

Article

Missing Data Imputation for Reservoir Inflow Flood Discharge of Dams Based on Improved Singular Value Decomposition

Yongjiang Chen ^{1,2}, Kui Wang ^{1,*} , Mingjie Zhao ^{1,2}, Gang Liu ¹ and Jianfeng Liu ^{1,2}

¹ Engineering Research Centre of Diagnosis Technology of Hydro-Construction, Chongqing Jiaotong University, Chongqing 400074, China

² School of Civil and Hydraulic Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

* Correspondence: anhuiwk@163.com

Abstract

Missing values commonly exist in dam inflow flood discharge monitoring data, which hinders flood analysis, risk assessment and reservoir scheduling. Aiming at the problems of insufficient imputation accuracy and the difficulty in adaptive threshold selection of traditional Singular Value Decomposition (SVD) in flood discharge data with strong fluctuations and high noise, this study introduces a method for filling in missing dam inflow flood discharge based on Dam Monitoring Data Reconstruction Model (DSVD). The method constructs a non-repeating sequence monitoring matrix, introduces a hard singular value threshold for adaptive denoising, and completes time series data imputation combined with a weight optimization model, which effectively improves the imputation accuracy of strongly fluctuating flood discharge data. Taking the measured inflow flood discharge data of Jinjiaba Reservoir in Chongqing as the research object, this study systematically analyzes the influence of column-to-row ratio (R_a) and data missing rate on imputation performance, and conducts a comparative verification against other models. Experimental results indicate that the optimal R_a value is 6. The coefficient of determination (R^2) stays above 0.830 within a missing rate range of 5–40%, showing strong robustness against data loss. Compared with other benchmark models, the method has the highest R^2 (0.875) and the lowest Root Mean Square Error (RMSE, 7.771), exhibiting stronger adaptability to mountainous flood discharge data with steep rise and fall characteristics. The research findings provide a new method for the high-precision recovery of missing dam inflow flood discharge data and reliable data support for reservoir flood risk analysis and safe operation.

Keywords: dam inflow flood discharge; missing data imputation; singular spectrum analysis; singular value decomposition; DSVD



Academic Editor: James E. Landmeyer

Received: 19 May 2026

Revised: 23 June 2026

Accepted: 24 June 2026

Published: 26 June 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

Reservoir inflow flood discharge is the core basic data for dam flood control safety, flood overtopping risk assessment and reservoir operation, and its completeness and accuracy directly determine the reliability of flood process simulation, flood peak prediction and flood control decision-making [1–3]. Affected by complex hydrometeorological conditions in mountainous areas, long-term operation loss of monitoring equipment, extreme weather interference and communication transmission failures, the monitoring series of dam inflow flood discharge often has varying degrees of missing or abnormal data, resulting in discontinuous time series and information distortion. This seriously impairs the accuracy of

flood process analysis and even affects the reliability of dam flood control safety evaluation results [4–6]. Therefore, research on high-precision imputation methods for missing dam inflow flood discharge data is of great engineering significance for ensuring safe reservoir operation and improving flood risk management capability.

At present, missing data imputation methods for hydrological time series mainly include traditional interpolation methods, statistical regression methods and intelligent algorithms. Traditional interpolation methods (e.g., linear interpolation, spline interpolation) are simple to calculate but difficult to capture the strong fluctuation characteristics of steeply rising and falling flood discharge, leading to large imputation errors at high missing rates; statistical regression methods rely on the assumption of time series stationarity and have poor adaptability to non-stationary and high-noise data such as mountain river floods. In recent years, deep learning models including Generative Adversarial Imputation Nets (GAIN) [7], Timing GAIN (TGAIN) [8], Dam Temporal Reconstruction Nets (DTRN) [9] and Dam Monitoring Data Reconstruction Network (DMDRN) [10] have been widely applied in time series data recovery. However, these models feature complex network structures and a large number of training parameters, and tend to suffer from overfitting and low imputation accuracy when processing highly fluctuating flood discharge data with limited samples.

As a classic matrix decomposition technique, Singular Value Decomposition (SVD) has strong data feature extraction and noise suppression capabilities, and has been gradually applied in hydrological data recovery [11,12]. Nevertheless, traditional SVD has obvious limitations: first, its matrix construction strategy is simplistic and not optimized to fit the strong fluctuation features of flood discharge data; second, the singular value threshold relies on manual trial calculation with poor adaptability; third, the imputation model ignores time series correlation, and noise is easily propagated into the recovery results, resulting in insufficient imputation accuracy for highly fluctuating data [13–17].

To address the above problems, this paper integrates the idea of singular spectrum analysis and introduces the Dam Monitoring Data Reconstruction Model (DSVD) [18]. The core objectives are to improve the recovery of missing dam inflow flood discharge data through non-repeating sequence matrix construction, hard singular value thresholding and adaptive imputation: a non-repeating sequence is adopted to construct the matrix, reducing computational cost and adapting to the independent fluctuation characteristics of flood discharge; a hard singular-value threshold (HSVT) method is introduced to adaptively determine the threshold based on the effective data ratio, realizing the accurate removal of strong noise; a weight optimization imputation model is constructed to minimize imputation error and suppress noise transmission. Taking the measured inflow flood discharge data of Jinjiaba Reservoir in Chongqing as the experimental object, this study systematically verifies the imputation performance of the method under different column-to-row ratios and missing rates, and conducts a comparative analysis with multiple mainstream models. This study aims to introduce a high-precision and highly adaptive recovery method for missing dam inflow flood discharge data and data support for reservoir flood control safety and risk assessment.

2. Methodology

2.1. Basic Principle of Singular Value Decomposition

Singular Value Decomposition (SVD) [19] is a classic matrix decomposition method that decomposes any matrix into a combination of orthogonal matrices and a diagonal

singular value matrix, enabling data feature extraction and information compression. For an arbitrary matrix A , its decomposition form is as follows:

$$A = U\Sigma V^T \quad (1)$$

where U and V are the left and right orthogonal singular vector matrices, respectively; Σ is a diagonal matrix composed of singular values arranged in descending order. The first several large singular values retain most of the effective information of the matrix, and discarding small singular values enables approximate matrix reconstruction. Based on this characteristic, SVD can be used for missing data imputation with the following general process: first, perform mean initial imputation on missing values to obtain a complete matrix; then, conduct an SVD and retain the main singular values for approximate matrix imputation; finally, replace the original missing values with the values at the corresponding positions of the approximate matrix, and iterate until convergence.

2.2. Missing Data Imputation Method for Dam Inflow Flood Discharge Based on DSVD

Aiming at the problems of strong fluctuation, high noise, difficulty in threshold determination and insufficient imputation accuracy of traditional SVD for dam inflow flood discharge data, this paper integrates the idea of singular spectrum analysis [20] and introduces the Dam Monitoring Data Reconstruction Model (DSVD) [18]. The core steps are non-repeating sequence matrix construction, HSVT and adaptive imputation to improve the recovery performance of missing dam inflow flood discharge data. The corresponding process is shown in Figure 1.

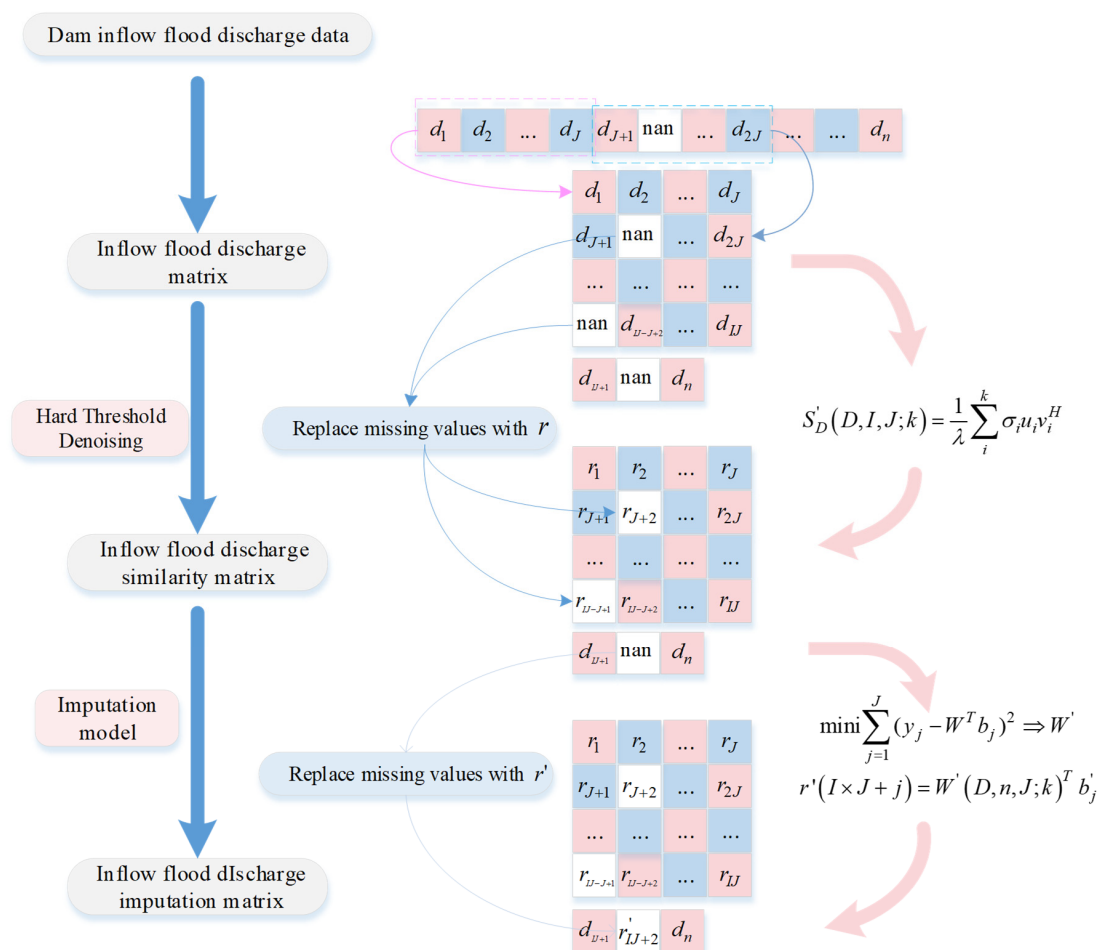


Figure 1. Flowchart of the Dam Inflow Flood Discharge Data Imputation Based on DSVD.

The detailed process is as follows:

(1) Construction of Dam Inflow Flood Discharge Monitoring Matrix

Let the dam inflow flood discharge data be $d = (d_1, d_2, d_3, \dots, d_n)$, with missing values marked as “nan”. The traditional Singular Spectrum Analysis (SSA) matrix [21–23] is arranged with only one element interval per row, which significantly increases the number of rows and columns, and the computational cost of the constructed matrix. For inflow flood discharge, which has large fluctuations and independent data points with weak dependence on previous values, a non-repeating sequence is adopted to construct the inflow flood discharge matrix. That is, the corresponding inflow data is arranged in order, and then entered from left to right, starting with the first row. Once the first row is filled, the second row is entered, and so on, until the matrix is complete. The specific arrangement is as follows (number of columns J ; number of rows I):

$$d = (d_1, d_2, \dots, d_J, d_{J+1}, d_{J+2}, \dots, d_{2J}, d_{2J+1}, \dots, d_{(I-1)J}, d_{(I-1)J+1}, \dots, d_{IJ})$$

$$\Rightarrow \begin{bmatrix} d_1 & d_2 & \dots & d_J \\ d_{J+1} & d_{J+2} & \dots & d_{2J} \\ \dots & \dots & \dots & \dots \\ d_{(I-1)J} & d_{(I-1)J+1} & \dots & d_{IJ} \end{bmatrix} \quad (2)$$

On this basis, the column-to-row ratio R_a (columns J /rows I) is defined to control the matrix dimensions, and the matrix is determined by the R_a as follows:

$$D = \begin{bmatrix} d_1 & d_2 & \dots & d_{n/I} \\ d_{n/I+1} & d_{n/I+2} & \dots & d_{2n/I} \\ \dots & \dots & \dots & \dots \\ d_{(I-1)n/I+1} & d_{(I-1)n/I+2} & \dots & d_n \end{bmatrix} \quad (3)$$

(2) Singular Value Decomposition and Hard Threshold Denoising

Temporarily set “nan” to 0 and perform SVD on the inflow flood discharge matrix:

$$S_D = \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \dots + \sigma_m u_m v_m^H \quad (4)$$

where σ denotes the singular value corresponding to the inflow flood discharge matrix, satisfying $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_m$; u and v are the corresponding left and right singular vectors, respectively.

Given the large fluctuations in dam inflow flood discharge data, a hard singular value threshold method [24] is selected to screen principal components to avoid signal distortion: set a threshold τ , retain singular values greater than τ , and directly set singular values less than τ to 0 to achieve strong noise removal. After threshold processing, a denoised singular value matrix Σ' is obtained, and a similarity matrix R is reconstructed:

$$R = U \Sigma_T V^T \quad (5)$$

(3) Imputation of Missing Dam Inflow Flood Discharge Data

Based on the denoised similarity matrix R , an adaptive imputation model is constructed: the weight matrix W is trained using the first $I - 1$ rows of the matrix, and model parameters are optimized by minimizing imputation error to avoid noise transmission. Using the trained model, imputation values r are generated for “nan” positions within the matrix one by one; for remaining time series missing points outside the matrix, supplementary values r' are generated based on the model’s generalization ability. Finally, all missing values are replaced with r and r' , and complete imputed time series data are outputted, realizing the high-precision recovery of fluctuating inflow flood discharge monitoring data.

The pseudocode for the process is shown in Table 1.

Table 1. The pseudocode for the process.

Input: Inflow data for dams with missing values $d = (d_1, d_2, d_3, \dots, d_n)$
Output: Inflow data for the reservoir after the dam has been filled
1: Start
2: Enter the raw monitoring data for inflow to the dam $d = (d_1, d_2, d_3, \dots, d_n)$
3: The corresponding missing values are marked as “nan”
4: Sorting using unique sequences
$d = (d_1, d_2, \dots, d_J, d_{J+1}, d_{J+2}, \dots, d_{2J}, d_{2J+1}, \dots, d_{(I-1)J}, d_{(I-1)J+1}, \dots, d_{IJ})$ $\Rightarrow \begin{bmatrix} d_1 & d_2 & \dots & d_J \\ d_{J+1} & d_{J+2} & \dots & d_{2J} \\ \dots & \dots & \dots & \dots \\ d_{(I-1)J} & d_{(I-1)J+1} & \dots & d_{IJ} \end{bmatrix}$
5: Constructing Inflow flood discharge matrix
$D = \begin{bmatrix} d_1 & d_2 & \dots & d_{n/I} \\ d_{n/I+1} & d_{n/I+2} & \dots & d_{2n/I} \\ \dots & \dots & \dots & \dots \\ d_{(I-1)n/I+1} & d_{(I-1)n/I+2} & \dots & d_n \end{bmatrix}$
6: Perform SVD $S_D = \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \dots + \sigma_m u_m v_m^H$
7: Perform an HSVT $S'_D(D, I, J; k) = \frac{1}{\lambda} \sum_i^k \sigma_i u_i v_i^H$
8: The reconstruction yields the similarity matrix $R R = U \Sigma_T V^T$
9: Building an adaptive filling model minimize $\sum_{j=1}^I (y_j - W^T b_j)^2 \Rightarrow W'$
10: Train the model
11: Replace the corresponding missing values with reconstructed values
12: Dam inflow data after imputation
13: End

2.3. Theoretical Expansion of the DSVD for Reservoir Inflow Data

The mathematical improvements in DSVD are closely coupled with the physical evolution law of reservoir inflow flood discharge, and the physical connotation of each module is explained through combination with hydrological characteristics:

(1) Non-Repeating Matrix and Flood Fluctuation Characteristics

Mountain reservoir flood discharge presents the physical characteristics of a steep rise and sharp fall, a short flood peak duration, and weak dependence between adjacent time points. The non-repeating matrix avoids over-fitting the weak temporal correlation caused by overlapping data, which physically matches the discrete fluctuation characteristics of mountain floods.

(2) HSVT Denoising and Hydrological Noise Sources

The noise in flood discharge monitoring data mainly comes from environmental interference, instrument jitter and short-term surface runoff fluctuations, which are due to the high-frequency random noise in physical nature. The HSVT module filters out such high-frequency components while retaining the low-frequency main trend of flood evolution (the core physical information of the flood process), which realizes the separation of effective hydrological signals and interference noise, and ensures that the reconstructed data conform to the physical law of basin runoff convergence.

In summary, the DSVD is theoretically applicable to filling the gaps in inflow monitoring data.

2.4. Experiments

2.4.1. Engineering Background

Jinjiaba Reservoir, located in Youyang County, Chongqing City, is situated in the middle reaches of Ganlong River, a primary tributary in the lower reaches of Wujiang River Basin. A schematic diagram of Jinjiaba Reservoir is shown in Figure 2. The Jinjiaba Hydropower Project consists of a concrete face rockfill dam, a bank spillway, diversion tunnels and a power plant. The dam site controls a drainage area of 1059 km², with a normal water level of 445 m corresponding to a storage capacity of 1.519×10^8 m³ and a total reservoir capacity of 1.581×10^8 m³. The dam has a crest elevation of 447 m, a wave wall crest elevation of 448.2 m, a crest width of 10 m, a maximum dam height of 102.5 m and a crest length of 315 m. An L-shaped concrete wave wall with a height of 2.7 m is installed on the dam crest, with its base connected to the concrete face. The upstream and downstream slopes of the face dam are 1:1.4, and the downstream slope adopts a three-tier slope with 3 m wide berms at elevations of 417.30 m and 387.30 m, respectively. The downstream dam surface is protected with dry-laid stone.

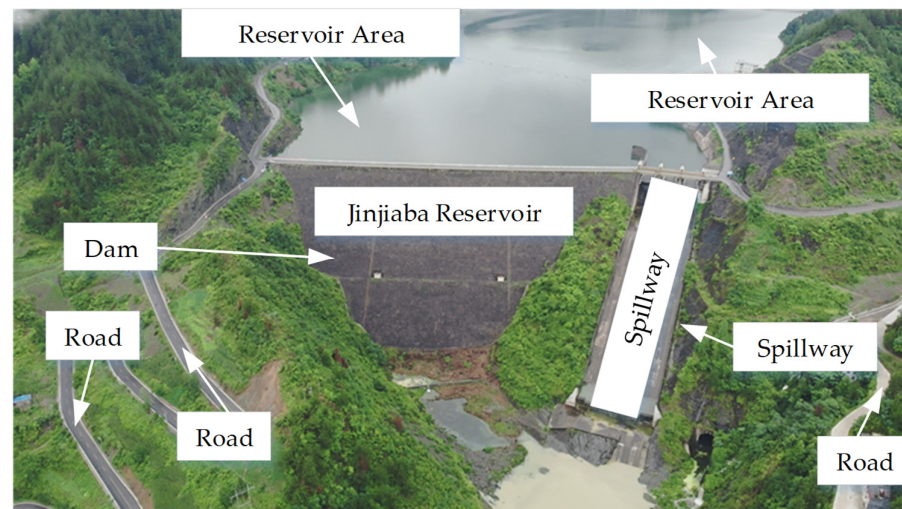


Figure 2. Schematic Diagram of the Jinjiaba Reservoir.

Floods in Ganlong River Basin are mainly caused by rainstorms. According to statistical data from the reference hydrological station (Zhuoheba Station on Apeng River), annual maximum discharge generally occurs from May to September, with rainstorm floods concentrated in June and July. Flood hydrographs are mostly single-peaked, with a typical flood duration of 2–3 days, exhibiting typical mountain river characteristics of steep rise and fall: the rising stage lasts less than 10 h, the flood peak duration is 1–2 h, and the recession stage lasts 1–2 days.

In view of the above regional hydrological rules, we adopted hourly monitoring data for Jinjiaba Reservoir from 00:00 on July 31 to 03:00 on 23 September 2023 for experimental analysis. The trend of inflow flood discharge monitoring data is shown in Figure 2. As can be clearly seen from Figure 3, this time period covers multiple complete flood events that conform to the local flood evolution laws. In terms of data volume, the total number of hourly monitoring records in this sequence reaches 1300, which is sufficient to support the model test and performance analysis. Meanwhile, the dataset also contains typical flow extremes such as flood peaks, which can effectively reflect the actual variation characteristics of local inflow discharge. On the other hand, the measured inflow flood discharge data used in this study includes both surface runoff generated by rainstorms and stable groundwater baseflow, which is consistent with the actual inflow composition of mountainous reservoirs. The model targets the total discharge time series

(containing baseflow) for missing data imputation, meeting the requirements of practical reservoir analysis.

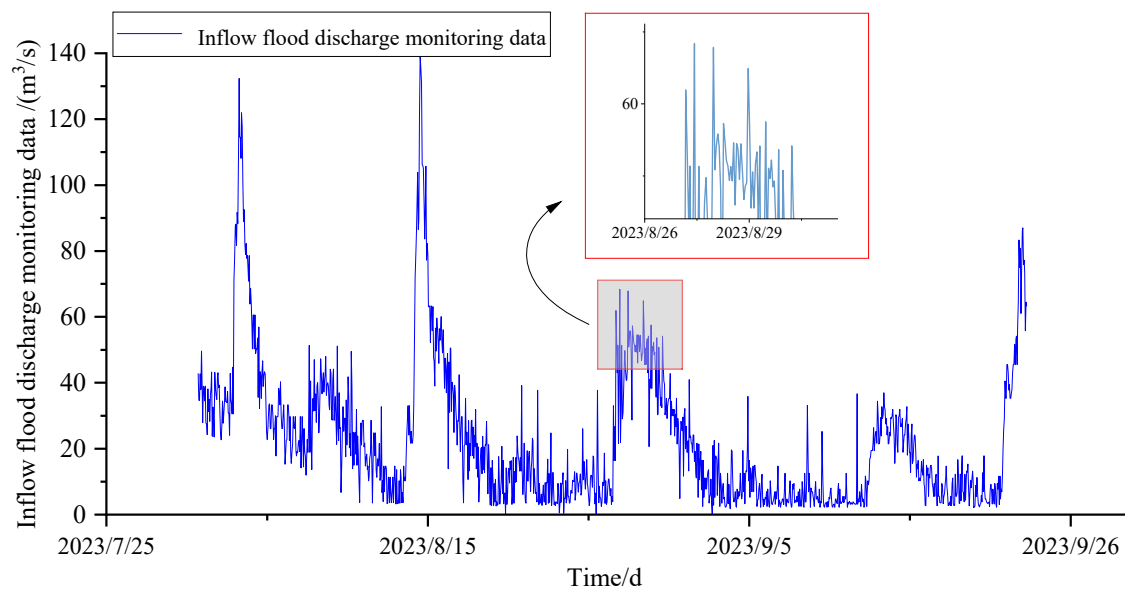


Figure 3. Trend of Inflow Flood Discharge Monitoring Data.

2.4.2. Data Analysis

To further analyze the applicability of the DSVD for monitoring inflow discharge, the inflow discharge monitoring data were analyzed using three methods: RAPS (Regularity Anomaly Partition Statistic), ITA (Inhomogeneity Test Algorithm), and IPTA (Interval Partition Anisotropy Test). The principles of these three methods are as follows:

(1) RAPS for Time Series Irregularity

RAPS is designed to quantify the high-frequency irregular noise of mountain flood discharge time series. The core logic is to separate the original hydrological sequence into three physical components: low-frequency trend component, periodic runoff component and random irregular noise component via moving average filtering.

- ① A 7-point equal-weight moving average filter extracts the varying flood base trend $trend_{ra}$, representing the basin background runoff evolution;
- ② Subtract the trend term from the original flow to obtain the periodic fluctuation term $periodic_{ra}$, reflecting seasonal rainfall–runoff periodicity;
- ③ The residual absolute deviation from the overall mean is defined as the irregular high-frequency component $irregular_{comp}$, mainly originating from instrument jitter, short-term surface runoff turbulence and the environmental interference noise mentioned in Section 2.3 of the original paper;
- ④ The irregular ratio = $std(irregular_{comp})/std(Q)$ is adopted as the quantitative index of RAPS. A threshold of 0.15 is set: when the ratio exceeds 0.15, the sequence contains prominent high-frequency noise.

(2) ITA for Time Series Inconsistency and Mutation

ITA detects abrupt shift points and evaluates the overall inhomogeneity of flood discharge series based on the sliding-window Mann–Whitney U nonparametric test, which adapts to non-stationary hydrological data without normal distribution assumption:

- ① Set a sliding window length $win = 10$; for each middle sampling point, split the sequence into front and back subsegments of equal window length;
- ② Conduct bilateral Mann–Whitney U test for paired subsegments, and record the corresponding p -value of each test position;

③ Positions with $p < 0.05$ are judged as statistically significant mutation points; the inhomogeneity strength index is calculated as the proportion of mutation points in all test windows.

A low inhomogeneity strength (<0.05) indicates the overall flood sequence has no obvious abrupt structural changes.

(3) IPTA for Time Series Anisotropy

IPTA quantifies the temporal anisotropy of mountain flood processes by equal-interval segmentation statistics, aiming to characterize the typical steep-rise–steep-fall feature of mountain river floods:

① Divide the full-length original time series into $seg_num = 5$ equal-length non-overlapping subseries;

② Calculate the mean, standard deviation, maximum and minimum discharge of each subseries separately;

③ Compute the coefficient of variation (CV) of subseries mean values and subseries standard deviations, and take the average of the two CVs as the anisotropy score;

④ Threshold classification: anisotropy score > 0.20 represents strong anisotropy, $0.10\text{--}0.20$ moderate anisotropy, <0.10 weak anisotropy.

Using these three methods to analyze the flood inflow data at Jinjiaba, the steps are as follows:

① Data Preprocessing: Extract hourly reservoir inflow discharge time series Q (total length $n = 1300$) from Jinjiaba Reservoir monitoring data, take complete valid measured samples without artificial missing values as the input original time series;

② Subseries Segmentation Rule for IPTA: Uniformly divide the complete original time series into five equal-length continuous non-overlapping subseries, and the segmentation boundary is evenly distributed along the time axis without manual truncation;

③ RAPS Decomposition Execution: Implement 7-point moving average filtering to separate trend, periodic and irregular noise components, calculate the irregular ratio index, and draw component decomposition curves;

④ ITA Sliding Mutation Detection: Adopt a sliding window width of 10, traverse all valid middle positions of the sequence, output the p -value sequence of the significance test, and count the number of significant mutation points and inhomogeneity strength;

⑤ IPTA Subseries Statistical Calculation: Compute mean and standard deviation of each segmented subseries, calculate the anisotropy score, and plot the standard deviation variation curve of each subseries;

⑥ Comprehensive Discriminant Analysis: Integrate RAPS irregularity, ITA inhomogeneity and IPTA anisotropy indexes to verify the adaptability precondition of the DSVD for mountain flood discharge data.

2.4.3. Influence of Column-to-Row Ratio on Imputation Performance

To analyze the influence of different column-to-row ratios (R_a) on imputation performance, experiments are conducted with inflow flood discharge data of a total length of 1300. The R_a is set to $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ to explore its impact on imputation accuracy.

Taking an R_a of 1 as an example, 40% of the inflow flood discharge monitoring data is randomly removed, and the incomplete dataset is input into the DSVD to construct an inflow flood discharge data matrix. The removed data are imputed using DSVD, and evaluation metrics including Root Mean Square Error (RMSE), Mean Square Error (MSE) and Coefficient of Determination (R^2) are calculated to assess the imputation performance.

Similarly, experiments are conducted on inflow flood discharge data with different column-to-row ratios using DSVD.

2.4.4. Influence of Missing Rate on Imputation Performance

Following the approach outlined in Section 2.4.3, inflow flood discharge monitoring data are randomly removed at different missing rates of {5%, 10%, 20%, 30%, 40%} and imputed using DSVD to analyze the imputation capability under varying missing rates, with the column-to-row ratio set to 6.

2.4.5. Performance Comparison of DSVD with Other Deep Learning Models

To validate the scientificity and advantages of the DSVD, inflow flood discharge data with a 40% missing rate are input into SVD, DTRN, GAIN, DMDRN and TGAIn for comparative analysis.

To further evaluate the corresponding interpolation results, the peak absolute error is introduced as one of the evaluation metrics. Its calculation formula is shown below:

$$E_{peak} = |Q_{imputed} - Q_{ture}| \quad (6)$$

where E_{peak} represents the peak discharge absolute error, $Q_{imputed}$ represents the peak discharge extracted from the imputed time series, and Q_{ture} represents the real flood peak discharge of the original complete monitoring sequence.

3. Results

3.1. Data Analysis Results

Following the process in Section 2.4.2, the data analysis results are shown in Figure 4.

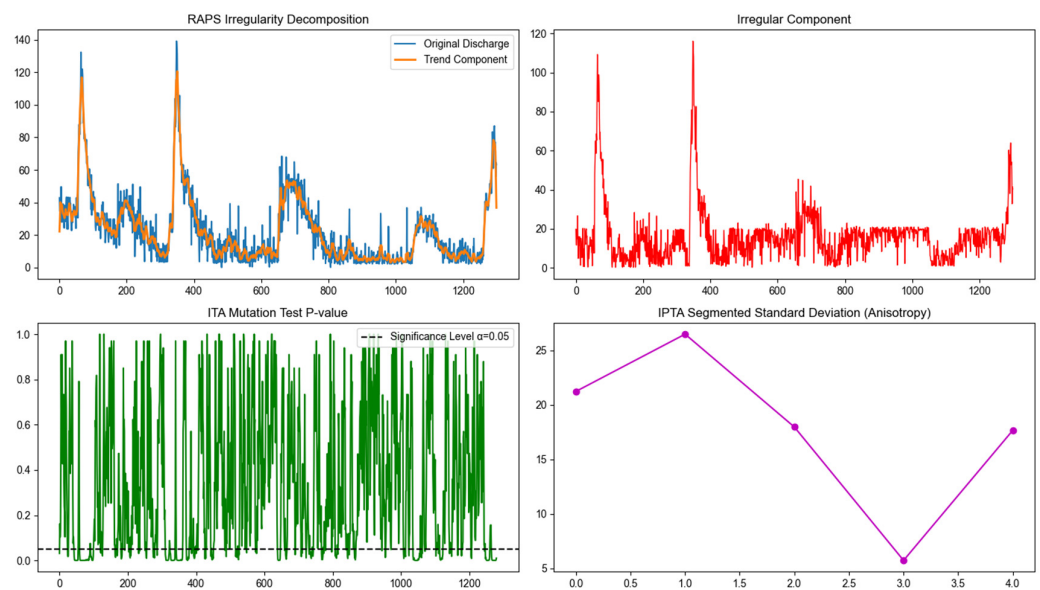


Figure 4. Results of irregularities, inhomogeneity, and anisotropy.

RAPS Test for Irregularity: The ratio of irregular components reaches 0.664, which indicates that the discharge time series contains remarkable high-frequency irregular fluctuations and noise. This is consistent with the actual characteristics of mountain river floods affected by rainstorms and complex terrain.

ITA Test for Inhomogeneity: The inhomogeneity intensity of the series is 0.252, with 322 significant mutation points detected. This proves that the hydrological sequence presents certain inhomogeneity during the study period.

IPTA Test for Anisotropy: The anisotropy index is 0.409, demonstrating that the flood series has strong anisotropy, which is a typical feature of mountainous rivers with rapidly rising and receding flood processes.

The above statistical test results interpret the inherent characteristics of the research data well. The following conclusions also fully explain the rationality and superiority of the DSVD method.

Aiming at the prominent high-frequency noise verified by RAPS, the hard singular-value threshold (HSVT) denoising module of DSVD is specially designed to filter irregular noise and retain valid hydrological signals.

Faced with the inhomogeneity of the time series, the optimized non-repeating sequence matrix structure enables the model to extract local features adaptively and reduce the adverse impact of data inhomogeneity on imputation accuracy.

3.2. Influence of Column-to-Row Ratio on Imputation Performance Results

The influence of Column-to-Row Ratio on imputation performance results is shown in Figure 5, and the evaluation metrics are summarized in Table 2.

Table 2. Imputation Performance with Different R_a .

R_a	R^2	MSE	RMSE
1	0.851	72.043	8.488
2	0.857	69.492	8.336
3	0.858	68.751	8.292
4	0.863	66.177	8.135
5	0.856	69.622	8.344
6	0.875	60.393	7.771
7	0.854	70.864	8.418
8	0.865	65.452	8.090
9	0.865	65.452	8.090
10	0.850	72.757	8.530

As shown in Table 2, when the column-to-row ratio ranges from 1 to 10, the R^2 of DSVD remains above 0.850, meeting the imputation requirements for inflow flood discharge data and confirming the good imputation performance of DSVD. The optimal imputation performance is achieved at a column-to-row ratio of 6, with the highest R^2 . Thus, a column-to-row ratio of 6 is adopted in subsequent experiments. Notably, the optimal column-to-row ratio of 6 may be specific to the experimental inflow flood discharge data; the optimal value may vary for different data types. Therefore, pre-experiments are recommended to determine the optimal R_a in practical inflow flood discharge monitoring data imputation. In the absence of pre-experiments, a column-to-row ratio of 6 is recommended based on this study.

3.3. Influence of Missing Rate on Imputation Performance Results

The influence of missing rate on imputation performance results is shown in Figure 6, and the evaluation metrics are summarized in Table 3.

Table 3. Imputation Performance of DSVD under Different Missing Rates.

Missing Rates	R^2	MSE	RMSE
5%	0.930	37.986	6.163
10%	0.872	37.131	6.093
20%	0.836	58.749	7.665
30%	0.882	50.688	7.120
40%	0.875	60.393	7.771

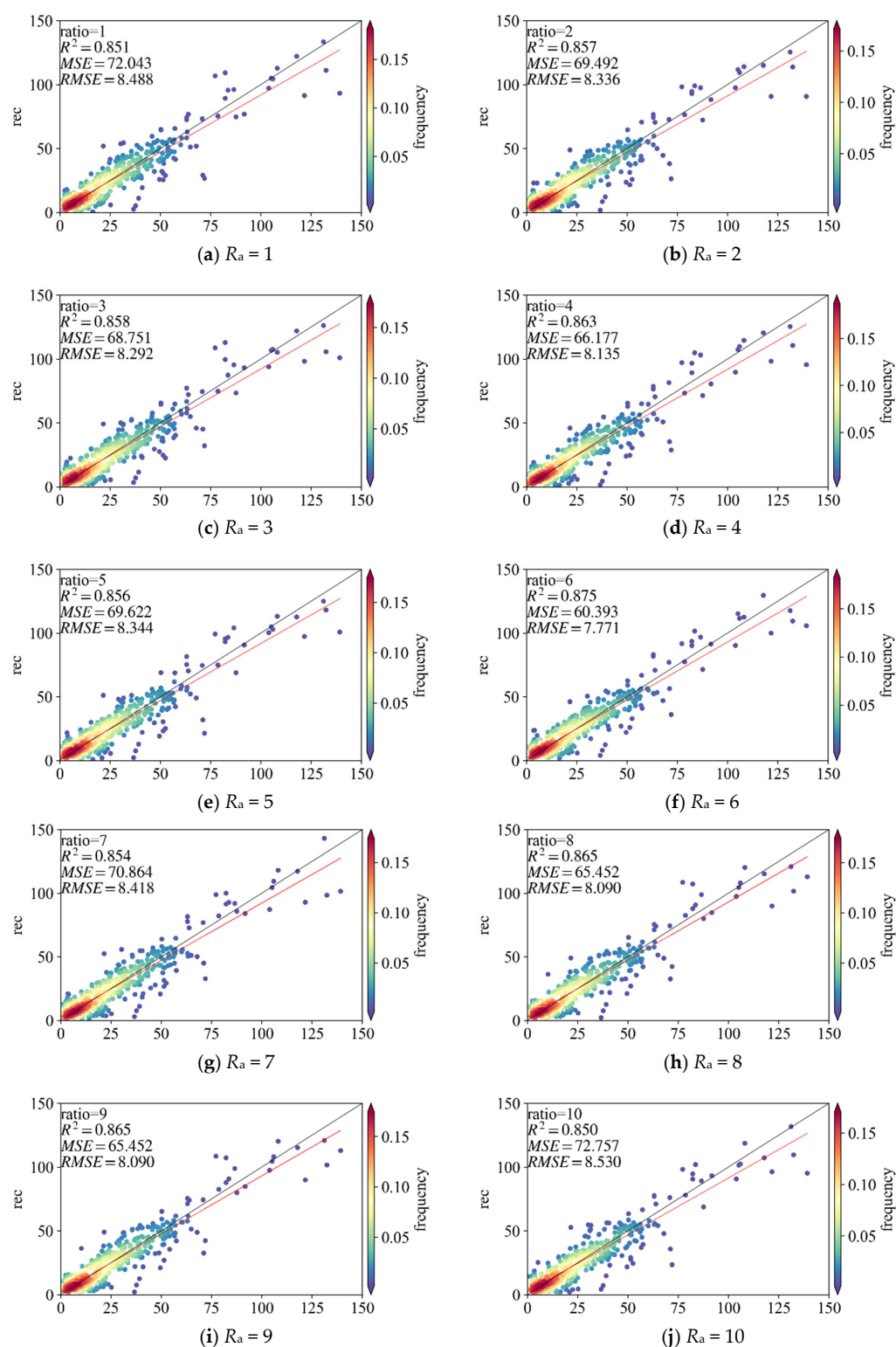


Figure 5. Imputation Results with Different R_a .

As shown in Table 3, when the missing rate of inflow flood discharge monitoring data increases from 5% to 40%, the R^2 remains above 0.830, verifying the good imputation performance of DSVD under varying missing rates for dam inflow flood discharge data.

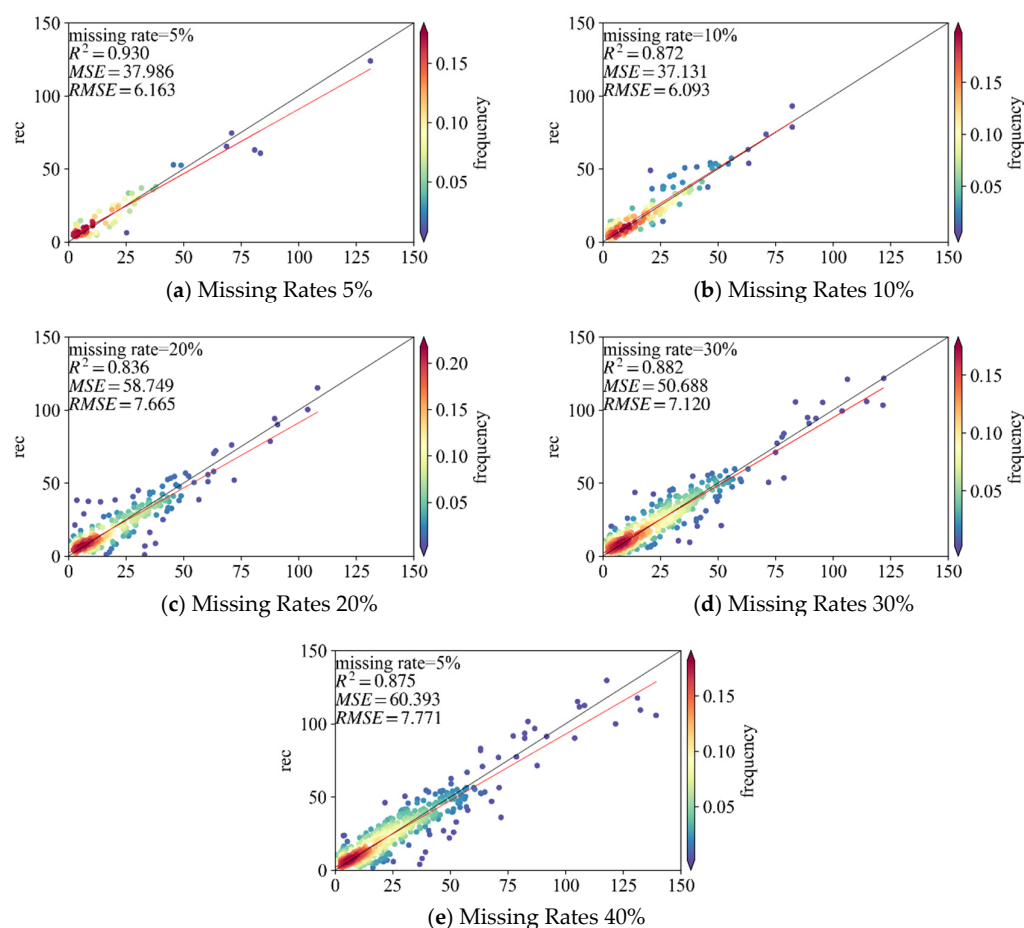


Figure 6. Imputation Results of DSVD under Different Missing Rates.

3.4. Performance Comparison of DSVD with Other Deep Learning Models Results

A performance comparison of DSVD with other deep learning models' results are shown in Figure 8, the corresponding time-series comparison chart is shown in Figure 8, and the evaluation metrics are summarized in Table 4.

Table 4. Imputation Performance of Different Models.

Models	R^2	MSE	RMSE	Peak Discharge Error
DSVD	0.875	60.393	7.771	33.483
SVD	0.828	83.436	9.134	50.130
DTRN	0.661	164.399	12.822	81.222
GAIN	0.569	208.596	14.443	60.724
DMDRN	0.163	405.188	20.129	45.575
TGAIN	0.727	132.401	11.507	88.493

As shown in Figure 8, the trend line interpolated using DSVD is closer to the original data, indicating that DSVD is more effective for interpolating dam inflow monitoring data. On the other hand, as shown in Table 4, DTRN exhibits poor performance with an R^2 of 0.661 for inflow flood discharge imputation. This is attributed to the strong fluctuations in inflow flood discharge monitoring data, limiting DTRN's ability to extract temporal features. The DSVD achieves the highest R^2 of 0.875, the lowest RMSE, and the lowest peak discharge error among all models, confirming its effectiveness and superiority in dam inflow flood discharge data imputation.

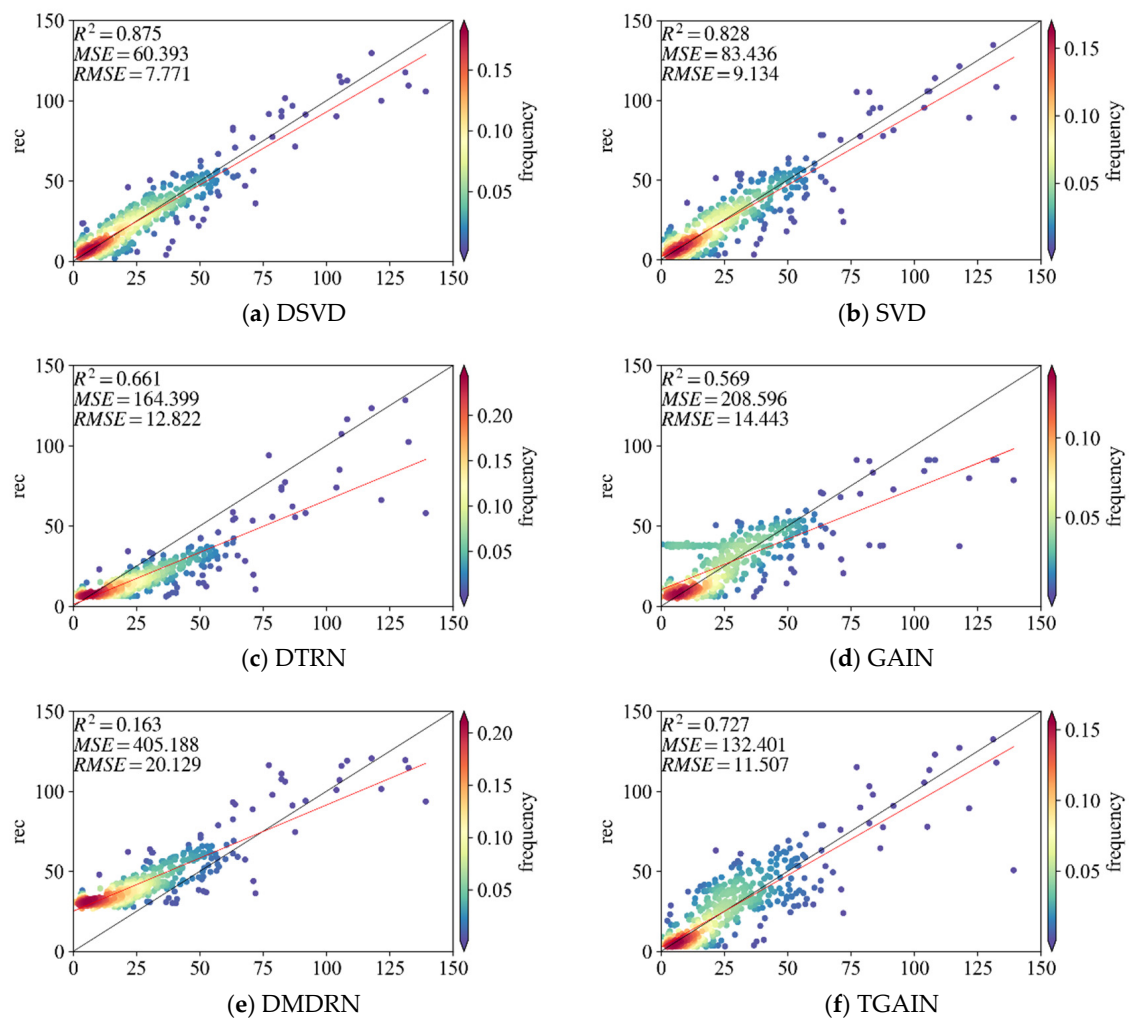


Figure 7. Imputation Results of Different Models.

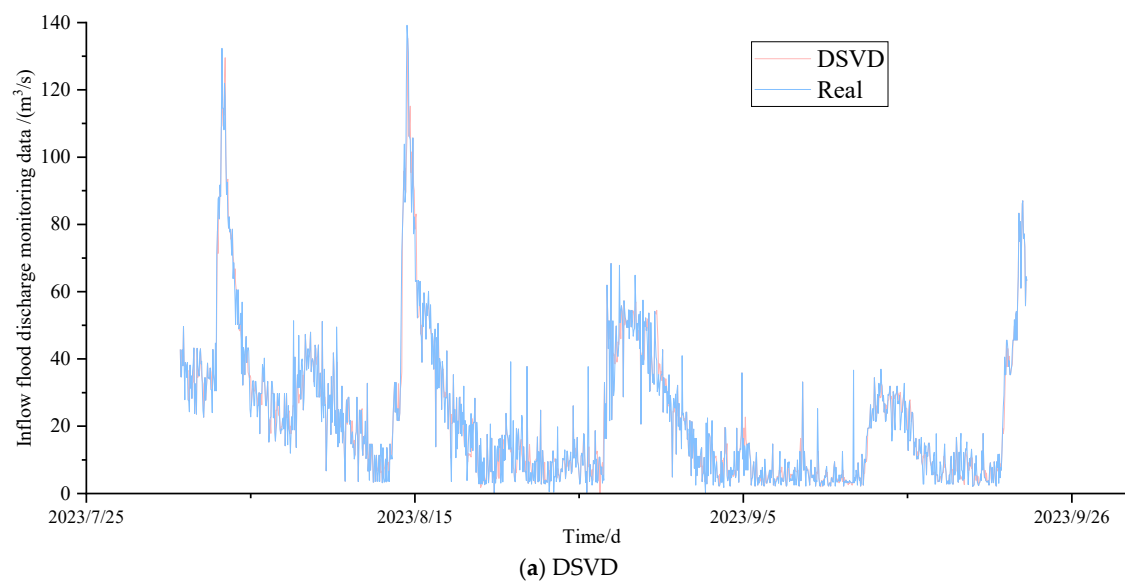
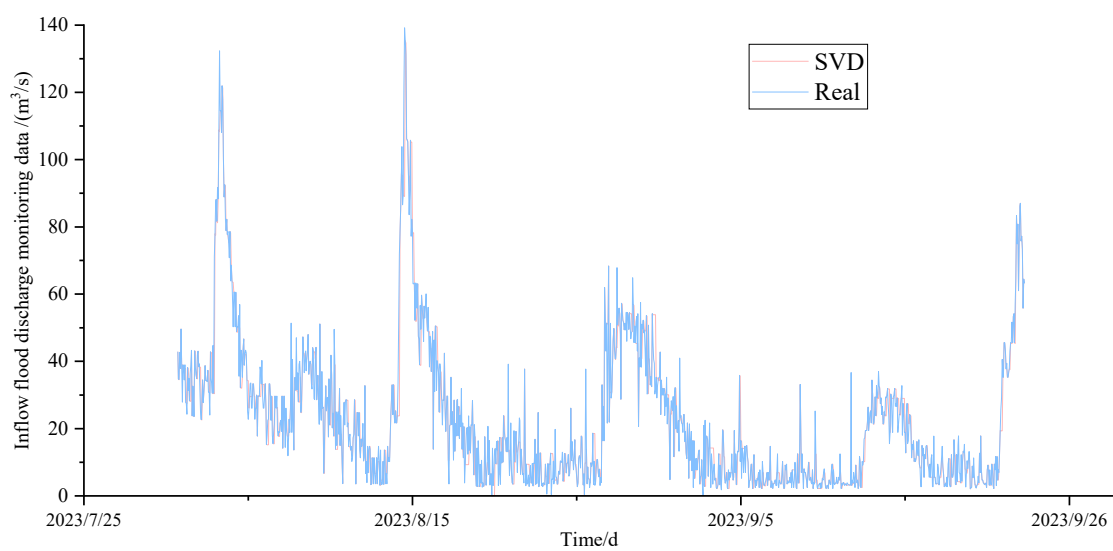
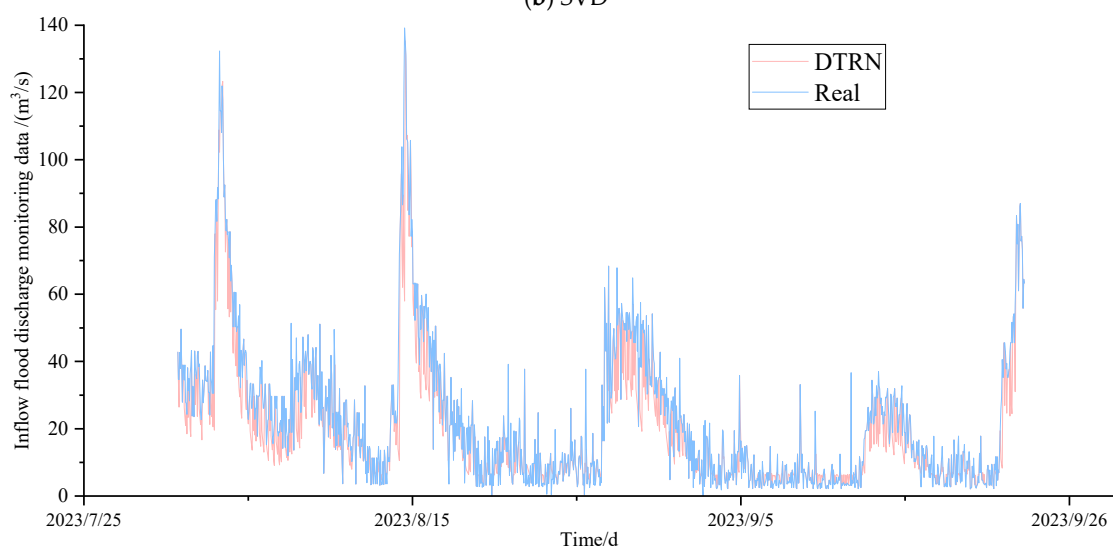


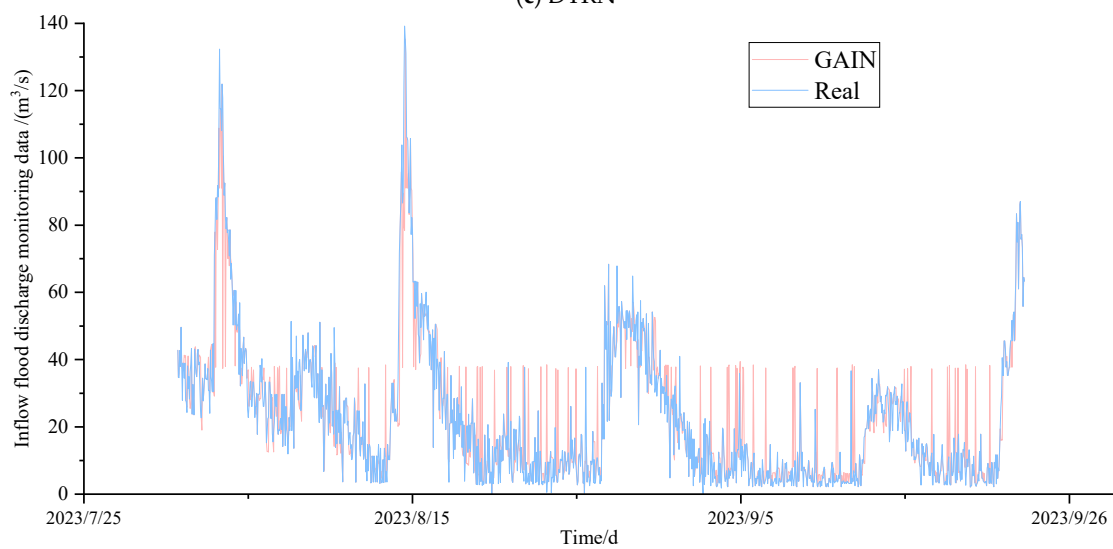
Figure 8. Cont.



(b) SVD



(c) DTRN



(d) GAIN

Figure 8. Cont.

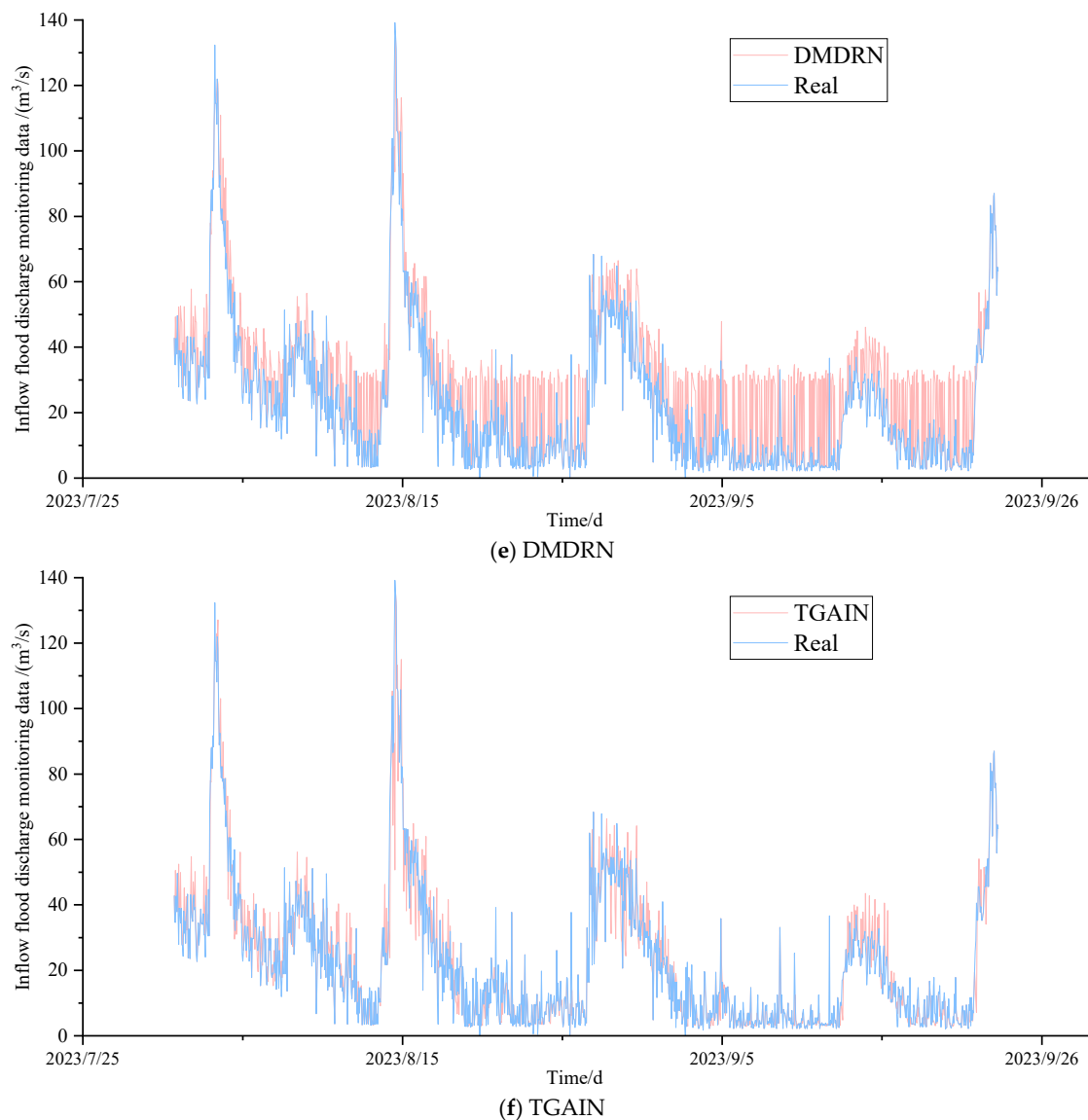


Figure 8. The time-series comparison chart.

4. Discussion

The DSVD effectively addresses the limitations of traditional SVD (insufficient imputation accuracy and difficult threshold determination for strongly fluctuating and high-noise dam inflow flood discharge data) through matrix structure optimization, adaptive hard threshold denoising and weight optimization imputation, demonstrating significant advantages in validation with measured data from Jinjiaba Reservoir. Parameter sensitivity analysis reveals that the column-to-row ratio is a critical parameter affecting DSVD performance. Within the range of 1–10, the optimal performance (highest R^2 ; lowest MSE and RMSE) is achieved at a column-to-row ratio of 6, which aligns with the temporal characteristics of Jinjiaba Reservoir flood discharge (steep rise and fall, and predominantly single-peaked hydrographs). A small column-to-row ratio results in insufficient matrix dimensions and inadequate feature extraction, while an excessively large ratio introduces redundant information and noise interference, reducing imputation accuracy.

Experiments under varying missing rates confirm the strong robustness of DSVD to missing data. The R^2 remains above 0.830 for missing rates of 5–40%, enabling the effective capture of temporal variations in flood discharge even under high missing rates. This is

attributed to the hard threshold denoising mechanism for noise suppression and the weight optimization model for temporal correlation fitting.

A comparative analysis with SVD, DTRN, GAIN, DMDRN and TGAIN demonstrates the superior imputation accuracy of DSVD. Traditional SVD has moderate imputation capability but higher error than DSVD due to the unoptimized matrix structure and threshold selection. Among deep learning models, DTRN, GAIN and DMDRN exhibit poor performance due to inadequate network structure adaptability for extracting the temporal features of strongly fluctuating flood discharge data. TGAIN outperforms the above three models but still underperforms compared to DSVD, primarily due to the lack of a targeted noise removal strategy and susceptibility to high-frequency noise interference in flood discharge data. Collectively, the DSVD integrates the stability of traditional matrix decomposition methods and the temporal fitting capability of intelligent models, exhibiting excellent adaptability to non-stationary, strongly fluctuating and high-noise data such as mountainous dam inflow flood discharge.

Despite the promising performance of the DSVD, this study has certain limitations: first, it should be noted that all missing datasets used for model validation in this paper are generated by fully random deletion at given missing ratios. Although this setting facilitates the controlled analysis of the sensitivity of column-to-row ratio and missing rate, it cannot reproduce the temporally clustered and consecutive missing segments commonly encountered in actual hydrological monitoring. In practical operation, sensor faults, communication breakdowns, regular equipment maintenance and extreme weather will result in continuous long-period data gaps, which are more challenging for time series imputation; second, the optimal column-to-row ratio is determined via pre-experiments, and future research may integrate intelligent optimization algorithms (e.g., genetic algorithm, particle swarm optimization) for automatic parameter optimization to enhance model automation; third, this study focuses solely on single-variable (flood discharge) missing data imputation, while practical hydrological monitoring data often involve coupled multi-variable missing issues. Future research may construct multi-variable joint imputation models to further improve data recovery capability in complex scenarios.

Future research directions include: first, establishing multiple real-world missing scenarios with different lengths of consecutive missing periods and hybrid missing modes to comprehensively test the robustness of the DSVD, so as to further validate its practical application value for actual reservoir monitoring data reconstruction; second, integrating optimization algorithms to realize the adaptive selection of key model parameters and reduce manual intervention; third, extending the model application scope by constructing a multi-variable collaborative missing data imputation framework combining multi-source monitoring data such as water level and rainfall, to provide technical support for improving the quality of full-factor hydrological monitoring data in reservoirs.

5. Conclusions

Aiming at the problems of strong fluctuation, high noise, an uncertain missing rate and insufficient accuracy of traditional imputation methods for dam inflow flood discharge data, this paper introduces DSVD for dam inflow flood discharge imputation. Experimental validation with measured data from Jinjiaba Reservoir yielded the following main conclusions:

(1) A non-repeating sequence is adopted to construct the inflow flood discharge matrix to adapt to the independent fluctuation characteristics of flood discharge; an HSVT is introduced to adaptively remove strong noise; a weight optimization model is constructed to suppress noise transmission, effectively improving the imputation accuracy of missing dam inflow flood discharge data.

(2) The column-to-row ratio is a critical parameter of the DSVD. Within the range of 1–10, the optimal imputation performance is achieved at a column-to-row ratio of 6, with $R^2 = 0.875$, $MSE = 60.393$ and $RMSE = 7.771$. The optimal column-to-row ratio for different data types can be determined via pre-experiments, with a recommended value of 6 in the absence of pre-experiments.

(3) The DSVD exhibits strong robustness to missing data. Under different missing rates ranging from 5% to 40%, the R^2 of the imputation results remains above 0.830, effectively adapting to various missing scenarios of dam inflow flood discharge data and meeting the data accuracy requirements for flood process analysis and risk assessment.

(4) Compared with SVD, DTRN, GAIN, DMDRN and TGAIN, the DSVD achieves the optimal imputation accuracy. At a 40% missing rate, its R^2 is 0.047 higher than that of traditional SVD and 0.148 higher than that of the best-performing deep learning model (TGAIN), with a significantly lower RMSE than all benchmark models. It exhibits stronger adaptability to mountainous flood discharge data with steep rise and fall characteristics.

In summary, the DSVD features a concise structure, adaptive parameters and high imputation accuracy, providing an effective technical means for the high-precision recovery of missing dam inflow flood discharge data. It offers reliable data support for reservoir flood overtopping risk analysis, flood control operation decision-making and safe dam operation, with good engineering application value and promotion prospects.

Author Contributions: Conceptualization, Y.C. and K.W.; Methodology, Y.C. and G.L.; Software, G.L.; Validation, Y.C.; Formal analysis, G.L.; Investigation, K.W. and J.L.; Data curation, G.L. and J.L.; Writing—original draft, Y.C. and G.L.; Writing—review and editing, Y.C., M.Z. and J.L.; Visualization, M.Z. and J.L.; Supervision, K.W., M.Z. and J.L.; Project administration, K.W. and M.Z.; Funding acquisition, K.W. and M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China, grant number 2024YFB2605405.

Data Availability Statement: The processed dataset supporting the findings of this study is available from the corresponding author upon reasonable request.

Acknowledgments: The touch-ups to this paper by the Editors are greatly acknowledged.

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

Abbreviations

The following abbreviations are used in this manuscript:

SVD	Singular-Value Decomposition
DSVD	Dam Monitoring Data Reconstruction Model
HSVT	Hard Singular-Value Threshold
RMSE	Root Mean Square Error
GAIN	Generative Adversarial Imputation Nets
TGAIN	Timing Generative Adversarial Imputation Nets
DTRN	Dam Temporal Reconstruction Nets
DMDRN	Dam Monitoring Data Reconstruction Network

References

1. Babaei, M.; Moeini, R.; Ehsanzadeh, E. Artificial Neural Network and Support Vector Machine Models for Inflow Prediction of Dam Reservoir (Case Study: Zayandehroud Dam Reservoir). *Water Resour. Manag.* **2019**, *33*, 2203–2218. [[CrossRef](#)]
2. Cho, E.; Ahmadisharaf, E.; Ahmadisharaf, A.; Nematirad, R.; Aghakouchak, A. Unraveling the Relationships between Trend of Dam Inflows, Hydrometeorological Variables, and Vegetation in Western and Southwestern United States. *J. Hydrometeorol.* **2024**, *25*, 1793–1808. [[CrossRef](#)]

3. Lee, E.H. Proactive dam operation based on inflow prediction by modified long short-term memory for improving resilience. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108525. [\[CrossRef\]](#)
4. Goodarzi, E.; Lee, T.S.; Ziaei, M. Risk and uncertainty analysis for dam overtopping-Case study: The Doroudzan Dam, Iran. *J. Hydro-Environ. Res.* **2014**, *8*, 50–61. [\[CrossRef\]](#)
5. Zhao, K.; Deng, Z.; Chen, S.; Zhong, Q.; Chao, Y.; Jiang, J. Effects of flow velocity and concentration on the overtopping failure mechanism of tailings dams. *Environ. Earth Sci.* **2025**, *84*, 480. [\[CrossRef\]](#)
6. Micovic, Z.; Hartford, D.N.D.; Schaefer, M.G.; Barker, B.L. A non-traditional approach to the analysis of flood hazard for dams. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 559–581. [\[CrossRef\]](#)
7. Yoon, J.; Jordon, J.; van der Schaar, M. GAIN: Missing Data Imputation using Generative Adversarial Nets. In *International Conference on Machine Learning*; Dy, J., Krause, A., Eds.; Jmlr-Journal Machine Learning Research: San Diego, CA, USA, 2018; Volume 80.
8. Li, H. Research on Traffic Flow Operation Risk Evaluation Based on Deep Learning. Ph.D. Thesis, Jilin University, Changchun, China, 2021. [\[CrossRef\]](#)
9. Chen, Y.; Wang, K.; Zhao, M.; Liu, J. Network models for temporal data reconstruction for dam health monitoring. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 2010–2031. [\[CrossRef\]](#)
10. Chen, Y.; Wang, K.; Zhao, M.; Xiong, Y.; Li, C.; Liu, J. Identification and reconstruction of anomalous data in dam monitoring considering temporal correlation. *Smart Mater. Struct.* **2023**, *32*, 115009. [\[CrossRef\]](#)
11. Yi, S.; Sneeuw, N. Filling the Data Gaps Within GRACE Missions Using Singular Spectrum Analysis. *JGR Solid Earth* **2021**, *126*, e2020JB021227. [\[CrossRef\]](#)
12. Drmac, Z. A QR-Preconditioned QR SVD Method for Computing the SVD with High Accuracy. *ACM Trans. Math. Softw.* **2017**, *44*, 11. [\[CrossRef\]](#)
13. Liu, Y.; Wu, F.; Yu, B.; Li, C. An incremental randomized algorithm for singular value decomposition of streaming data matrices. *Appl. Math. Lett.* **2026**, *174*, 109822. [\[CrossRef\]](#)
14. Wang, C.; Zhao, X.L.; Zheng, Y.B.; Li, B.Z.; Ng, M.K. Functional Tensor Singular Value Decomposition. *SIAM J. Sci. Comput.* **2025**, *47*, A2180–A2204. [\[CrossRef\]](#)
15. Kanchi, R.S.; He, S. Differentiable singular value decomposition (SVD). *Mech. Syst. Signal Proc.* **2025**, *237*, 112817. [\[CrossRef\]](#)
16. Weiss, S.; Proudler, I.K.; Barbarino, G.; Pestana, J.; Mcwhirter, J.G. On Properties and Structure of the Analytic Singular Value Decomposition. *IEEE Trans. Signal Process.* **2024**, *72*, 2260–2275. [\[CrossRef\]](#)
17. Liu, Y.; Wu, F.; Miao, J.; Li, C. A subspace-orbit randomized algorithm for quaternion tensor singular value decomposition based on Qt-product. *Appl. Math. Comput.* **2026**, *512*, 129780. [\[CrossRef\]](#)
18. Chen, Y.; Wang, K.; Zhao, M.; Liu, J.; Cheng, Y. A reconstruction method for dam monitoring data based on improved singular value decomposition. *Mech. Syst. Signal Process.* **2025**, *224*, 112217. [\[CrossRef\]](#)
19. Cui, L.B.; Hu, W.L.; Yuan, J.Y. Iterative refinement method by higher-order singular value decomposition for solving multi-linear systems. *Appl. Math. Lett.* **2023**, *146*, 108819. [\[CrossRef\]](#)
20. Agarwal, A.; Alomar, A.; Shah, D. On Multivariate Singular Spectrum Analysis and Its Variants. *ACM SIGMETRICS Perform. Eval. Rev.* **2022**, *50*, 79–80. [\[CrossRef\]](#)
21. Xu, X.; Zhang, M.; Luo, M.; Yang, J.; Qu, Q.; Tan, Z.; Yang, H. Echo Signal Extraction Based on Improved Singular Spectrum Analysis and Compressed Sensing in Wavelet Domain. *IEEE Access* **2019**, *7*, 67402–67412. [\[CrossRef\]](#)
22. Zhang, H.; Lu, W.; Wei, J.; Huang, X.; Yang, X.; Lu, X. Efficient Singular Spectrum Mode Ensemble for Extracting Wide-Band Components in Overlapping Spectral Environments. *IEEE Trans. Signal Process.* **2024**, *72*, 4666–4681. [\[CrossRef\]](#)
23. Bozzo, E.; Carniel, R.; Fasino, D. Relationship between Singular Spectrum Analysis and Fourier analysis: Theory and application to the monitoring of volcanic activity. *Comput. Math. Appl.* **2010**, *60*, 812–820. [\[CrossRef\]](#)
24. Gavish, M.; Donoho, D.L. The Optimal Hard Threshold for Singular Values is $4/\sqrt{3}$. *IEEE Trans. Inf. Theory* **2014**, *60*, 5040–5053. [\[CrossRef\]](#)

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.