zhao, J.; Zhang, Y. Connectivity shapes implicit regularization in matrix factorization models for matrix completion. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 45914–45955. <https://doi.org/10.52202/079017-1460>.
43. Li, Z.; Sun, T.; Wang, H.; Xu, Z. AIR-Net: Adaptive and implicit regularization neural network for matrix completion. *arxiv* **2021**, arXiv:2110.07557. <https://doi.org/10.48550/arXiv.2110.07557>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.