Article

Time Series Prediction Model of Landslide Displacement Using Mean-Based Low-Rank Autoregressive Tensor Completion

Chenhui Wang [1,2,]* and Yijiu Zhao [1]

1 School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
2 Center for Hydrogeology and Environmental Geology Survey, China Geological Survey, Baoding 071051, China
* Correspondence: wangchenhui@mail.cgs.gov.cn

Abstract: Landslide displacement prediction is a challenging research task that can help to reduce the occurrence of landslide disasters. The frequent occurrence of extreme weather increases the probability of landslides, and the subsequent increase in the superimposed economic development level exacerbates disaster losses, emphasizing the importance of landslide prediction. The collection of landslide monitoring data is the foundation of landslide displacement prediction, but the lack of various data severely limits the effectiveness of the landslide monitoring system. To address the issue of missing data during the landslide monitoring process, this paper proposes a time series prediction model of landslide displacement using mean-based low-rank autoregressive tensor completion (MLATC). Firstly, the reasons for the missing data of landslide displacement are analyzed, and the corresponding dataset of missing data is designed. Then, according to the characteristics and internal correlation of landslide displacement monitoring data, the establishment process of mean-based low-rank tensor completion prediction model is introduced. Finally, the proposed method is used to complete and predict the missing data for the random missing and non-random missing landslide displacement. The results show that the data completion and prediction results of the model are essentially consistent with the original displacement monitoring data of the landslide, and the accuracy and precision are relatively high. It shows that the model has good landslide displacement completion and prediction effects, which can provide a certain reference value for the missing data processing and landslide displacement prediction.

Keywords: time series; missing data; tensor completion; autoregressive norm; displacement prediction

1. Introduction

Landslides are one of the many common natural disasters in the world [1,2]. A frequent occurrence of extreme weather increases the likelihood of landslides. Rapid economic and social development further aggravates the loss of landslide disasters [3,4]. Therefore, accurate landslide displacement prediction is becoming increasingly important in order to prevent and mitigate the damage caused by landslides [5]. In the process of automated real-time monitoring of landslides, data acquisition and transmission are generally performed using different types of sensors and other electronic devices [6]. Automated monitoring equipment is always in the open-air environment, and most of them inevitably suffer from tear, aging, power loss and other phenomena, all of which can lead to missing monitoring data. In addition, most landslide disasters are located in a relatively harsh geological environment, such as heavy rainfall, hail, dense fog, electromagnetic interference, etc., and the installation and deployment of geological hazard monitoring equipment in open fields will inevitably be affected by the abovementioned harsh environment. Randomness or prolonged interruptions in the operation of the monitoring device can cause the monitoring device to fail to properly send monitoring data to the server, which leads to the problem of missing monitoring time series in the server. Time series forecasting is a valid basis for

making accurate discriminations. To enhance the accuracy of landslide prediction, it is necessary to construct a corresponding accurate data completion method.

For the problem of completing and predicting missing landslide data, most traditional time series models have focused on models such as regression analysis and exponential smoothing. The problem of incompleteness for missing time series data can be broadly classified as either deletion or padding. Data deletion is used as anomalous data to remove some objective abnormal monitoring data, which is primarily used for anomaly detection and feature analysis, while filling is utilized to find the long-term time series change pattern of monitoring data and to supplement the missing monitoring data. The main methods include missing value filling algorithm based on nearest neighbor method, cyclic neural network, random forest and matrix decomposition, but more data are required for machine learning training [7,8]. Statistical filling is more effective for data series with less dimensions that can establish a maximum model, provided that the relationship between missing eigenvalues and existing eigenvalues can be established through observation. The main methods of machine learning include missing value filling algorithm based on the nearest neighbor method, cyclic neural network, and matrix decomposition. Matrix decomposition can effectively explore the correlation between different time series for different dimensions of long time series problems. The matrix decomposition method is used to learn the overall characteristics of the time series matrix, which can be used to approximate the matrix with time characteristics in low-rank, and then complete the missing data [9].

Large-scale time series data are always accompanied by the missing problem. Therefore, the tensor completion method has been introduced into this field to complement the traditional data completion method based on probability and statistics [10–12]. The data completion scheme based on simple quantitative statistics has a relatively simple and efficient processing effect for small datasets and simple regression models, but it is not feasible for massive data in the era of data explosion. Modern research not only has many kinds of data variables and long time series, but also requires fast processing speed, high universality and portability. Multiple variables and long time series can better describe the complex causal relationship between each other [13,14]. Now, neural network technology is often used to deal with the above situations and delete the missing data in order to form a complex intelligent model, but it is not the best solution for missing data because deletion may strengthen or weaken the connection degree of a causal relationship. Based on this, scholars have explored the application of tensor decomposition technology to data completion in multivariate long time series, trying to improve the resolution speed and data missing problems.

Recent studies have found that low-rank matrices have certain advantages in the analysis of multivariate long time series data [11,12], including the sequence tensor completion method [15]. The sequence tensor complement restores the potential tensor from the sampling structure of the time series, allocates the position of the missing items as needed, seamlessly integrates the future value of the time series into the framework of the missing data and improves the data completion accuracy [15]. The low-rank matrix completion method performs singular spectrum analysis and singular value decomposition on the time series in order to complete the low-rank completion of the missing data of the time series, although its calculation is large [16]. Therefore, by adding a time dimension to transform the status time series into a high-dimensional tensor, the cost of computing complexity is better solved. This is also in line with the law of human activities, both short and long term activities, so there are studies using tensors (sensors $\times$ 1 day $\times$ 24 h) which indicate the above activity mode [17,18]. The dependency between sensors is preserved, providing a new feasible scheme for capturing local and global time patterns [16]. More scholars have combined the autoregressive moving average model with the tensor model to propose the low-rank matrix autoregressive tensor completion model and have achieved good results in the completion and prediction of financial time series data [19].

The mean-based low-rank autoregressive tensor completion mainly includes completing low-rank matrix decomposition/tensor completion and constructing time series autoregressive models, as well as processing the missing data with the neighboring data mean instead of zero before the operation. The low-rank matrix completion model uses the underlying low-rank structure to recover incomplete matrices (assuming that the long-term landslide data sequence is incomplete) [20]. Considering that the deformation displacement of landslide has a great correlation with the previous deformation, the autoregressive model is constructed to represent the deformation law of landslide displacement with time. The autoregressive regularizer is introduced in the low-rank matrix decomposition to characterize the temporal dynamics in landslide displacement deformation, and the learned autoregressive regularizer is implemented to predict the temporal factor matrix, thus realizing the landslide displacement monitoring data completion and predictive modeling [12].

The purpose of this study is to establish a new method for completing and predicting landslide displacement data based on MLATC. In this paper, the causes of data loss of landslide displacement are analyzed. Taking the Shuizhuyuan landslide in the Three Gorges Reservoir area as an example, the data completion and prediction algorithm are designed by using MLATC. Then, the landslide displacement data are divided into training set and test set, and the random missing and non-random missing are selected for corresponding data completion and prediction. The designed model can achieve an accurate completion and prediction of landslide displacement. Finally, a comparative analysis with existing models verifies the effectiveness of the model.

2. Theory and Method

2.1. Reasons and Analysis of Missing Data

2.1.1. Reasons of Missing Data

The reasons for the missing of landslide monitoring data are complex, which may be caused by the process of data collection, transmission, storage and analysis. Time series with missing data typically share the same characteristics, such as noisy, incomplete and abnormally abrupt data in the time series. Missing data in the time series can have a significant impact on future landslide data analysis, monitoring and early warning. The data should thus be preprocessed in the process of data analysis, i.e., data cleaning, and the processing of missing data is one of the key elements in data cleaning.

Data deficiencies in landslide monitoring time series are broadly characterized as follows. (1) Long time span and large amount of data: Landslide monitoring requires the data collection of the different factors affecting landslide displacement and deformation, and the continuous extension of monitoring time leads to the increasing amount of data. (2) Randomness: The data acquisition and transmission in the landslide monitoring system are completed automatically by the field equipment, and the components and modules used for acquisition and transmission are all electronic components. It is highly susceptible to rainfall, hail, electromagnetic interference and other natural factors, and the above natural phenomena have randomness. (3) Spatial correlation: The mechanism of landslide disasters is complex. Corresponding monitoring devices will be deployed in different areas of the landslide, and there will be some correlation between different sensors. Therefore, missing data completion needs to consider the spatial correlation among the sensors in the time series.

Due to the diverse properties of the landslide time series and the different types of monitoring data, there are missing data in the time series for a variety of causes, which can be roughly summarized as follows. (1) Data are not available: The landslide monitoring process is affected by the harsh natural environment. Problems with the power supply of the equipment and the transient failure of the sensors can lead to a certain period of time or a certain moment of failure to complete the data acquisition, thus leading to the lack of data. (2) Data transmission failure: The transmission of landslide monitoring data mainly relies on a wireless communication network; when the wireless signal is interrupted, the data

for a certain period of time or a certain moment cannot be transmitted to the background monitoring center. (3) Human interference: When technicians perform data preprocessing operations, the relevant data may be transmitted incorrectly for various reasons, which may also lead to data loss.

2.1.2. Types of Missing Data

According to the existing studies, different patterns of missing data occur in different time series. Considering the actual situation of landslide monitoring, the types of missing data in landslide monitoring time series can be usually classified into the following two types.

(1) Random missing (RM): Randomization can occur at any time during the monitoring process. Such data loss has no regularity or symptoms. The time of data loss is random and occasional, as shown in Figure 1.
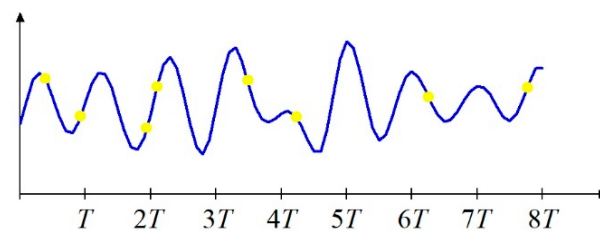
Figure 1. Random missing. The yellow dots represent missing data.

(2) Non-random missing (NM): In the monitoring process, regular or periodic phenomena such as cloudy weather and signal interruptions may occur with the same probability in a specific period of time, and such missing data has a certain regularity, as shown in Figure 2.
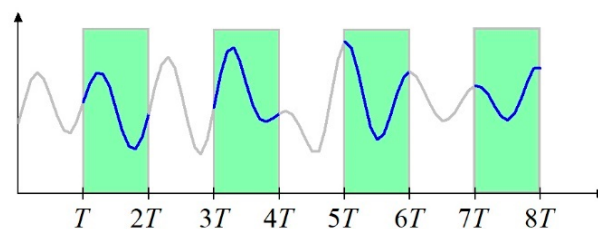
Figure 2. Non-random missing. The curves in the green area represent missing data.

2.2. Construction of Low-Rank Tensor Completion Model

2.2.1. Tensor Composition

In the landslide monitoring system, the monitoring data of each on-site monitoring sensor is collected through the intelligent terminal, and then the on-site monitoring data are transmitted to the background data monitoring center. Considering that the collected monitoring data include date, time, displacement and other information, it can be considered that the acquired landslide displacement time series is high-dimensional. For the landslide time series, this paper constructs a three-dimensional missing tensor $\mathcal{T} \in \mathrm{R}^{l_1,l_2,l_3}$, where l_1 is the number of on-site sensors for landslide, l_2 is days collected, and l_3 is the frequency of sampling every day. Therefore, the tensor represents the displacement monitoring value of each sensor at the corresponding time. This paper aims to identify the missing data position in tensor $\mathcal{T}$ and complete data completion and displacement prediction, respectively.

2.2.2. Construction of Low-Rank Tensor Completion Model

The low-rank matrix completion (LRMC) model uses the low-rank structure at the bottom to restore the incomplete matrix [21], defining the rank of the tensor $\mathcal{X} \in \mathrm{R}^{l_1, l_2, l_3}$ as $rank(\mathcal{X})$ and the objective function of low-rank tensor completion as:

$$\begin{cases} \min_{\mathcal{X}} rank(\mathcal{X}) \\ s.t.\mathcal{X}_\Omega = \mathcal{T}_\Omega \end{cases} \tag{1}$$

where $\mathcal{X} \in \mathrm{R}^{l_1, l_2, l_3}$ is a tensor to be solved; $\mathcal{T} \in \mathrm{R}^{l_1, l_2, l_3}$ is a missing tensor; Ω is the observed landslide data in T. Then, the tensor kernel norm is replaced by the rank minimum [22], and the concept of tensor kernel norm is defined:

$$\| \mathcal{X} \|_* = \sum_{i=1}^{3} \alpha_i \| X_{(i)} \|_* \tag{2}$$

where α_i is the weighting factor and $X_{(i)}$ is the matrix expanded along the i-th mode for tensor $\mathcal{X}$. The three pattern expansion matrices of the data tensor $\mathcal{X} \in \mathrm{R}^{l_1, l_2, l_3}$ are shown in Formula (3):

$$\begin{cases} X_{(1)} \in \mathrm{R}^{l_1 \times (l_2 \times l_3)} \\ X_{(2)} \in \mathrm{R}^{l_2 \times (l_1 \times l_3)} \\ X_{(1)} \in \mathrm{R}^{l_3 \times (l_1 \times l_2)} \end{cases} \tag{3}$$

where $X_{(1)}, X_{(2)}, X_{(3)}$ are the module expansion matrix for the staggered sampling of three types of expansion patterns, respectively.

The features of different orders between data are fused with each other through three module expansion matrices, effectively ensuring the high-precision completion of data [23]. Below is the singular value decomposition of each pattern expansion matrix [20]:

$$\begin{cases} X_{(1)} = U_{l_1 \times l_1} \overset{(1)}{\underset{l_1 \times l_2 l_3}{\sum}} M^T{}_{l_2 l_3 \times l_2 l_3} \\ X_{(2)} = V_{l_2 \times l_2} \overset{(2)}{\underset{l_2 \times l_1 l_3}{\sum}} L^T{}_{l_1 l_3 \times l_1 l_3} \\ X_{(3)} = W_{l_3 \times l_3} \overset{(3)}{\underset{l_3 \times l_1 l_2}{\sum}} K^T{}_{l_1 l_2 \times l_1 l_2} \end{cases} \tag{4}$$

where U, V and W are left singular matrices; M, L and K are right singular matrices; and Σ is a diagonal matrix. It is defined as:

$$\sum{}^{(q)} = diag\left(\sigma_1{}^{(q)}, \sigma_2{}^{(q)}, \cdots, \sigma_{n_q}{}^{(q)}\right), q = 1, 2, 3 \tag{5}$$

where n_q is the total number of singular values of the q-th module expansion matrix. Based on the definition of tensor kernel norm, the objective function is:

$$\begin{cases} \min_{\mathcal{X}} \sum_{i=1}^{3} \alpha_i \| X_{(i)} \| \\ s.t.\mathcal{X}_\Omega = \mathcal{T}_\Omega \end{cases} \tag{6}$$

where the matrix core norm is defined as [24]:

$$\| X \|_* = \sum_{k=1}^{j} \sigma_k(X) \tag{7}$$

where $\|\cdot\|_*$ is the kernel norm of X; j is the rank of matrix X; σ_k is the k-th singular value of the matrix X in order of size. Because of the existence of interdependent matrix kernel norm terms, the objective function of low-rank tensor completion is difficult to be solved by ordinary methods. Alternating direction method of multipliers (ADMM) [23,25] is a more widely used optimization method for constrained problems in machine learning. It is an extension of the augmented Lagrange method. ADMM algorithm provides a framework for solving constrained optimization problems with linear equations, which is convenient for disassembling the original optimization problem into several relatively solvable sub-optimization problems for iterative solutions. With continuous research, the low-rank tensor completion method is further developed. The purpose of reducing the computational complexity is to use QR decomposition instead of singular value decomposition. The decomposed low-rank tensor completion reduces the time complexity of the low-rank tensor completion algorithm. The nonlinear set CG algorithm based on Riemannian manifold reduces the decomposition of large-scale singular values in low-rank tensor completion. The tensor completion algorithm based on the Douglas–Rachford separation technology further considers the existence of the noise of the source data [25], which greatly enhances the robustness of the model. In addition, low-rank tensor completion also has a series of characteristics such as fast convergence speed and high computational accuracy.

2.3. Construction of MLATC

The model first transforms time series matrix into a tensor, which is recorded as sign $\mathcal{Q}(\bullet)$, such as $\mathcal{Q}(Y)$, transforming matrix $Y \in \mathrm{R}^{M \times N}$ into a third-order tensor of size $M \times N \times J$ [12].

2.3.1. Zero-Valued Low-Rank Autoregressive Tensor Completion Core Model

$$\min_{\mathcal{X}, \mathcal{Z}, A} \| \mathcal{X} \|_* + \lambda \| \mathcal{Z} \|_{A, \mathcal{X}}$$
$$s.t. \begin{cases} \mathcal{X} = \mathcal{Q}(Z) \\ \mathcal{P}_{\Omega}(Z) = \mathcal{P}_{\Omega}(Y) \end{cases} \tag{8}$$

where Y is the input observation matrix, and the matrix size of Z and Y is the same. $\| \mathcal{X} \|_*$ represents the kernel norm of the tensor $\mathcal{X}$, and the calculation method is $\| \mathcal{X} \|_* = \sum_k a_k \| X_{(k)} \|_*$; $\| X \|_* = \sum_{i=1}^{\min\{M,N\}} \sigma_i$. σ_i is the i-th largest singular value of matrix X; A is the variable to be estimated; λ is a control objective function to balance the weight parameters of the first and second items. $\| \mathcal{Z} \|_{A\mathcal{H}}$ represents the autoregressive norm of matrix Z, whose value is determined by Equation (9).

$$\| \mathcal{Z} \|_{A, \mathcal{H}} = \sum_{m,t} \left(z_{m,t} - \sum_i a_{m,t} z_{m,t-h_i} \right)^2 \tag{9}$$

$\mathcal{P}_{\Omega}(Z)$ represents the orthogonal projection of matrix Z in the field Ω, and the calculation formula is:

$$[\mathcal{P}_{\Omega}(Z)]_{m,n} = \begin{cases} z_{m,n} & if(m,n) \in \Omega \\ 0 & if(m,n) \notin \Omega \end{cases} \tag{10}$$

2.3.2. Mean-Based Low-Rank Autoregressive Tensor Completion

The time series matrix $Y \in \mathrm{R}^{M \times (N=I \times J)}$ with partially missing data can be compressed into a tensor $Y \in \mathrm{R}^{M \times I \times J}$, thus transforming the matrix completion problem into a tensor completion problem. The traditional tensor completion method assigns the missing data to zero. Considering the time accumulation and continuity of landslide displacement and deformation, the average value of the data before and after the missing data are taken instead of assigning the value to zero when completing the landslide displacement data.

The specific difference is that the projection of matrix Z in the domain Ω is further required to obtain a useful non-orthogonal projection:

$$\left[\mathcal{P}_\Omega\left(Z'\right)\right]_{m,n}{}' = \begin{cases} [\mathcal{P}_\Omega(Z)]_{m,n} & if(m,n) \in \Omega \\ \frac{1}{2t}\sum\limits_{i=n-t}^{n+t}[\mathcal{P}_\Omega(Z)]_{m,i} & if(m,n) \notin \Omega \end{cases} \tag{11}$$

where t is a hyperparameter representing the time sequence length of taking the mean value, in this case, 10 is taken. The matrix Z' conforming to such constraints is called a close orthogonal matrix of Y in the field Ω (recorded as $Z' \overset{\perp}{\sim} Y_\Omega$). For ease of understanding, Z' is still noted as Z.

MLATC is a simple linear combination of LRMC and vector auto-regressive (VAR) models. The first term of the objective function is the LRMC objective function, and the second term is the VAR objective function. The Lagrange function in matrix form is constructed according to the optimization problem:

$$L(\mathcal{X}, Z, A) = \sum_k a_k \| \mathcal{X}_{k(k)} \|_* + \lambda \| \mathcal{Z} \|_{A,\mathcal{H}} + \left\langle \mathcal{Q}^{-1}(\mathcal{X}) - Z, \mathcal{Q}^{-1}(\tau_k) \right\rangle + \frac{\rho}{2} \| \mathcal{Q}^{-1}(\mathcal{X}) - Z \|_F^2 \tag{12}$$

Then, Question (12) is transformed into $arg\min_{\mathcal{X}_{k(k)} Z,A} L(\mathcal{X}, Z, A, \tau_k)$, where X and Z are variables, and τ_k is the parameters to be learned. When solving the tensor $\mathcal{X}_k$, $\frac{\partial L(\mathcal{X},Z,A)}{\partial \mathcal{X}} = 0$ should hold. According to Formula (9), since $\frac{\partial \|Z\|_{A,\mathcal{H}}}{\partial \mathcal{X}} = 0$, $arg\min_{\mathcal{X}} L(\mathcal{X}, Z, A) \Leftrightarrow arg\min_{\mathcal{X}} L(\mathcal{X}, Z, A) - \| Z \|_{A,\mathcal{H}}$. Thus, $\mathcal{X}_k^{l+1}$ can be resolved by Formula (13):

$$\mathcal{X}_k^{l+1} = arg\min_{\mathcal{X}} a_k \| \mathcal{X}_{(k)} \|_* + \frac{\rho}{2} \| \mathcal{Q}^{-1}(\mathcal{X}) - Z^l \|_F^2 + \left\langle \mathcal{Q}^{-1}(\mathcal{X}) - Z^l, \mathcal{Q}^{-1}\left(\tau_k^l\right) \right\rangle \tag{13}$$

In Equation (13), the number l of iterations of the algorithm is expressed, and the VAR model is used to solve the matrix Z [12,21,26].

The MLATC achieves a more accurate completion of the original data than the zero-filled algorithm, and then the prediction of landslide displacement data can be completed on this basis.

3. Case Study

3.1. Experimental Dataset

The displacement monitoring data of the Shuizhuyuan landslide in the Three Gorges Reservoir area is selected as the time series' experimental dataset of this study. This time series is collected from seven GNSS monitoring sensors at the Shuizhuyuan landslide. The time interval is from 15 July 2017 to 1 December 2021, with a total of 1600 days. The data monitoring cycle is to obtain one monitoring data at the site every day, so the tensor size of the constructed dataset is $7 \times 1600 \times 1$ (sensors $\times$ time points $\times$ day) tensor structure. The monitoring data are shown in Figure 3.

3.2. Missing Data Processing for Time Series

In order to objectively describe the validity of the prediction model, two different data missing processing methods, random and non-random, are used to carry out certain data missing processes on the original time series dataset of landslides by combining the characteristics and main types of missing landslide monitoring data.

Random missing (RM): Random missing indicates that the landslide monitoring equipment has sporadic and random data loss in the operation process. The whole time series is divided according to the proportion of random missing data, with 5%, 10%, 20% and 40% set, respectively, representing different scales and levels of data loss of monitoring equipment.

Non-random missing (NM): Non-random missing means that the landslide monitoring equipment has regular equipment failure or regular network interruption during operation.

The data missing is also simulated according to 5%, 10%, 20% and 40% and compared with random missing.

The original spatiotemporal landslide displacement dataset is artificially treated with missing data, and the effectiveness of different models in completing and predicting the missing data can be objectively evaluated by two different treatments, random as well as non-random. For landslide disasters with relatively complex genetic mechanisms, there must be certain trend characteristics and a correlation between different sensors on the landslide. Therefore, better data recovery and prediction for missing datasets after non-random missing processing is an important indicator of this model.

Shuizhuyuan landslide is currently in the mean-slow deformation, and the first 1300 days are selected as the training set of the time series, and the last 300 days are selected as the test set. The MAPE $= \sum_{i=1}^{n} \left| \frac{\hat{x}_i - x_i}{x_i} \right| \times \frac{100\%}{n}$ and RMSE $= \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{x}_i - x_i)^2}$ are used to test the accuracy and precision of the algorithm model in terms of prediction. Moreover, the MLATC model is compared with the high-accuracy low-rank tensor completion (HALRTC) and temporal regularized matrix factorization (TRMF) methods. The completion and prediction results with a smaller MAPE and RMSE are considered to be better.
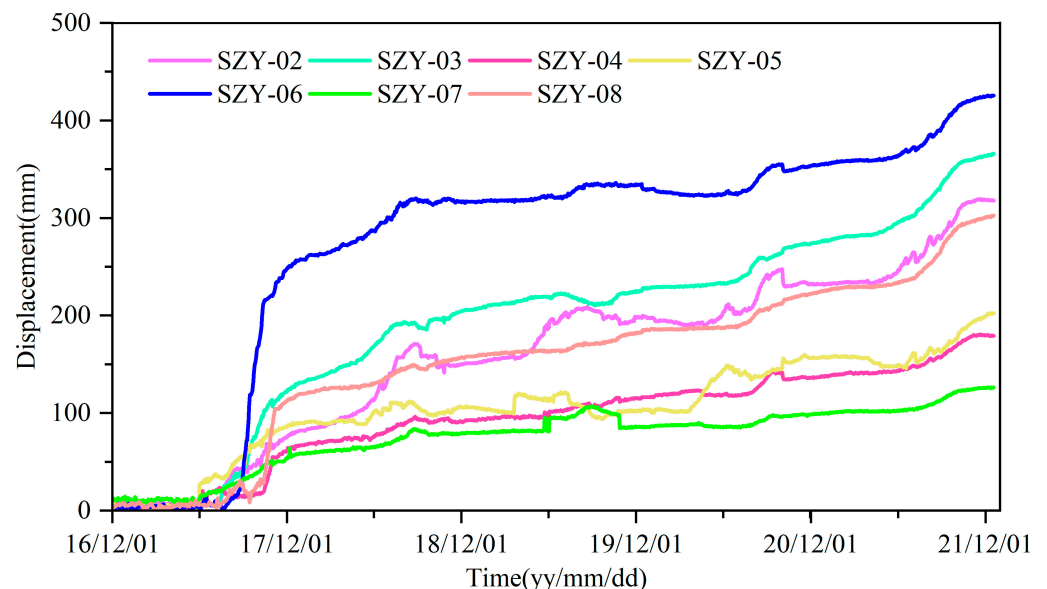
**Figure 3.** Landslide displacement monitoring data.

3.3. Data Completion and Analysis

The experimental results of four sensors numbered SZY-02, SZY-03, SZY-06 and SZY-08 are used as examples to introduce the effect of landslide monitoring data completion. After NM40% processing of the dataset of Shuizhuyuan landslide, the data completion effect of the completed training set of SZY-08 is shown in Figure 4. In Figure 4, the blue curve represents the measured data, the yellow dots represent the missing data, the red curve represents the recovered data, X-axis 0–260, 260–520, 520–780, 780–1040, 1040–1300 represents data recorded for each day and Y-axis is the displacement data. Considering the large amount of data for each sensor, only SZY-08 shows the complementary results for the entire time series, and the other sensors are the complementary results for the first 260 days. Figure 5 shows the complementary effect of the other sensors for the first 260 days.

As can be seen in Figures 4 and 5, in the absence of raw landslide displacement data for many consecutive days, the data have been completely processed into missing data, which is challenging for tensor completion. Missing data for several consecutive days has lost the correlation and trend characteristic information between displacement data. The prediction results show that the MLATC model still achieves a very good complementary

effect, which is essentially consistent with the original displacement data and achieves a good effective data recovery in some completely missing data days. This is very helpful to analyze the deformation law and deformation trend of the landslide, and also to provide data reference for understanding the relationship between different deformation areas of the landslide.
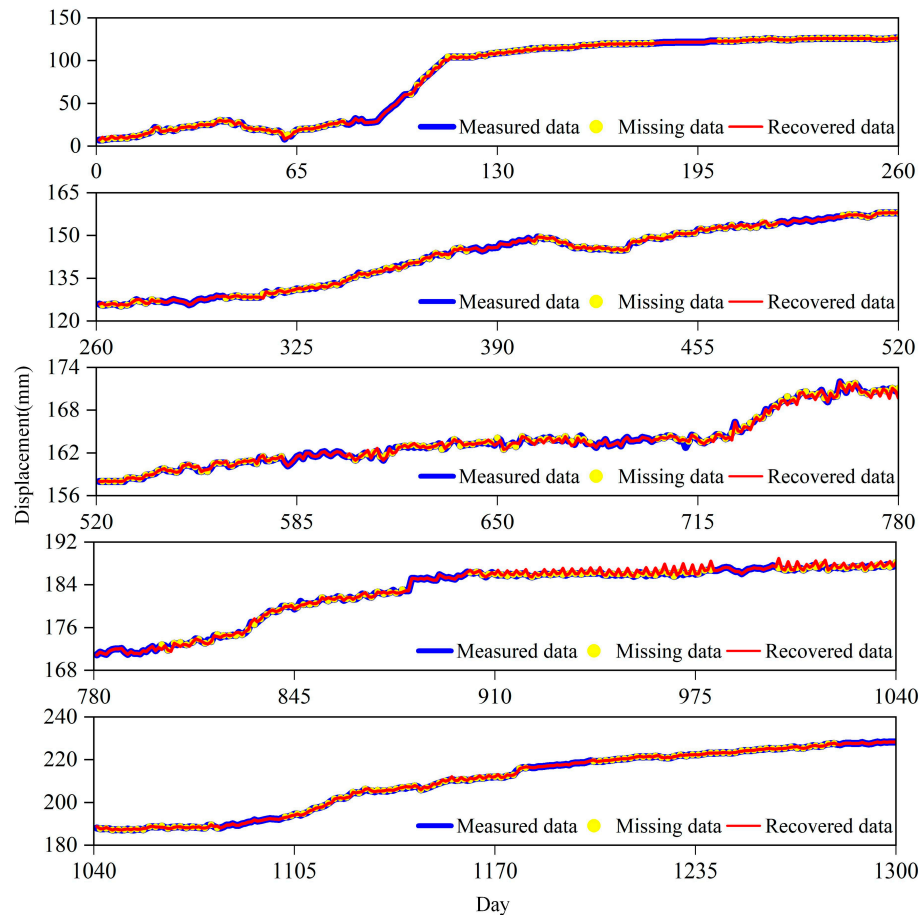
Figure 4. Completion of displacement data of SZY-08 (NM40%). The blue curve represents the measured data, the yellow dots represent the missing data and the red curve represents the recovered data.

Table 1 shows the evaluation metrics for data completion in the non-random missing case. The results in Table 1 show that the MLATC model has the best MAPE and RMSE for four different scenarios of NM5%, NM10%, NM20% and NM40%, indicating that the MLATC model has a greater improvement in data recovery performance compared to the HALRTC and TRMF models.

Table 1. Evaluation metrics for data completion in the NM case (MAPE/RMSE).

| Model	TRMF	HALRTC	MLATC
NM 5%	11.8542/14.0631	1.3161/3.8315	0.9066/0.9196
NM 10%	10.2681/16.5992	1.1953/3.6270	0.7770/0.8958
NM 20%	11.7050/18.1097	11.7976/25.2256	0.7708/0.9616
NM 40%	13.8412/25.9528	15.0777/38.6958	0.7928/1.0070

From Table 1, it can be concluded that the MLATC model is better than the HALRTC and TRMF models in terms of completion effect, indicating that the tensor autoregressive kernel parametrization can effectively replace the rank function, which enables the low-rank tensor completion model to obtain a more accurate completion effect. In addition,

the MLATC model introduces an autoregressive norm on the basis of the HALRTC model, which can make full use of the structural correlation and local trend feature information between the high-dimensional multivariate time series data, as well as more clearly correlate the multivariate time series data, which also proves that the addition of the autoregressive norm can further improve the complementary performance and accuracy of the model in the low-rank tensor complementary structure. Since the tensor structure well maintains the structural information of the spatiotemporal data in the time dimension and exploits the correlation between the daily displacements of different sensors in the time series, the tensor completion results of the MLATC model are better than those of TRMF in the matrix form. The experimental results with different missing ratios simultaneously verify this inference and confirm that effective completion and rolling prediction of displacement monitoring data can be achieved by using global time series datasets. After analyzing the reasons, the TRMF model only analyzes and calculates the matrix structure and does not use the time domain smoothness of multivariate time series and the potential correlation information between the series in the time and space dimensions. MLATC and HALRTC both use the tensor structure to complete the time series prediction through the method of quantitative completion. In particular, the MLATC model introduces the autoregressive norm, which not only preserves the structural information of the original time series through the tensor structure, but also makes full use of the correlation and trend characteristic information between the time series.
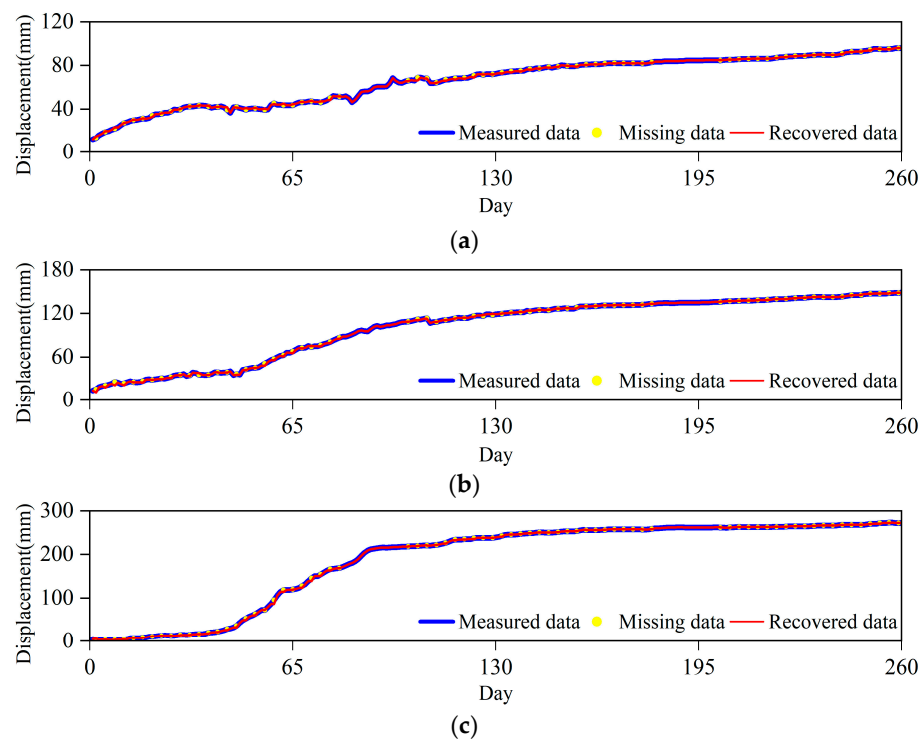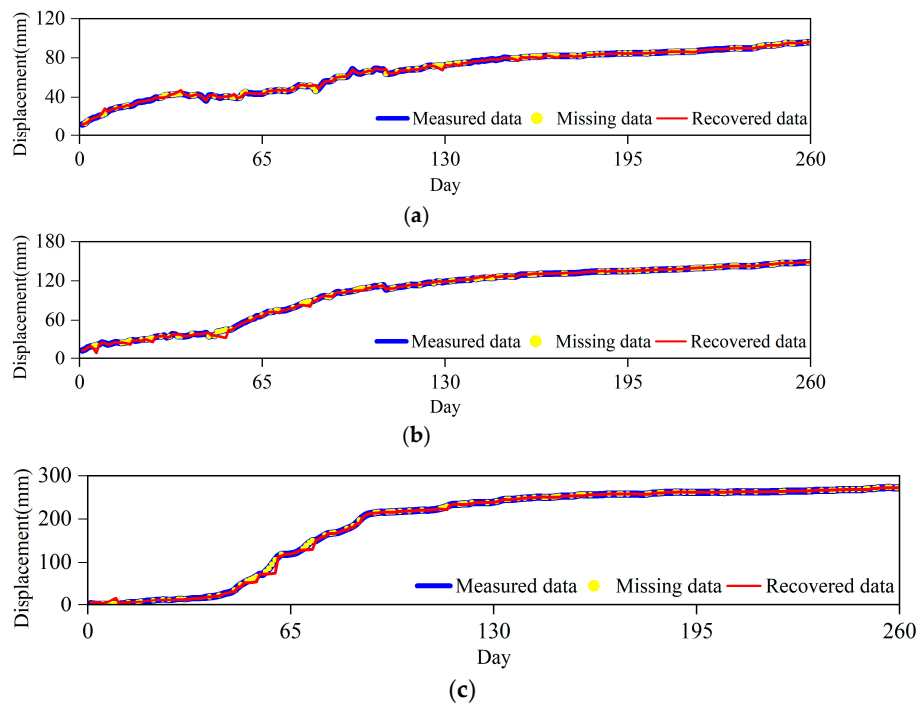

Figure 5. Completion of displacement data (NM40%). (**a**) Point SZY-02; (**b**) point SZY-03; (**c**) point SZY-06. The blue curve represents the measured data, the yellow dots represent the missing data and the red curve represents the recovered data.

Similarly, the experimental results of the four sensors numbered SZY-08, SZY-02, SZY-03 and SZY-06 are shown in Figures 6 and 7 after RM 40% processing of the dataset of the Shuizhuyuan landslide; the data completion performance are shown in Table 2. The prediction model has a very good recovery effect on missing data and shows a better data fit in terms of time series trend prediction.
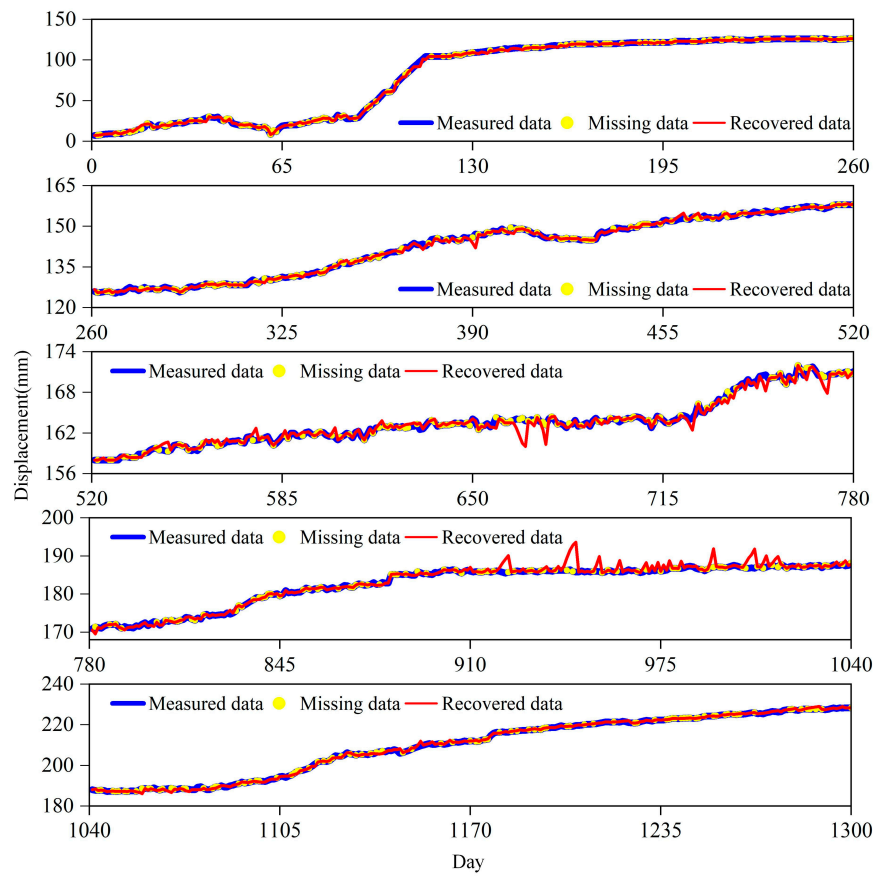
Figure 6. Completion of the displacement data of SZY-08 (RM40%). The blue curve represents the measured data, the yellow dots represent the missing data and the red curve represents the recovered data.
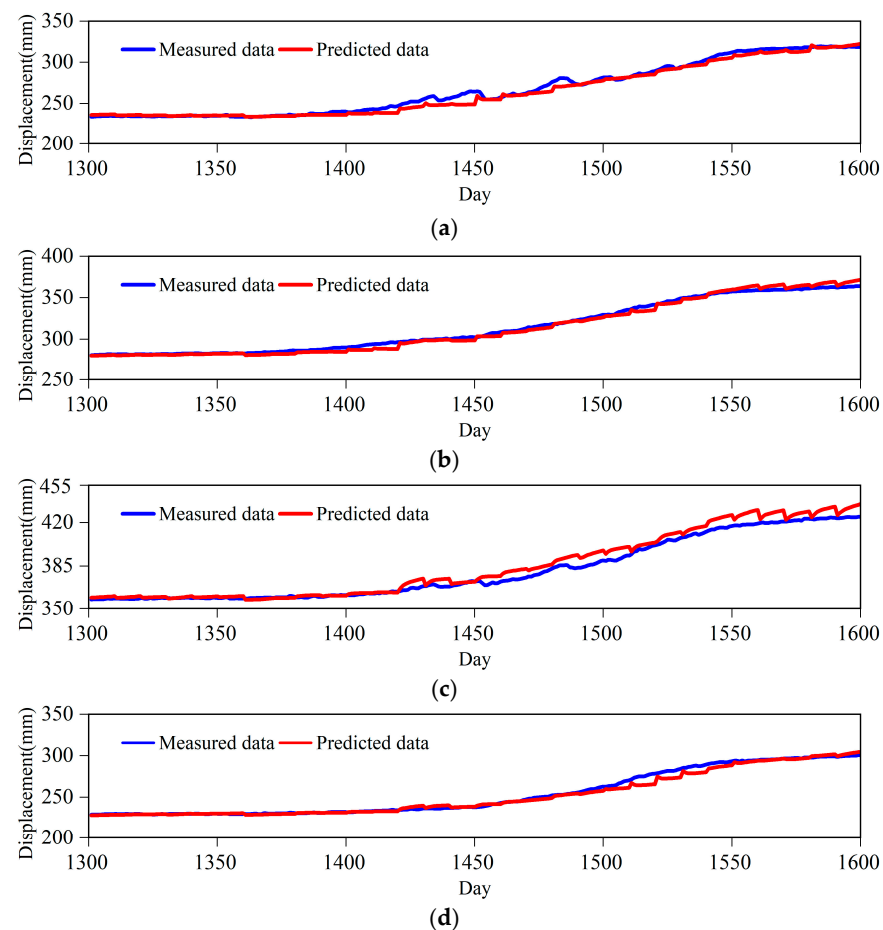

Figure 7. Completion of displacement data (RM40%). (**a**) Point SZY-02; (**b**) point SZY-03; (**c**) point SZY-06. The blue curve represents the measured data, the yellow dots represent the missing data and the red curve represents the recovered data.

Table 2. Evaluation metrics for data completion in the RM case (MAPE/RMSE).

| Model	TRMF	HALRTC	MLATC
RM 5%	11.8535/14.0634	0.8087/2.7628	0.7676/0.9880
RM 10%	10.2782/16.5925	0.9729/3.0545	0.7263/1.0339
RM 20%	11.7037/18.1142	11.7976/25.2256	0.9494/1.7648
RM 40%	13.8347/25.9556	15.0777/38.6958	1.4817/1.8889

3.4. Data Prediction and Analysis

In the case of random missing, the time series of the Shuizhuyuan landslide is predicted for 300 days after data missing processing is performed with a missing ratio of 40%. The time series of four sensors numbered SZY-02, SZY-03, SZY-06 and SZY-08 are analyzed separately, and the prediction results are shown in Figure 8. The analysis shows that the prediction results are ultimately consistent with the original displacement data in the case of the 40% missing data ratio. The MLATC model realizes well the deformation trend feature fitting of displacement, and effectively predicts the displacement data.

Figure 8. Prediction of displacement data (RM40%). (**a**) Point SZY-02; (**b**) point SZY-03; (**c**) point SZY-06; (**d**) point SZY-08. The blue curve represents the measured data and the red curve represents the predicted data.

Similarly, after processing the dataset of the Shuizhuyuan landslide with 40% non-random missing, the data completion effect is shown in Figure 9. The prediction results of MLATC model are also ultimately close to the original displacement data. Although there are some fluctuations in the prediction data, the overall effective fitting of displacement is still achieved.
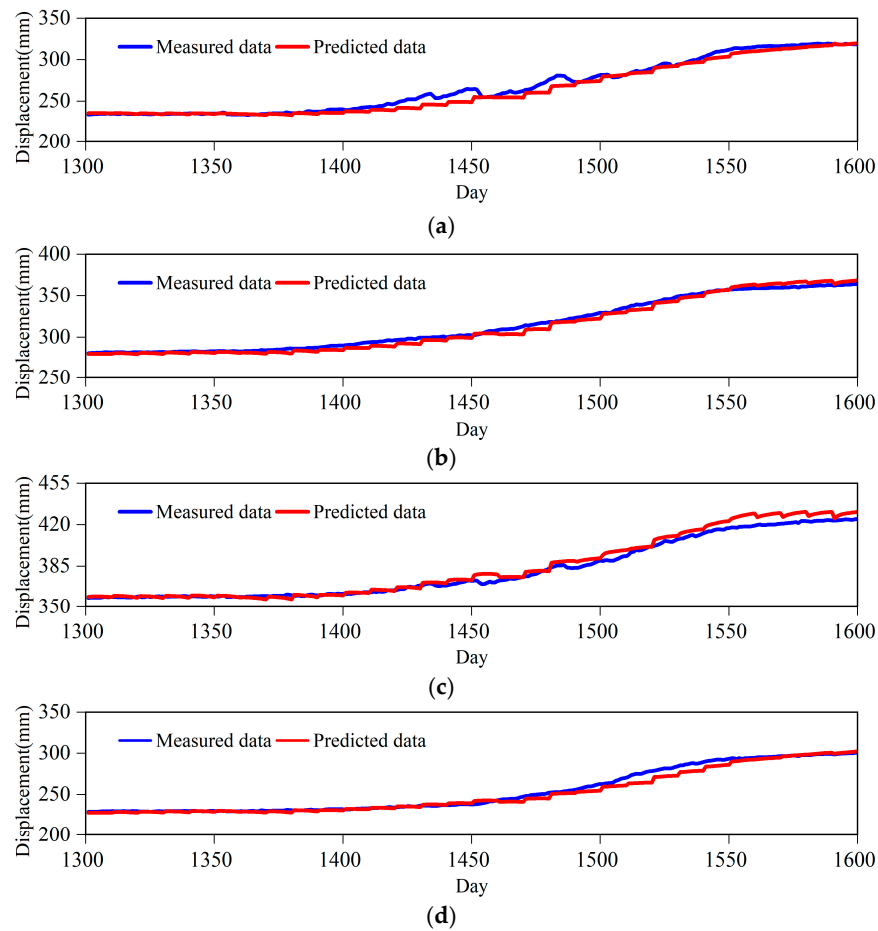
Figure 9. Prediction of displacement data (NM40%). (**a**) Point SZY-02; (**b**) point SZY-03; (**c**) point SZY-06; (**d**) point SZY-08. The blue curve represents the measured data and the red curve represents the predicted data.

The prediction effect of MLATC model under random and non-random missing is shown in Tables 3 and 4.

Table 3. Evaluation metrics for data prediction in the RM case (MAPE/RMSE).

Model	TRMF	HALRTC	MLATC
RM5%	39.6023/102.415	15.2062/39.8024	1.1084/3.6774
RM10%	41.1064/106.681	14.8872/39.0352	1.0792/3.6587
RM20%	50.2660/127.588	14.8241/39.1718	1.1120/3.7032
RM40%	83.7915/213.040	13.8827/37.1245	1.1012/3.6665

Table 4. Evaluation metrics for data prediction in the NM case (MAPE/RMSE).

Model	TRMF	HALRTC	MLATC
NM5%	39.4527/102.047	14.9328/38.9776	1.1079/3.6676
NM10%	42.7754/110.746	14.2427/37.2552	1.1037/3.6588
NM20%	49.9770/126.921	12.6523/33.3975	1.1099/3.6785
NM40%	83.7004/212.867	11.5458/30.6736	1.1694/3.8755

By analyzing the prediction results of each model, it can be proven that the prediction accuracy of the time series model is effectively improved after converting the original time series into a tensor structure in this method. The prediction accuracy of both HALRTC model and MLATC model has improved significantly over the prediction accuracy of

TRMF. This indicates that although the TRMF model captures the global consistency of the time series, the model mainly acts on the latent layer matrix data, which may cause the local trend feature information between different sensors to be ignored, thus affecting the prediction accuracy. Due to the intervention of the autoregressive norm, the prediction of HALRTC and MLATC models is significantly better than that of TRMF model, and the MAPE and RMSE accuracy of the models are higher. This demonstrates that the spatiotemporal monitoring data between displacement sensors in different areas of the landslide are utilized under the framework of tensor complementary structure, and the deformation law characteristics of short-term correlation and long-term deformation consistency in the inherent deformation process of the landslide are fully considered.

4. Discussion

Affected by the complex environment in the field, there will be missing data in the process of landslide monitoring, which will affect the accurate analysis of landslide displacement. In landslide disaster monitoring, landslide displacement deformation is relatively complex, and the displacement changes of each deformation area on the landslide are not the same. From the analysis of landslide local deformation characteristics, there is a certain correlation between the daily displacement deformation of different monitoring points. In addition, landslide displacement is a continuous cumulative process, and the landslide displacement occurring in the preceding period will have a certain influence on the landslide displacement occurring subsequently. Therefore, in order to better realize landslide data completion and prediction, a novel model based on MLATC method is proposed, and the RM and NM cases are designed, respectively, so as to verify the validity and reliability of the designed model. In addition, from the construction of the model, the initial time series matrix is converted into a third-order tensor structure. Under the assumption that the time series data satisfy approximate low-rankness, the time series completion and prediction problem is transformed into low-rank tensor completion and prediction. The tensor-based completion method fully considers the time series of landslide displacement data, which not only preserves the correlation between different displacements, but also better fits the landslide displacement deformation characteristics, making the displacement completion more accurate.

To verify the effectiveness of the algorithm, a comparative analysis with the existing RTMF and HALRTC models was performed. The MAPE and RMSE of the MLATC model are 0.9066, 0.9196 and 0.7676, 0.9880 for the NM5% and RM5% data completion, respectively. Similarly, the MAPE and RMSE for NM5% and RM5% data prediction are 1.1079, 3.6676 and 1.1084, 3.6774, respectively. The analysis results show that under the conditions of 5%, 10%, 20% and 40% missing data, the data completion and prediction effects of the MLATC model are better than those of other models, which also confirms the significant data completion and prediction effects of the MLATC model. In this study, the completion and prediction model were constructed by considering the intrinsic correlation between landslide displacement data and using the low-rankness of the completion tensor. Data completion is an iterative calculation of the landslide displacement time series based on the entire original data. In order to verify the effectiveness of the data completion model, random and non-random data missing cases are designed, and time series prediction is based on the predicted experimental data as the missing value, which also belongs to a special case of missing data. By default, this part of the experimental data is not involved in iterative calculations. Therefore, the data prediction for time series is the same as the method used for data completion. The model is implemented in a rolling prediction method with cyclic and iterative computation, which leads to a loss in the amount of data in the prediction case. The MAPE and RMSE values of NM and RM also show that the prediction effect of the MLATC model is lower than that of data completion, but both can achieve satisfactory results, which are more in line with the actual needs of landslide monitoring.

5. Conclusions

A novel method for landslide displacement data completion and prediction based on MLATC model is proposed for missing landslide displacement data and time series prediction. The tensor structure in the MLATC model well perpetuates the structural information of landslide spatiotemporal data in the time dimension, makes full use of the correlation of landslide displacement time series in different time scales, and solves the problem of serious data loss caused by the destruction of the original matrix structure. The data completion and prediction models are implemented in a tensor structure framework, combining VAR models and autoregressive paradigms, and different proportions of random and non-random missing cases are selected for experimental analysis. The experimental results of the Shuizhuyuan landslide prove that the model can also achieve the effective completion of missing data and displacement trend prediction of landslide displacement time series without the need of complete original landslide displacement data. The reliability and accuracy of the model data completion and prediction are verified, further improving the feasibility of the model, which likewise helps to issue timely and accurate early warning forecast signals to remind people who are in the danger area to evacuate and avoid casualties and property damage. The method can be extended and applied in data completion and in the prediction of such landslides.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Abbreviations

Symbols	Description
$\mathcal{X} \in \mathrm{R}^{n_1 \times \cdots \times n_1}$	d-order tensor
$X_{(i)}$	The matrix that expanded along the i-th mode for tensor $\mathcal{X}$
$\parallel \mathcal{X} \parallel_*$	The kernel norm of the tensor $\mathcal{X}$
$\mathcal{Q}(\bullet)$	Transform the time series matrix into a third-order tensor
$\mathcal{P}_\Omega(Z)$	The orthogonal projection of matrix Z in the field Ω
Abbreviations	**Full Name**
MLATC	Mean-based Low-rank Autoregressive Tensor Completion
LRMC	Low-Rank Matrix Completion
HALRTC	High-Accuracy Low-Rank Tensor Completion
TRMF	Temporal Regularized Matrix Factorization
ADMM	Alternating Direction Method of Multipliers
VAR	Vector Auto-Regressive
RM	Random Missing
NM	Non-random Missing
MAPE	Mean Absolute Percentage Error
RMSE	Root-Mean-Squared Error

References

1. Wang, J.; Nie, G.; Xue, C. Landslide displacement prediction based on time series analysis and data assimilation with hydrological factors. *Arab. J. Geosci.* **2020**, *13*, 460. [CrossRef]
2. Liu, Z.-Q.; Guo, D.; Lacasse, S.; Li, J.-H.; Yang, B.-B.; Choi, J.-C. Algorithms for intelligent prediction of landslide displacements. *J. Zhejiang Univ. Sci. A* **2020**, *21*, 412–429. [CrossRef]
3. Zhang, Y.; Tang, J.; He, Z.; Tan, J.; Li, C. A novel displacement prediction method using gated recurrent unit model with time series analysis in the Erdaohe landslide. *Nat. Hazards* **2020**, *105*, 783–813. [CrossRef]
4. Li, H.; Xu, Q.; He, Y.; Deng, J. Prediction of landslide displacement with an ensemble-based extreme learning machine and copula models. *Landslides* **2018**, *15*, 2047–2059. [CrossRef]
5. Yang, B.; Yin, K.; Lacasse, S.; Liu, Z. Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides* **2019**, *16*, 677–694. [CrossRef]
6. Segoni, S.; Piciullo, L.; Gariano, S.L. Preface: Landslide early warning systems: Monitoring systems, rainfall thresholds, warning models, performance evaluation and risk perception. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 3179–3186. [CrossRef]
7. Zhu, X.; Zhang, S.; Jin, Z.; Zhang, Z.; Xu, Z. Missing value estimation for mixed-attribute data sets. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 110–121. [CrossRef]
8. Li, L.; Zhang, J.; Wang, Y.; Ran, B. Missing value imputation for traffic-related time series data based on a multi-view learning method. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2933–2943. [CrossRef]
9. Kolda, T.G.; Bader, B.W. Tensor decompositions and applications. *SIAM Rev.* **2009**, *51*, 455–500. [CrossRef]
10. Jing, P.; Su, Y.; Jin, X.; Zhang, C. High-order temporal correlation model learning for time-series prediction. *IEEE Trans. Cybern.* **2018**, *49*, 2385–2397. [CrossRef]
11. Sen, R.; Yu, H.; Dhillon, I.S. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1–14. [CrossRef]
12. Chen, X.; Sun, L. Bayesian temporal factorization for multidimensional time series prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 4659–4673. [CrossRef] [PubMed]
13. Yu, H.; Rao, N.; Dhillon, I.S. High-dimensional time series prediction with missing values. *arXiv* **2015**, arXiv:1509.08333.
14. Zhang, P.; Ren, P.; Liu, Y.; Sun, H. Autoregressive matrix factorization for imputation and forecasting of spatiotemporal structural monitoring time series. *Mech. Syst. Signal Process.* **2022**, *169*, 108718. [CrossRef]
15. Zhu, H.; Ding, X.; Liu, G. Rainfall prediction based on tensor complement. *Comput. Appl. Softw.* **2022**, *39*, 218–222+280. [CrossRef]
16. Chen, K.; Dong, H.; Chan, K.-S. Reduced rank regression via adaptive nuclear norm penalization. *Biometrika* **2013**, *100*, 901–920. [CrossRef]
17. Tan, H.; Wu, Y.; Shen, B.; Jin, P.J.; Ran, B. Short-term traffic prediction based on dynamic tensor completion. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2123–2133. [CrossRef]
18. Figueiredo, M.; Ribeiro, B.; de Almeida, A. Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home. *IEEE Trans. Instrum. Meas.* **2013**, *63*, 364–373. [CrossRef]
19. Li, D.; Yu, J.; Gao, W.; Chen, S.; Zhu, F. Financial time series prediction algorithm combining delay transformation and tensor decomposition. *Comput. Eng. Des.* **2022**, *43*, 1295–1303. [CrossRef]
20. Cai, J.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [CrossRef]
21. Chen, X.; Sun, L. Low-rank autoregressive tensor completion for multivariate time series forecasting. *arXiv* **2020**, arXiv:2006.10436.
22. Liu, J.; Musialski, P.; Wonka, P.; Ye, J. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell* **2013**, *35*, 208–220. [CrossRef] [PubMed]
23. Zhao, H.; Shou, P.; Ma, L. A Tensor Completion Method of Missing Data in Transformer District. *Proc. CSEE* **2020**, *40*, 7328–7336. [CrossRef]
24. Candes, E.; Recht, B. Exact matrix completion via convex optimization. *Commun. ACM* **2012**, *55*, 111–119. [CrossRef]
25. Ouyang, W.; Peng, Y.; Yao, Y.; Zhang, J.; Deng, B. Anderson Acceleration for Nonconvex ADMM Based on Douglas-Rachford Splitting. *Comput. Graph. Forum* **2020**, *39*, 221–239. [CrossRef]
26. Xia, H.; Dong, Q.; Chen, Y.; Zheng, J.; Gao, C.; Wang, Z. QoS Prediction Based on the Low-Rank Autoregressive Tensor Completion. In Proceedings of the 2022 International Conference on Networking and Network Applications (NaNA), Urumqi, China, 3–5 December 2022; pp. 265–269.