

## Research Article

# A Study of Missing Collaborative Data Imputation Models based on Same-City Delivery

Xintong Zou  and Hui Jin 

*School of Automotive and Traffic Engineering, Liaoning University of Technology, Jinzhou 121000, China*

Correspondence should be addressed to Hui Jin; [jinhui5868@163.com](mailto:jinhui5868@163.com)

Received 27 April 2022; Accepted 15 June 2022; Published 11 July 2022

Academic Editor: Wen Liu

Copyright © 2022 Xintong Zou and Hui Jin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With advent of the postepidemic era, the development of digital logistics operations management is imminent. Among the various logistics delivery methods, same-city delivery is chosen by the vast majority of customers for its timeliness and safety. Online ordering and delivery methods for same-city delivery are also gaining increasing attention from enterprises which need to know the inventory balance of all same-city warehouses in time for early deployment and response. However, in practice, the inventory balance of each warehouse can be affected by other warehouses in the same city, and there is often a lack of data in the inventory management system due to equipment and other issues resulting in a poor response from the company to handle emergencies. To address these issues, an improved matrix decomposition model was designed to interpolate the missing data by taking into account the spatiotemporal correlation between warehouses. The L-curve criterion was used to select hyperparameter values, the spatiotemporal regularize was used to capture the time dependence of the time series, and the model performance was evaluated using root mean square error and mean absolute percentage error. Comparisons with classical interpolation techniques were made to validate the improved performance of the proposed method.

## 1. Introduction

The emergence of the meta-universe concept in recent years has promoted digital development. With the advent of digital logistics, customers are increasingly demanding from platform-based e-commerce companies [1]. Due to many physical warehouses of platform-based e-commerce enterprises, orders are fragmented, statistics are untimely, equipment errors are unable to be transmitted in real time, and other realities, which make it difficult for companies to know the inventory balance of the same-city physical warehouses in time. It can easily lead to problems such as low sensitivity of inventory scheduling and the inability of companies to respond in time after customers adjust their delivery time. Integration of same-city physical warehouses data with enterprise systems is urgent [2]. Completeness of the same-city physical warehouse data has a direct impact on the integration performance. Interpolation missing data in same-city physical warehouses is of importance [3].

However, current research methods for the interpolation of missing data in logistics are less developed than in other fields such as transportation [4], meteorology [5], and so on. The main reason is the redundancy of logistics data and is influenced by synergies across regions. The current approach would result in an additional computational burden on the model. Regularized decomposition as a fast measure to deal with missing data by using different regular term functions to deal with low-rank matrix decomposition problems and have being decreased the time cost dramatically. The flexible use of regularized decomposition clearly achieves a better performance [6]. To improve the interpolation performance, researchers at home and abroad have carried out research on regularized data interpolation models.

In the logistics field, new technology, new markets, new business models, and new customer expectations cause the logistics industry to face immense change, which brings both risk and opportunities. Big data technology makes the logistics transportation process easier to manage, so that goods

can be quickly transported to the designated location, to ensure the efficiency of logistics transportation [7].

For example, to explore time-series analysis of the dynamic variables and their interrelated influence in both the short and long run on the relationship between modern logistic industry and economic growth, it implies that logistic industrial development is comparatively quicker in the geographical areas where economic growth is higher than those areas where economic growth is low [8]. Logistics 4.0 details the strategic direction of future logistics in the industrial sector [9]. For logistics inventory, BP neural networks can be used to explore the complex relationship between inventory demand and various influencing factors to obtain effective measures for inventory control [10]. Long Short-Term Memory (LSTM), on the other hand, as a variant of RNN, has been applied in the field of logistics data processing because it is ideally suited for problems that are highly correlated with time series. For example, LSTM is used to build a prediction model for stock price forecasting in the logistics industry [11]. To improve the overall efficiency of the logistics system and the level of customer service proposes the coordinated development of urban intelligent transportation data system and supply chain management. Aims to study the importance and advantages of the coordinated development of urban intelligent transportation data system and supply chain management [12]. Unfortunately, there are few of these analyses for missing data in logistics.

In machine learning, different issues arise during the statistical or data-driven control of interpolation methods, which result in discrepancies between the interpolation and the true values. Most popular imputation methods rely on discriminative models such as decision trees, regression models, and neighborhood-based methods, which are relatively interpretable. Unfortunately, systematically missing data may yield biased sample statistics, which causes these methods to generalize poorly. To address this limitation, different imputation methods turn to generative models to improve imputation accuracy, the structural assumptions encoded in the models can lead to better generalization. A multilayer perceptron-multivariate imputation of chain equation regression imputation method optimized by the limit-memory-BFGS algorithm is proposed, considering the temporal and spatial characteristics of traffic volume [13]. The vehicle dispatching system is one of the most critical problems in online ride-hailing platforms, which require adapting the operation and management strategy to the dynamics of demand and supply. In this paper, we propose a single-agent deep reinforcement learning approach for the vehicle-dispatching problem called deep dispatching, by reallocating vacant vehicles to regions with a large demand gap in advance [14]. Bayesian interpolation model is proposed to characterize the data generation process and to learn the underlying statistical patterns in the traffic data [15]. Considering the spatiotemporal nature of IoT data and the uncertainty of the data collected by sensors, Bayesian maximum entropy (BME) which to impute missing values utilizing as a convenient means to estimate the missing data from IoT applications [16]. On the other hand, state-of-the-

art imputation methods are based on graph Laplacian (GL) transform for a semisupervised feature extraction; the scatter matrix fused features update the output weights in real-time online [17]. Training on limit learning machines, with the ability to extract features, realize the multivariate real-time prediction of the time series and complements the model according to the general matrix and low-rank matrix decomposition by the regularization matrix factorization (MF). The decomposed matrix factors are weighted by a common symmetric matrix new matrix completion model and regularization weighted functions are obtained, and then the optimal completion matrix for data completion is obtained [18].

The model software implementation, Spss, uses mean fill which is not suitable for nonrandom missing datasets. Amos uses the original highest likelihood method which is suitable for randomly missing datasets. Solas uses multiple fill regression which is not suitable for nonrandom missing datasets. Matlab is suitable for fast running and needs to display the results step by step. Python is suitable for cases where the data package can be called directly.

These studies have highlighted that most of the models have a strong reliance on mathematics, and there is little research on models for interpolation of missing datasets in logistics. Modeling errors due to different data types can have an impact on the actual results. To addresses problem of large-scale fine-grained traffic state prediction, a deep learning architecture is proposed to handle the challenge of data scale, granularity, and sparsity. Based on the domain knowledge of traffic engineering, the propagation of traffic state is analyzed, and four modules are specially designed and elaborated in the aspects of the temporal propagation of traffic state, spatial propagation of traffic state, etc. [19]. To achieve higher performance for interpolation missing data in logistics, this paper designs an improved LSTM and Graph Laplacian Regularized Matrix Factorization (LSTM-GL-ReMF) system [20]. The L-curve, the temporal regularizer LSTM, and the spatial regularizer GL are combined to build the entity interpolation model for the inventory balance dataset of logistics entities warehouses in each region of the same city. The L-curve is used to select hyperparameters to take values [21]. The temporal regularizer, LSTM, is to capture temporal dependencies in time-series data and the spatial regularizer, GL, is to capture spatial correlations between different warehouses, solving the problem that interpolation results considering only individual warehouses are far from reality by considering spatiotemporal correlation. The model is evaluated using two evaluation metrics, RMSE and MAPE, and compared with other state-of-the-art decomposition models to prove the superior interpolation performance of the improved LSTM-GL-ReMF model in the field of missing logistics data. The main contributions of this paper are summarized as follows:

- (i) The L-curve criterion is used to select the values of the regularized hyperparameters and to increase the interpretability of the hyperparameter values.
- (ii) Fusing the L-curve criterion with the LSTM and GL algorithm to construct the model.

- (iii) Apply the improved model to the domain of logistics operations management.

## 2. Fundamentals

This section presents the basic theory and modeling principles of this thesis. Includes the definition of operations, matrix decomposition steps, and LSTM-GL-ReMF model theory, where matrix operations are the basis of matrix decomposition and the improved models are optimized on the basis of the basic principles.

### 2.1. Matrix Decomposition

*Definition 1.* Given two matrices  $\mathbf{A} \in R^{m_1 * n_1}$  and  $\mathbf{B} \in R^{m_2 * n_2}$ , the Kronecker product is as follows:

$$\mathbf{A} * \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1m_2}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m_11}\mathbf{B} & \dots & a_{m_1m_2}\mathbf{B} \end{pmatrix}. \quad (1)$$

*Definition 2.* Given two matrices  $A = (a_1, a_2, \dots, a_r) \in R^{m * r}$  and  $\mathbf{B} = (\beta_1, \beta_2, \dots, \beta_r) \in R^{\omega * r}$  which have the same number of columns, the KR product is as follows [22]:

$$\mathbf{A}\phi\mathbf{B} = (a_1 * \beta_1, a_2 * \beta_2, \dots, a_r * \beta_r) \in R^{(m\omega) * r}. \quad (2)$$

The matrix decomposition steps (Algorithm 1) are as follows:

There are three matrices  $\mathbf{r}(|U| * |D|)$ ,  $\mathbf{p}(|U| * K)$ ,  $\mathbf{q}(|D| * K)$  which are trained using either random gradients or batch gradient and combined with a gradient descent method with repeated sampling. The specific formula is as follows: first calculate the minimum differential as shown in equations (3) and (4), where the potential characteristic is  $K$  and the minimum value of the difference is  $e$  [23]:

$$\frac{\delta}{\delta \mathbf{p}_{ij}} e_{ij}^2 = -2(\mathbf{r}_{ij} - \hat{\mathbf{r}}_{ij})(\mathbf{q}_{kj}) = -2e_{ij}\mathbf{q}_{kj}, \quad (3)$$

$$\frac{\delta}{\delta \mathbf{q}_{ij}} e_{ij}^2 = -2(\mathbf{r}_{ij} - \hat{\mathbf{r}}_{ij})(\mathbf{q}_{kj}), = -2e_{ij}\mathbf{p}_{kj}. \quad (4)$$

Get iterative matrix elements  $p'_{ik}$  and  $q'_{kj}$ :

$$p'_{ik} = p_{ik} + \alpha \frac{\delta}{\delta q_{ij}} e_{ij}^2 = p_{ik} + 2\alpha e_{ij} q_{kj}, \quad (5)$$

$$q'_{kj} = p_{kj} + \alpha \frac{\delta}{\delta q_{kj}} e_{ij}^2 = p_{kj} + 2\alpha e_{ij} p_{ik}.$$

Finally, set the threshold value and reduce the error by the error function.

*2.2. LSTM-GL-ReMF.* In the LSTM-GL-ReMF, its temporal regularizer introduces the recurrent neural network LSTM with extremely strong learning performance to discover hidden correlations in the samples. Hochreiter

and Schmidhuber proposed the LSTM which could overcome long-term dependencies and determine the best time window automatically [24]. Calculate the hidden layer sequence (6) and output sequence (7) by equations  $\varepsilon$  to  $\gamma$ :

$$\varepsilon_t = f_a(w_{xe}x_t + w_{\varepsilon\varepsilon}\varepsilon_{t-1} + \bar{\mathbf{b}}_\varepsilon), \quad (6)$$

$$\gamma_t = w_{\varepsilon\gamma}\varepsilon_t + \bar{\mathbf{b}}_\gamma, \quad (7)$$

where the activation function is  $f_a$ , the given sequence is  $x$ , the weight matrix is  $\omega$ , the bias vector is  $\bar{\mathbf{b}}$  and the moment is  $t$  [25]:

$$\begin{aligned} i_t &= \sigma(\omega_{xi}x_t + \omega_{ei}\varepsilon_{t-1} + \omega_{ci}c_{t-1} + \bar{\mathbf{b}}_i), \\ \zeta_t &= \sigma(\omega_{x\zeta}x_t + \omega_{\varepsilon\zeta}\varepsilon_{t-1} + \omega_{c\zeta}c_{t-1} + \bar{\mathbf{b}}_\zeta), \\ c_t &= \zeta_t c_{t-1} + i_t \tan \varepsilon(\omega_{xc}x_t + \omega_{\varepsilon c}\varepsilon_{t-1} + \bar{\mathbf{b}}_c), \\ O_t &= \sigma(\omega_{xo}x_t + \omega_{\varepsilon o}\varepsilon_{t-1} + \omega_{co}c_t + \bar{\mathbf{b}}_o), \\ \varepsilon_t &= o_t \tan \varepsilon(c_t), \end{aligned} \quad (8)$$

where the input gate is  $i$ , the forgetting gate is  $\zeta$ , the cell state is  $c$ , and the output gate is  $O$ . The sigmoid activation function is  $\sigma$  and the tangent hyperbolic activation function is  $\tan \varepsilon$ . Spatial regularizers are based on GL spatial regularization [26] and this can be able to combine complex spatial and temporal dependencies into the decomposition to increase the accuracy of the interpolation performance. Data smoothing is measured using  $\|f_k\|_M^2$ :

$$\|f_k\|_M^2 = \int_{x \in M} \|Mf_k\|^2 dp_x(x), \quad (9)$$

where  $\Delta M$  is used to denote the gradient of  $f_k$  along the stream shape  $M$ . Define the edge weight matrix  $\mathbf{w}_{ij}$  as follows:

$$\mathbf{w}_{ij} = \begin{cases} 1, & \text{if } x_i \in N_\rho(x_j) \text{ or } x_j \in N_\rho(x_i), \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where  $N_\rho(x_i)$  denotes the set of  $x_i$  nearest neighbors of  $\rho$ . Subsequently, the discrete approximation is calculated as follows:

$$R_k = \frac{1}{2} \sum_{i,j=1}^N (f_k(x_i) - f_k(x_j))^2 \mathbf{w}_{ij} = v_k^T L v_k. \quad (11)$$

Of these, Graph Laplacian is  $L$  and the mapping function is  $v_k$ . Then the objective function is as follows:

$$O = \|X - UV^T\|_F^2 + \lambda \text{Tr}(V^T L V). \quad (12)$$

Of these, matrix trajectory is  $\text{Tr}$ , regularization parameter is  $\lambda$  [27].

## 3. Model Development and Refinement

This section describes the improvements made to the model. The improved model is more applicable to the type of dataset presented in this paper.

**3.1. Problem Definition.** With the development of e-commerce, city-city distribution has been favored by more users, such as booking delivery in city-city delivery. However, due to various external factors, customers' predetermined time is easy to change. At this time, enterprises need to have an overall control of the same-city inventory, timely, and reasonable scheduling. At present, when most enterprises use Warehouse Management System (WMS) system for data integration, there is often some missing data, resulting in enterprises being unable to control the overall inventory in the same city. Based on the same-city warehouse inventory in its time and space of the close connection, a modified LSTM-GL-ReMF algorithm was designed to interpolate the missing data.

### 3.2. L-Curve Guidelines

**Lemma 1.** Operator equations on Banach spaces  $X, Y$ .

$$Kx = y. \quad (13)$$

Of these,  $x \in X$  and  $y \in Y$ ,  $K$  is a bounded linear tight operator from  $X$  to  $Y$ ,  $y$  is the measurement data item,  $y^\delta$  is the test data after being perturbed by the error, error level  $\delta > 0$ .

$$\|y - y^\delta\| \leq \delta, \quad y^\delta \in Y. \quad (14)$$

**Lemma 2.** In the plane, all points  $(\|Kx^{\alpha,\delta} - y^\delta\|, \|x^\alpha\|)$  form a monotonically decreasing L-curve with the point of greatest curvature being the corner of the L-curve [28]:

$$\begin{aligned} u(\alpha) &= \log \|Kx^{\alpha,\delta} - y^\delta\|, \\ v(\alpha) &= \log \|x^{\alpha,\delta}\|. \end{aligned} \quad (15)$$

Of these,  $x^{\alpha,\delta} = R_\alpha y^\delta$  and  $R_\alpha$  is a family of consistent-bounded operators. The curvature function of the L-curve with  $\alpha$  as a parameter [29]:

$$K(\alpha) = \frac{|u'v'' - u''v'|}{[(u')^2 + (v')^2]^{3/2}}, \quad (16)$$

The triple spline interpolation method can be used to approximate the corner angles and thus select the regularization parameters. The main idea is to form a better approximation to the L-curve by gradually increasing the number of nodes and to use the parameter corresponding to the point of maximum curvature on the final strip as the regularization parameter. The steps are as follows [30]:

*Step 1:* given the initial regularization parameter,  $\alpha_i$  and its corresponding  $(u_i, v_i)$ .

*Step 2:* interpolation of three splines about  $\beta_i = \log \alpha_i$  for  $u_i$  and  $v_i$ , respectively, note that the interpolation functions are  $u(\beta)$  and  $v(\beta)$ .

*Step 3:* calculate the point on the curve  $(u(\beta), v(\beta))$  with the greatest curvature and let it be  $(u(\beta_0), v(\beta_0))$ , note that  $\alpha_0 = 10^{\beta_0}$ .

*Step 4:* solve the regularization problem corresponding to  $\alpha_0$  and note that the regularization solution is  $x^{\alpha_0,\delta}$ . From the regularization solution, we can obtain the new point pair  $(u_0, v_0)$ , of these  $u_0 = \|Kx^{\alpha_0,\delta} - y^\delta\|$ ,  $v_0 = \|x^{\alpha_0}\|$ .

*Step 5:* add the new point pairs  $(u_0, v_0)$  and  $\alpha_0$  to the previous point pairs and parameters.

*Step 6:* repeat steps 2 to 5 for the point pairs and parameters obtained in step 5 until convergence.

**3.3. Model Construction.** The LSTM is applied to the time dimension of the training set, the GL is applied to the spatial dimension, and the L-curve criterion is introduced to take values for the regularization hyperparameters. Based on this, an improved model is used to interpolate the data for the test set of time series. The schematic diagram of the improved model is shown in Figure 1.

As shown in Figure 1, the powerful learning capability of the LSTM is used to discover concealed correlations in the data. The GL space regularizer computes discrete approximations from the nearest neighbor graph of the scatter of data points, determination of the corner angle of the L-curve by means of cubic spline interpolation and hence the value of the regularization hyperparameter.

The regularizer is able to combine complex spatial and temporal dependencies into the decomposition process by setting up an error calculator, embedding the current time and interpolating the missing data in the current observation matrix. The complete time-series dataset is used as the training set to train the model to access the model parameters, and then the dataset containing the missing data is used as input to the model. The data interpolation result is the interpolation result considering the temporal spatial correlation.

## 4. Simulation Analysis

This section presents the experiments conducted in order to evaluate the algorithm proposed in this work.

### 4.1. Simulation Preparation

**4.1.1. Dataset.** The training set selected for this thesis is obtained by collation in the WMS intelligent warehouse management system. The software in the WMS intelligent warehouse management system refers to the software part that supports the operation of the whole system, including pick operation, shelf management, receive processing, replenishment management, matrix charging, platform management, warehouse operation, cycle counting, over-stock operation, RF operation, and process management, [31]. The WMS process is shown in Figure 2.

The data are in the form of  $4 * 744$ . The number 4 denotes four warehouses and 744 denotes a record of stock balances for each hour of the day for a month. Four warehouses are located in different areas of the chosen city. The four warehouses are in a circular configuration, three of which are closer to each other and on one side of the ring, the

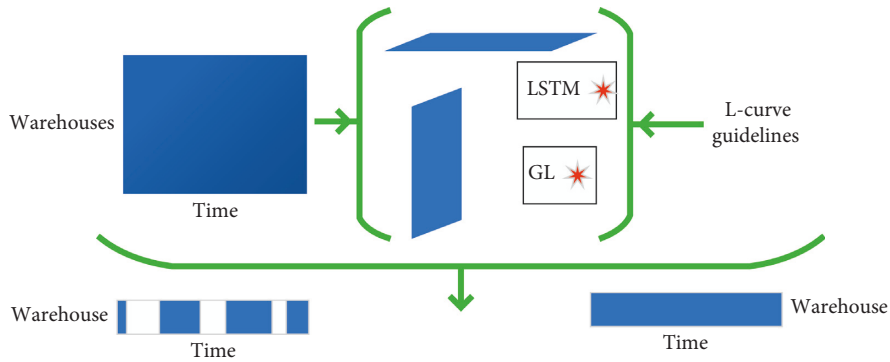


FIGURE 1: Illustration of improved LSTM-GL-ReMF.

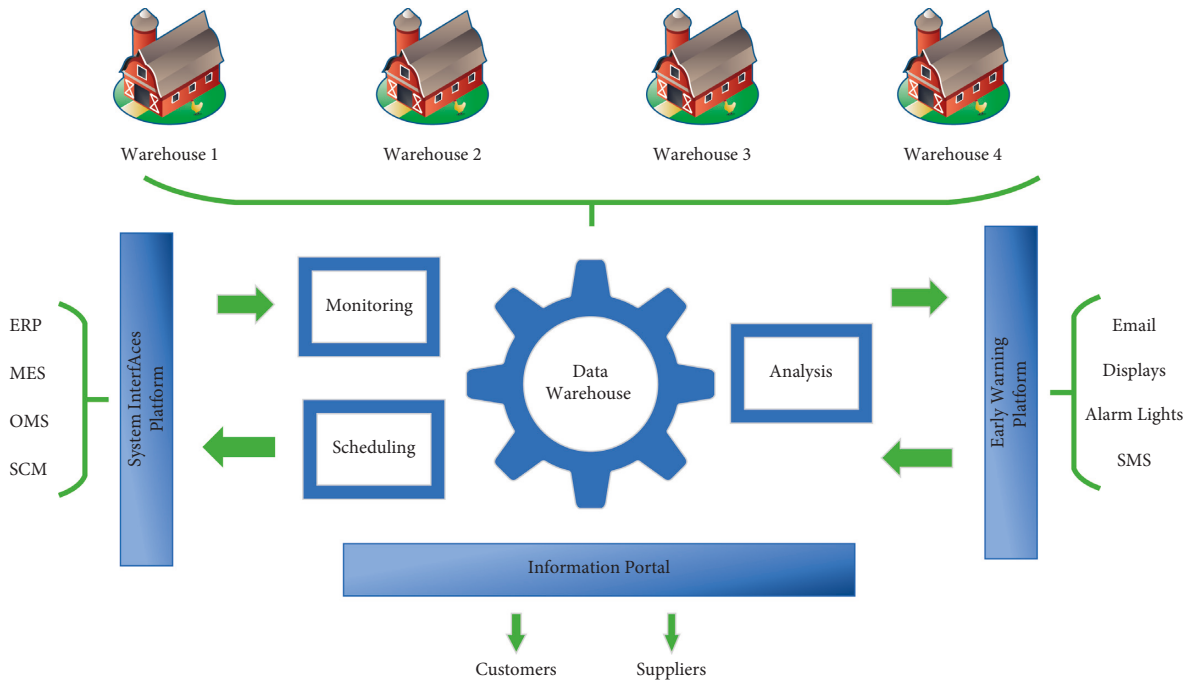


FIGURE 2: The WMS process.

**Input** Scoring Matrix  
**Step 01:** Using a mechanism to put-back to draw some data from the initial sample  
**Step 02:** Based on the above data, optimize the objective function and update the parameters  
**Step 03:** Duplicate data extraction  
**Output** User matrix and Item matrix

ALGORITHM 1: Matrix decomposition.

remaining warehouse is on the other side of the ring structure. The test set is one month’s stock balance data for certain good in a warehouse extracted directly from the WMS intelligent warehouse management system.

4.1.2. *Evaluation Indicators.* The commonly used evaluation metric in machine learning regression problems is the RMSE, which is more intuitive as a model evaluation metric in terms of order of magnitude. In this paper, there are no

zeros in the original data but to prevent the existence of zero in the interpolated values, MAPE is also chosen as the evaluation metric [32].

The MAPE was chosen, the number of measurements was  $n$ , and the number of measurement set scores was  $\hat{y}_i$ . The specific formula was as follows:

$$MAPE = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100. \tag{17}$$

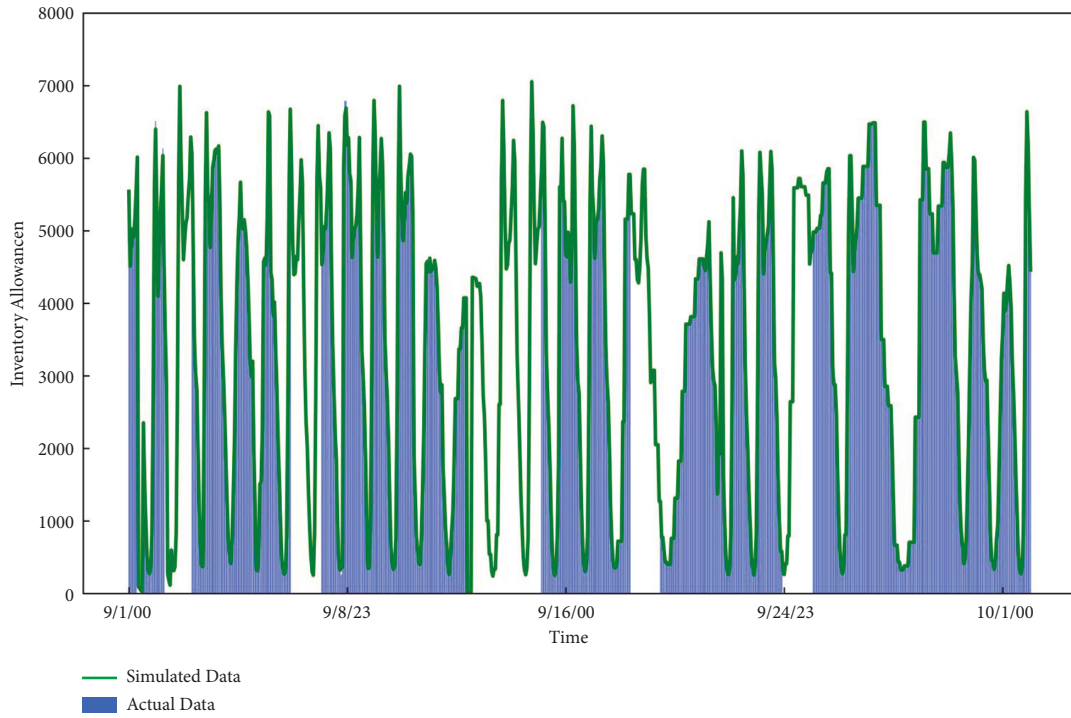
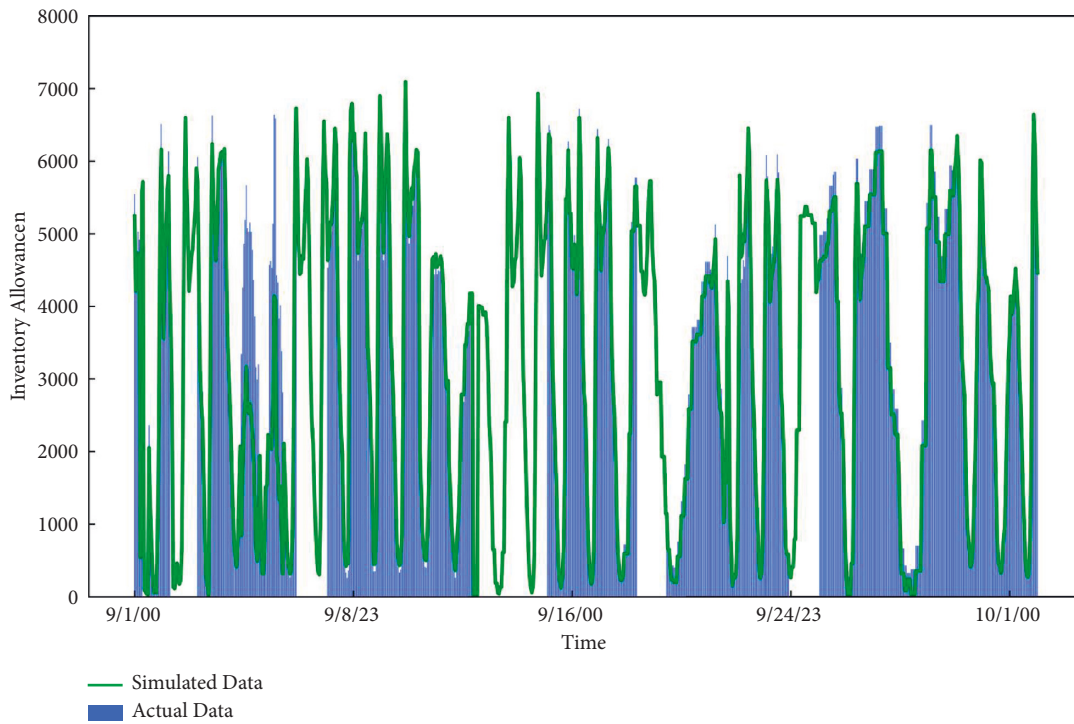
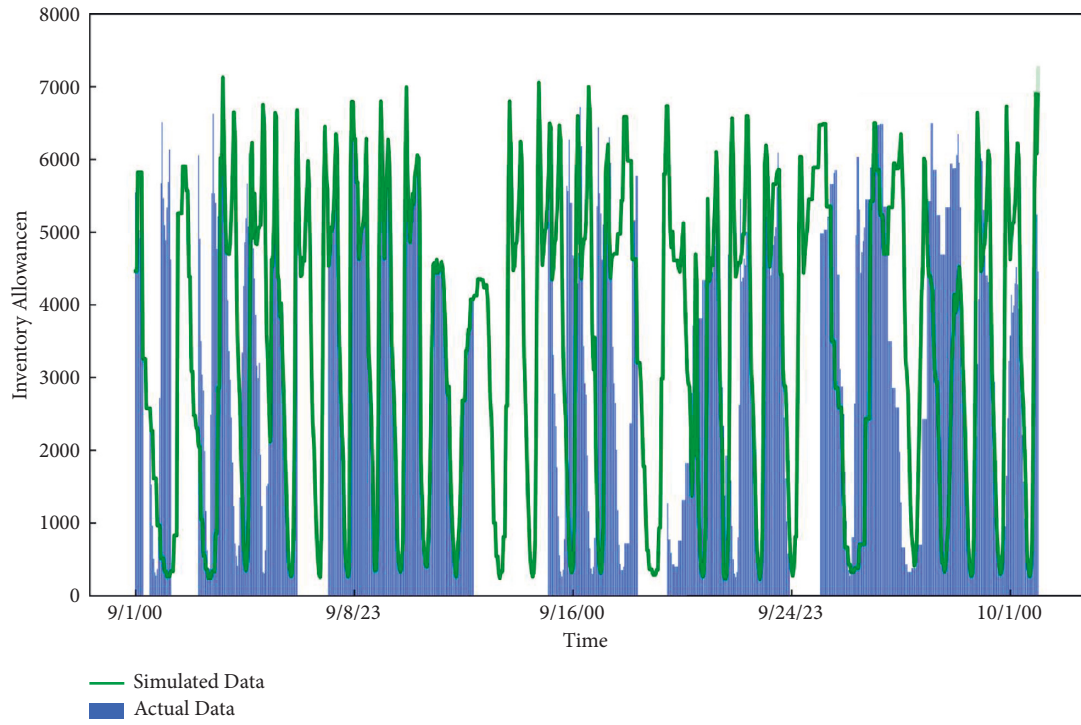


FIGURE 3: Improved LSTM-GL-ReMF model interpolation results.

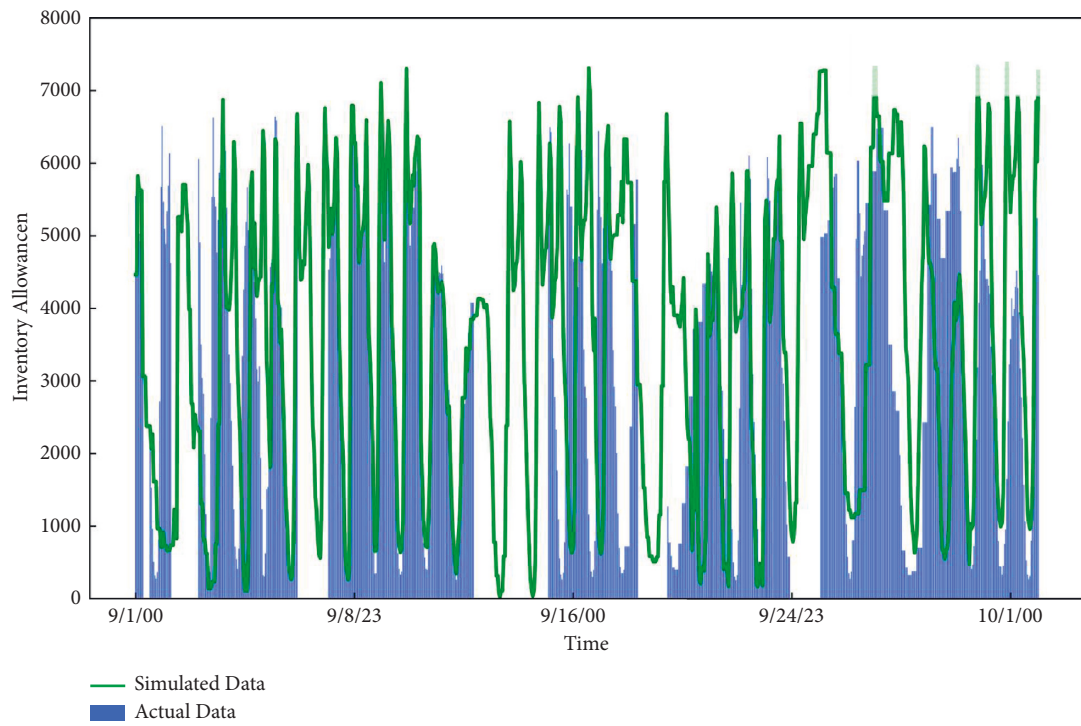


(a)

FIGURE 4: Continued.

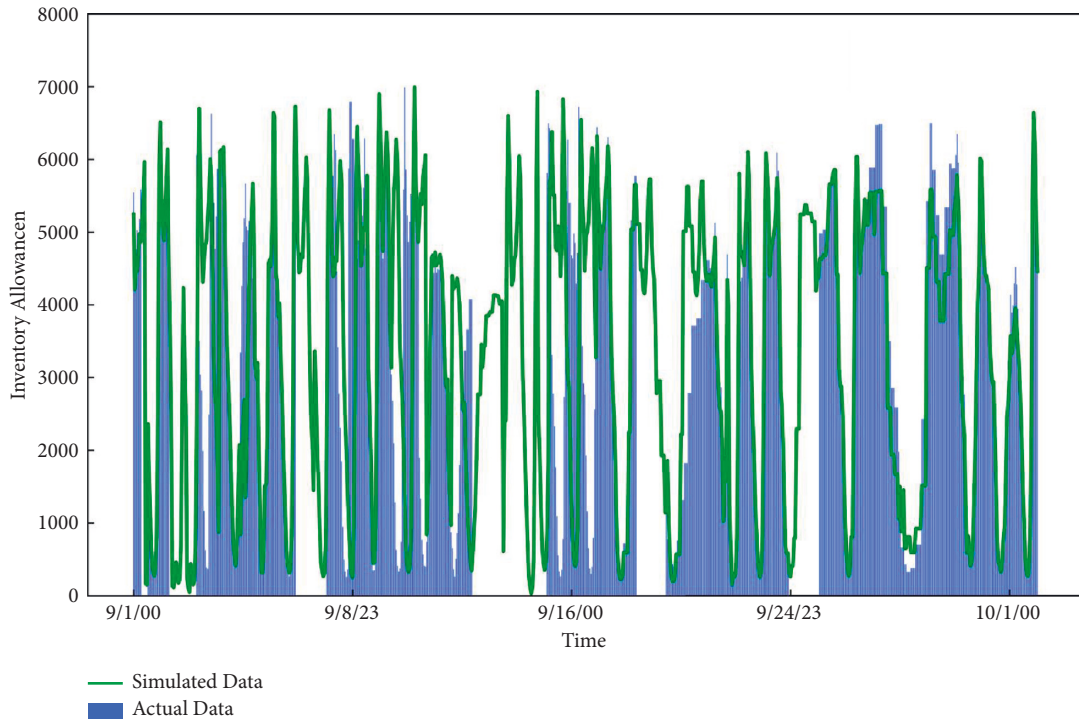


(b)

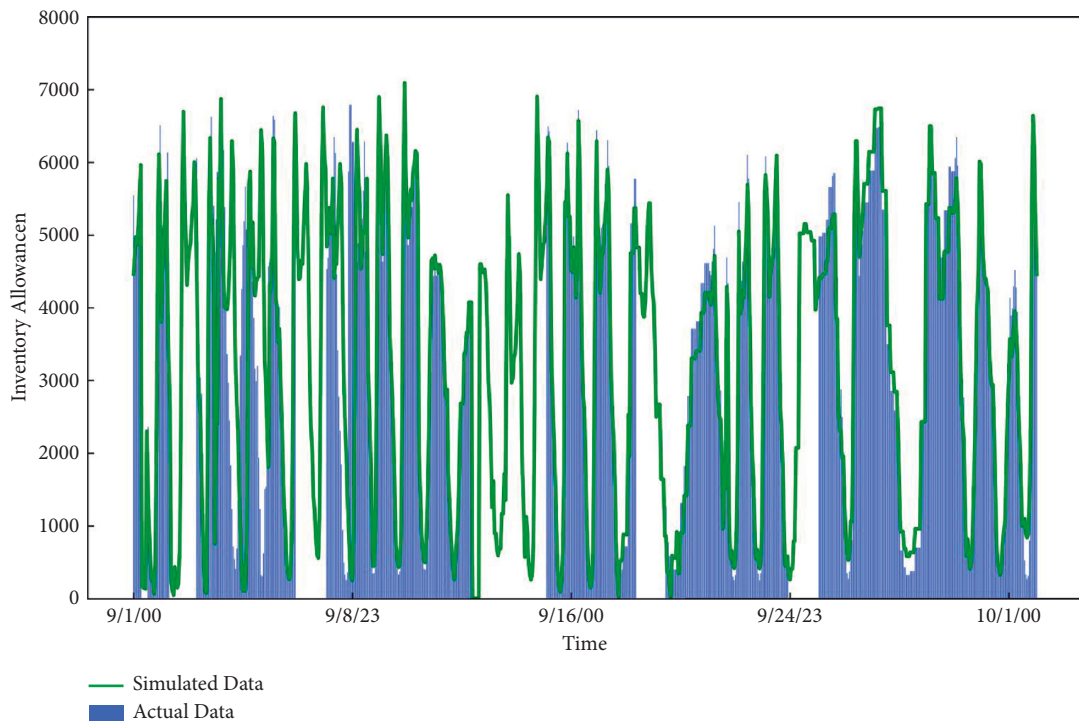


(c)

FIGURE 4: Continued.



(d)



(e)

FIGURE 4: Continued.



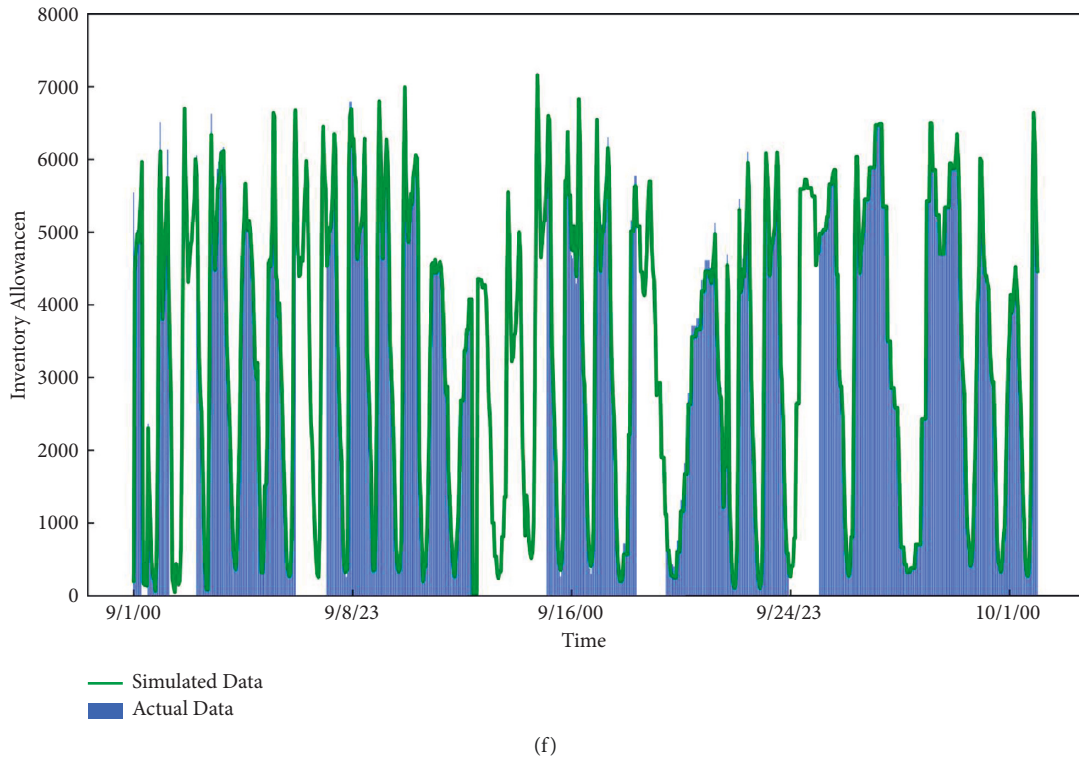


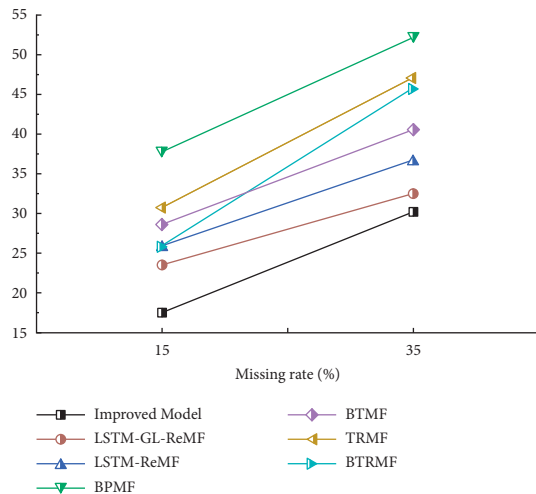
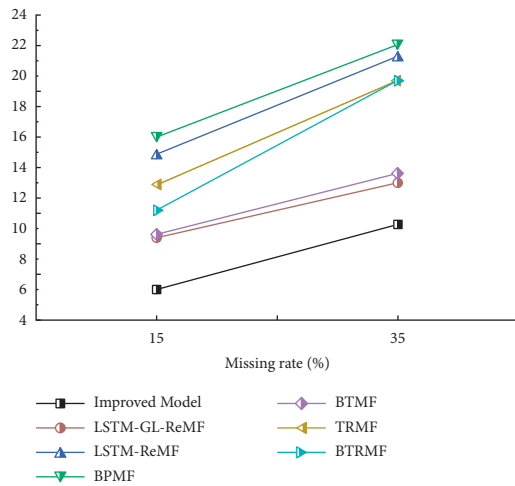
FIGURE 4: Compare the results of the models: (a) TRMF, (b) BTMF, (c) BPMF, (d) BTRMF, (e) LSTM-ReMF, and (f) LSTM-GL-ReMF.

TABLE 1: Table of values for each parameter.

Parameters	Values
Missing rates/%	25
Regularization parameter	50
Lambda	50
Sampling rate	1.0
Maxiter	100
Patience	6

TABLE 2: Model evaluation.

Evaluation indicators	MAPE/%	RMSE
Improved model	27.24	9.22
LSTM-GL-ReMF	30.52	12.16
LSTM-ReMF	27.91	16.13
BPMF	47.51	20.15
BTMF	33.27	11.23
TRMF	40.17	16.85
BTRMF	28.67	12.37



(a)

(b)

FIGURE 5: Model evaluation for different absence rates. (a) RMSE. (b) MAPE (%).

The RMSE was also chosen with the following equation [33]:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (18)$$

**4.2. Simulation Studies.** To validate the performance of the proposed improved model, a dataset of one month's inventory balance of certain good in a warehouse was selected and set up in an Intel i5-7200u cpu @2.70 GHz processor and Jupyter Notebook environment. Python3 was used for the experiments. The values of some of the parameters are shown in the table for each parameter in Table 1.

To verify the effectiveness and superiority of the improved model, simulations are conducted and the simulation results are quantitatively profiled in this paper. The simulation results are shown in Figure 3 for the improved LSTM-GL-ReMF model interpolation.

The interpolation results of the improved LSTM-GL-ReMF model are shown in the figure, where the blue bars indicate the actual data. Green straight lines indicate the simulated data, which can clearly show that the model interpolates the missing data. The interpolation results show that the interpolated lines overlap with the true values to a high degree. Because of the spatial correlation, the existence of zero values in the interpolated part is in line with the actual situation. The fact that the interpolated stock balance fluctuates greatly due to the different demand for goods by customers in different regions at different times of the year is also in line with the actual situation. The interpolated values for the end of September 2021 in this improved model have a zero value, which may be related to regional user activities such as end of month clearance and delayed start of the school season due to the epidemic.

A comparison is made with other models. The problem studied in this paper is a regression problem and the Bayesian correlation algorithm is highly used in regression problems and the time matrix is also a commonly used model for time series. Bayesian probabilistic matrix factorization (BPMF), temporal regularized matrix factorization (TRMF), Bayesian temporal matrix factorization (BTMF), Bayesian temporal regularized matrix factorization (BTRMF) [34], LSTM-ReMF, and LSTM-GL-ReMF are chosen for comparative analysis. The simulation results are shown in Figure 4.

The blue bar charts in Figure 4 show the actual data. Vacancy parts show the missing data. Green straight lines show the model simulation data, which can clearly show the interpolation of the missing data by the model. RMSE, MAPE was used as an evaluation indicator and the results are shown in Table 2.

To further investigate the generalization ability of the model, simulation experiments were conducted for data with missing rate of 15% and 35%, respectively. Error evaluation was obtained as shown in Figure 5.

From the above results, it can be seen that the interpolation results with the inclusion of spatiotemporal analysis are more realistic. The improved model in this paper performs better compared to the original LSTM-GL-ReMF model, but is still affected by the missing rate of data, with the larger missing rate the larger the model error. The comprehensive simulation results show that the improved LSTM-GL-ReMF model can effectively interpolate the missing inventory balance data in the logistics field with high performance. From a practical point of view, there is a relationship between inventory balances and interregional user behavior in different regions of the same city. The improved model in this paper has very good performance.

## 5. Conclusion

In this paper, a method for interpolating missing data in logistics based on regularized decomposition is proposed. There is empirical evidence that our research has some contribution to make. Regularization techniques are used to interpolate missing data in logistics to improve the LSTM-GL-ReMF model by considering spatiotemporal data characteristics and apply L-curve method to select regularization parameters. It is provided proof for the introduction of missing data interpolation models into the logistics domain. In addition, a rigorous theoretical derivation of the model's operating mechanism has been carried out, which provides theoretical support for the model algorithm. The results show that the interpolation model, which takes into account the spatial correlation between the inventory balances of different same-city physical warehouses, is closer to its real situation and has a higher interpolation accuracy. Future experiments will consider different types of missing data to find the range of missing data with the best interpolation capability and thus improve the model interpolation performance.

## Data Availability

The data used to support the findings of this study have not been made available because it is confidential within the company.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] B. Yi, "The development of smart warehousing in 2020 and future outlook," in *China's Warehousing and Distribution Industry Development Report in 2021*, pp. 108–122 + 9, Blue Book, Beijing, China, 2021.
- [2] N. U. Handayani, D. P. Sari, Y. Widharto, and G. Basyir, "Requirements analysis for the disaster logistics inventory information system to improve the effectiveness and efficiency of handling emergency response periods," *IOP Conference Series: Materials Science and Engineering, Bali Indonesia*, vol. 6, no. 4, pp. 212–224, 2019.

- [3] H. Ahn, K. Pio Kim, and K. P. Kim, "Comparison of missing data imputation methods in time series forecasting," *Computers, Materials & Continua*, vol. 70, no. 1, pp. 767–779, 2022.
- [4] X. Zhang, *Research on data repair and short-time traffic flow prediction methods for urban traffic flow*, PhD Dissertation, Lanzhou University of Technology, Lanzhou, China, 2021.
- [5] I. A. M. Jafri, N. M. Noor, A. Z. Ul-Saufie, and A. Suwardi, "Prediction of missing data in rainfall dataset by using simple statistical method," *IOP Conference Series: Earth and Environmental Science*, vol. 616, no. 1, Article ID 012005, 2020.
- [6] H. F. Yu, N. Rao, and I. S. Dhillon, "High-dimensional time series prediction with missing values," *Journal of Physics Conference Series*, vol. 179, no. 1, Article ID 012069, 2015.
- [7] S. Guan, "Smart E commerce logistics construction model based on big data analytics," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 3015–3023, 2021.
- [8] S. Mu, D. Baig, S. Alam, and K. M. Alam, "A correlative analysis of modern logistics industry to developing economy using the VAR model: a case of Pakistan," *Journal of Advanced Transportation*, vol. 2020, Article ID 8861914, 10 pages, 2020.
- [9] C. S. Veelenturf and L. P. Veelenturf, "The strategic role of logistics in the industry 4.0 era," *Transportation Research Part E: Logistics and Transportation Review*, vol. 129, pp. 1–11, 2019.
- [10] C. Wang and Q. Wang, "Research on commercial logistics inventory forecasting system based on neural network," *Neural Computing & Applications*, vol. 33, no. 2, pp. 691–706, 2021.
- [11] Y. M. Tang, K. Y. Chau, W. Wan, and T. W. Wan, "Forecasting economic recession through share price in the logistics industry with artificial intelligence (AI)," *Computation*, vol. 8, no. 3, p. 70, 2020.
- [12] Q. Chen, "Coordinated development of urban intelligent transportation data system and supply chain management," *Journal of Advanced Transportation*, vol. 2022, Article ID 6308728, 11 pages, 2022.
- [13] X. Wang, Y. Ma, S. Xu, and Y. Xu, "Data imputation for detected traffic volume of freeway using regression of multilayer perceptron," *Journal of Advanced Transportation*, vol. 2022, Article ID 4840021, 16 pages, 2022.
- [14] Y. Liu, F. Wu, C. Lyu, S. Li, J. Qu, and X. Qu, "Deep dispatching: a deep reinforcement learning approach for vehicle dispatching on online ride-hailing platform," *Transportation Research Part E: Logistics and Transportation Review*, vol. 161, Article ID 102694, 2022.
- [15] X. Chen, Z. Sun, and L. Sun, "A bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 73–84, 2019.
- [16] A. González-Vidal, P. Rathore, A. S. Rao, J. Mendoza-Bernal, M. Palaniswami, and A. F. Skarmeta-Gómez, "Missing data imputation with bayesian maximum entropy for internet of things applications," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 16108–16120, 2020.
- [17] X. Y. Zou, "Time series prediction algorithm based on graph Laplace transform and limit learning machine," *Computer Applications and Software*, vol. 38, no. 4, pp. 288–294, 2021.
- [18] F. Sun, M. Xu, X. Hu, and X. Jiang, "Graph regularized and sparse nonnegative matrix factorization with hard constraints for data representation," *Neurocomputing*, vol. 173, pp. 233–244, 2016.
- [19] Y. Liu, C. Lyu, Y. Zhang, Z. Liu, W. Qu, and X. Qu, "DeepTSP: deep traffic state prediction model based on large-scale empirical data," *Communications in Transportation Research*, vol. 1, Article ID 100012, 2021.
- [20] J. M. Yang, Z. R. Peng, and L. Lin, "Real-time spatiotemporal prediction and imputation of traffic status based on LSTM and Graph Laplacian regularized matrix factorization," *Transportation Research Part C*, vol. 129, 2021.
- [21] V. A. Boeva, "L-curve method for evaluating the optimal parameter of a smoothing cubic spline," *Mezhdunar. nauch. issled. zurn.*, vol. 11, no. 113, pp. 6–13, 2021.
- [22] Z. Xing, Y. Ma, X. Yang, and F. Nie, "Graph regularized nonnegative matrix factorization with label discrimination for data clustering," *Neurocomputing*, vol. 440, pp. 297–309, 2021.
- [23] J. Q. Ji and Y. Chung, "Missing data modeling based on matrix factorization of implicit feedback dataset," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 23, no. 5, pp. 495–507, 2019.
- [24] W. W. Zhu, J. L. Wu, T. Fu, J. H. Wang, J. Shangguan, and Q. Shangguan, "Dynamic prediction of traffic incident duration on urban expressways: a deep learning approach based on LSTM and MLP," *Journal of intelligent and connected vehicles*, vol. 4, no. 2, pp. 80–91, 2021.
- [25] H. D. Nguyen, K. P. Tran, S. Hamad, and M. Hamad, "Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management," *International Journal of Information Management*, vol. 57, Article ID 102282, 2021.
- [26] Y. Z. Zhu and Z. G. Chen, "Fault diagnosis of rolling bearings based on graph Laplacian matrix and improved K-mean clustering," *Electronic Measurement Technology*, vol. 44, no. 10, pp. 45–51, 2021.
- [27] C. Qiao, X. Y. Hu, L. Xiao, V. D. Wang, and Y. P. Wang, "A deep autoencoder with sparse and graph Laplacian regularization for characterizing dynamic functional connectivity during brain development," *Neurocomputing*, vol. 456, pp. 97–108, 2021.
- [28] G. Zhou, M. Yuan, X. Li et al., "Optimal regularization method based on the L-curve for solving rational function model parameters," *Photogrammetric Engineering & Remote Sensing*, vol. 87, no. 9, pp. 661–668, 2021.
- [29] I. A. Blatov and A. I. Zadorin, "Application a cubic spline to calculate derivatives in the presence of a boundary layer," *Journal of Physics: Conference Series*, vol. 179, no. 1, pp. 10–12, Article ID 012069, 2021.
- [30] E. X. Zheng, "Several regularization methods for ill-posed problems and their numerical realization," Master Thesis, Jilin University, Jilin, China, 2009.
- [31] F. Ud Din, D. Paul, J. Ryan, F. Wallis, and M. Wallis, "Aosr 2.0: a novel approach and thorough validation of an agent-oriented storage and retrieval WMS planner for SMEs, under industry 4.0," *Future Internet*, vol. 13, no. 6, p. 155, 2021.
- [32] S. Prayudani, A. Hizriadi, Y. Y. Lase, and Y. Fatmi, "Analysis accuracy of forecasting measurement technique on random K-nearest neighbor (RKNN) using MAPE and MSE," *Journal of Physics: Conference Series, Medan Indonesia*, vol. 1361, no. 1, Article ID 012089, 2019.
- [33] V. B. Kamble and S. N. Deshmukh, "Comparison between accuracy and MSE, RMSE by using proposed method with imputation technique," *Oriental Journal of Computer Science and Technology*, vol. 10, no. 4, pp. 773–779, 2017.
- [34] P. Ren, X. Chen, L. Sun, and H. Sun, "Incremental Bayesian matrix/tensor learning for structural monitoring data imputation and response forecasting," *Mechanical Systems and Signal Processing*, vol. 158, Article ID 107734, 2021.