Computer Fraud and Security
ISSN (online): 1873-7056

Dynamic Time Series Modeling Method based on Bayesian Learning and Gray Forecasting

Botao Zhang[1#], Aiying Zhao[1#], Wenjing Wang[2], Tiande Wang[1], Hanyu Wang[2], Junping Yan[1*], Jiaxing Wang[3]

1 College of Sciences, ShiHeZi University, Shihezi 832003, China

2 College of information Science and Technology, ShiHeZi University, Shihezi 832003, China

3 College of life Sciences, ShiHeZi University, Shihezi 832003, China

[#]Botao Zhang and Aiying Zhao contributed equally to this work and they are both first authors.

Corresponding Author: Junping Yan, Email: 19556706259@163.com

Abstract: As real-life time series data are usually complex, the traditional dynamic time series model cannot achieve satisfactory prediction results. In this paper, we combine gray prediction with dynamic time series to construct a dynamic time series model based on gray prediction to improve the prediction accuracy of dynamic time series. Aiming at the missing data problem often faced in the process of dynamic time series modeling, we design the missing data recovery algorithm for dynamic time series based on Bayesian learning, and model the missing data prediction problem in dynamic time series as a multi-sparse vector recovery problem based on the theory of compressed sensing. Carrying out dynamic time series prediction simulation experiments, the residual sum of squares, residual median error, and average relative error of the dynamic time series model based on gray prediction in this paper are 0.616, 0.307, and 0.297, respectively, which are lower than those of the comparative traditional GM(1,1) model, and the time series analysis model. And in the missing data recovery simulation experiments, this paper's algorithm has a smaller RMSE at any data missing rate, and the RSME value of this paper's algorithm remains lower than 0.2 when the data missing rate reaches the highest 95%.

Keywords: dynamic time series; Bayesian learning; gray prediction; compressed perception theory

1 Introduction

Obtaining data through a series of observations at a point in time is a commonplace activity, and in social, economic and other activities, many data are generated based on chronological order, such as weekly interest rates, daily stock closings, monthly price indices, annual sales volumes, etc., as observed in business [1-3]. Therefore, time series are widely covered within the fields of finance and economics, meteorology and hydrology, engineering and technology, natural sciences, social sciences, etc [4]. The study of such time-varying data has given rise to the discipline of time series analysis, which has contributed to the development of time series modeling and forecasting. Time series analysis is the theory and methodology established to analyze systematically observed time series data in order to seek its trend of change [5]. Statistical patterns of the object of study are obtained through time series analysis, so as to predict its possible future values and make decisions about possible future problems for the purpose of controlling the whole system [6-8]. So far, this discipline has been fully developed, with a complete theoretical system and strong application, which is a good means to solve probability statistics and other related problems.

Since the linear operation is closed in the Bayesian space, discussing the time series modeling problem under the algebra of Bayesian space can effectively overcome the adverse effects of various constraints on the time series data [9-10]. And based on the linear operation and inner product operation in Bayesian space, it can give the definition of numerical characteristics of time series data, which lays the foundation for subsequent statistical analysis [11-13]. In the face of the uncertainty problem of "little data" and "poor information", the gray system theory occupies an important low-dimensional time series modeling in the absence of data [14-15]. The advantage of gray time series modeling is that there are no special requirements and restrictions on data, and it mainly extracts valuable information by mining part of the known information, so as to realize the correct description and effective supervision of the system's operation behavior and evolution law [16-18]. Therefore, the combination of Bayesian learning and gray forecasting to build a time series analysis model, and determine the dynamic relationship between the trend time series, to provide managers with scientific and reasonable basis and methods.

In order to solve the problem that the traditional dynamic time series model can not accurately predict the complex data, this paper combines the gray prediction with the dynamic time series, establishes the gray prediction model according to the collected data, constructs the dynamic gray model on the basis of which the trend term in the time series is fitted, generates the smooth time series by using the inverse order method, determines the retrospective order in the time series model as well as the parameter of the model, and establishes the time The time series model is established. The optimal criterion function fixed-order method is used to determine the backward order in the model, and the dynamic gray model is used to predict the trend term in the sequential residual column, fit the dynamic time series fluctuations, and construct a time series combination model based on gray prediction. The missing data problem encountered in the dynamic time series modeling process, this paper introduces the sparse Bayesian learning principle in Bayesian network theory. Applying the theory of compressed perception, the missing data prediction problem in time series is modeled as a multiple sparse vectors recovery problem, and an efficient recovery algorithm based on sparse Bayesian learning is designed for the characteristics of dynamic time series, which obtains part of the support information through learning, thus solving the recovery problem of multiple sparse vectors at the same time. The prediction performance of the dynamic time series model proposed in this paper is examined by means of simulation experiments, and the data recovery capability of the dynamic time series missing data recovery algorithm is explored. Finally, the application practice of dynamic time series prediction of raw coal production is carried out to study the effect of the dynamic time series model based on Bayesian learning and gray prediction constructed in this paper in the real time series prediction work.

2 Dynamic time series model based on gray forecasting

Gray prediction model has the characteristics of small amount of required data and good macro change trend of predicted data, but it is not adaptable enough in the face of data fluctuations and cannot analyze the causes and consequences of data changes [19]. At the same time, due to the complex and dynamic changes in the time series scenario, and the existence of seasonal components, trend components and other complex structural features in the time series data, the traditional time series can not meet the increasingly diverse and complicated data analysis tasks. For this reason, this paper will combine gray prediction with time series to achieve dynamic time series modeling, enhance the adaptability to data fluctuations, and improve the accuracy of time series prediction.

2.1 Principle of gray prediction model

2.1.1 Gray prediction model

Gray model (GM) is through a small amount of incomplete information, the establishment of gray differential prediction model, the development of the law of things to make a fuzzy long-term description, is a fuzzy prediction in the field of the theory and method of the more perfect branch of forecasting.

There are primitive series:

$$\{X^{(0)}(i)\}, i=1,2,\cdots n \tag{1}$$

Add them up to generate a new series:

$$\{X^{(1)}(i)\}, i=1,2,\cdots n \tag{2}$$

The corresponding differential equation is:

$$\frac{dX^{(1)}}{dt}+aX^{(1)}=u \tag{3}$$

The cumulative matrix is:

$$B=\begin{bmatrix} -\frac{1}{2}[X^{(1)}(1)+X^{(1)}(2)] & \cdots & 1 \\ -\frac{1}{2}[X^{(1)}(2)+X^{(1)}(3)] & \cdots & 1 \\ \vdots & & \vdots \\ -\frac{1}{2}[X^{(1)}(n-1)+X^{(1)}(n)] & \cdots & 1 \end{bmatrix} \tag{4}$$

The constant vector is:

$$Y_n=[X^{(0)}(2),X^{(0)}(3),\cdots,X^{(0)}(n)]^{\mathrm{T}} \tag{5}$$

The coefficients of the solution are obtained by the least squares method:

$$\alpha=\begin{bmatrix} a \\ u \end{bmatrix}=(B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}Y_n \tag{6}$$

A further derivation reduces to:

$$\dot{X}^{(0)}(t+1) = -a\left(X^{(0)}(1) - \frac{u}{a}\right)e^{-au} \tag{7}$$

Equation (7) is the prediction equation of the GM(1, 1) model. At this point its actual predicted value is Eq. (8):

$$\dot{X}^{(0)}(t+1) = \dot{X}^{(1)}(t+1) - \dot{X}^{(1)}(t) \tag{8}$$

2.1.2 Dynamic Gray Model

A dynamic gray model is a model that uses the latest monitoring data to guide and predict subsequent monitoring. The process of constructing a dynamic gray model is described in detail below.

In the absence of known information, the original series can be morphed into the following form:

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(N)\} \tag{9}$$

For the sequence in Eq. (9) follow the modeling process above to build the GM(1, 1) model, which in turn leads to the predicted value $\dot{X}^{(0)}(N+1)$ at the moment of $(N+1)$, which is added to the original modeling series, then the new modeling sequence is Eq. (10):

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(N), X^{(0)}(N+1)\} \tag{10}$$

Then the next prediction is made for the GM(1, 1) model, and so on.

Modeling prediction according to the above method, it will be found that the data of modeling sequence $X^{(0)}$ will become more and more large, and the prediction accuracy will be affected. In order to solve the problem of larger data volume, this paper adopts equal-dimensional data for constraints, i.e., the latest monitoring data is used to replace the earlier monitoring information in the monitoring cycle for modeling. The new modeling sequence can be obtained after doing isodimensional data processing on the original monitoring data, as shown below:

$$X^{(0)} = \{X^{(0)}(2), \cdots, X^{(0)}(N), \hat{X}^{(0)}(N+1)\} \tag{11}$$

2.2 Dynamic time series modeling

2.2.1 Smoothness test for dynamic time series

When modeling time series for application data, it is necessary to monitor the series as a smooth time series, so the following focuses on the process of generating a smooth time series using the reverse order method. The specific process is as follows:

1) First, establish a variance series based on the residual values of the predicted values derived from the gray model, and divide the time series into M segment according to a certain length of time based on the length of the variance series, assuming that the resulting series is $y_1, y_2, \cdots, y_M$.

2) Calculate the total number of backward sequences of the mean sequence obtained above. For a $y_i(i=1,2,\cdots,M-1)$ -mean sequence, if the value of the later sequence is greater than the value of the previous sequence, i.e., $y_j(j>i)$, the comparison process is said to be an inverse order count. Assuming that the number of inverse sequences of a sequence is A_i, the total number of inverse sequences of the mean value sequence is:

$$A=\sum_{i=1}^{M-1}A_i \tag{12}$$

3) Calculate statistics for statistical tests:

$$E(A)=\frac{1}{4}M(M-1) \tag{13}$$

$$D(A)=\frac{M(2M^2+3M-5)}{72} \tag{14}$$

And there are statistics:

$$Z=\frac{\left[A+\frac{1}{2}-E(A)\right]}{\sqrt{D(A)}} \tag{15}$$

asymptotically obeys the $N(0,1)$ distribution. Therefore, for the mean series $y_1,y_2,\cdots,y_M$ can rely on equations (12), (13) and (14) to find out the value of statistic Z. In the case of significant level $\alpha=0.05$, if $|Z|<1.96$, the mean series is considered to have no significant trend, i.e., the mean series is a smooth time series; otherwise, it is considered that the mean series is non-smooth. After the verification of the smooth time series, it is also necessary to zero-mean the original series as follows:

$$e_i^{'}=e_i-\bar{e}_i \tag{16}$$

2.2.2 Dynamic time series modeling

After the smoothing and homogenization of the dynamic time series, the time series can be modeled starting from the use of a certain length of smooth series samples. The first step is to determine the order of backtracking in the time series model and the parameters to be taken in the model, and then build the time series model, which is shown by the inverse function [20]:

$$X_t=I_1X_{t-1}+I_2X_{t-2}+\cdots+a_t=\sum_{j=1}^{\infty}I_jX_{t-j}+a_t \tag{17}$$

In order to reduce the cumbersome computational process and at the same time ensure that the prediction accuracy of the model will not be degraded Ben, this paper utilizes the higher-order $AR(n)$ model as an alternative to the time series model.

2.2.3 Dynamic time series model ordering

The two key parameters for establishing a time series model have been mentioned above, and for the method of determining the number of retrospective orders in the model, this paper adopts the optimal criterion function to determine the order method. The optimal criterion function method, that is, through a criterion function to determine the time series model of the retrospective order in the case of meeting the forecasting requirements, to the optimal value, and so that the criterion function to reach a very small value of the retrospective order of the model is the best model.

For time series $\{X_t, 1 \le t \le N\}$, this paper describes it with a higher order $AR(n)$ model, defining the AIC criterion function as follows:

$$AIC(n) = \ln \sigma_a^2(n) + \frac{2n}{N} \tag{18}$$

The highest order of the backtracking order $M(N)$ is usually some number between $[N/3]-[2N/3]$. After determining $M(N)$, the number of backtracking orders is determined by the following equation, if the following equation holds:

$$AIC(n_0) = \min_{1 \le n \le M\ |N|} AIC(n) \tag{19}$$

Then n_0 is taken as the backward order of the time model. However, when the length of the time series is not enough or the lagged data is more, the retrospective order determined by the AIC method will have a large deviation; therefore, this paper chooses another criterion calculation method, which is referred to as AICC. the expression is as follows:

$$AICC(n) = N\ln\hat{\sigma}_a^2(n) + \frac{N+n}{N-n-2} \tag{20}$$

2.3 Dynamic time series combination model based on gray forecasting

The key point of the combination of time series and equal-dimensional dynamic gray model is that the equal-dimensional dynamic gray model predicts the trend term in the column of ordinal residuals to build a time series model to fit the fluctuations in the series, and finally the fitted value of the time series plus the dynamic predicted value of the gray model is used to correct the subsequent prediction results and to improve the accuracy of the prediction. The combination can be written in the following form:

$$X_t = d_t + y_t \tag{21}$$

Where, d_t is the trend term of the equidimensional dynamic gray model, and at the same time the series is non-smooth need to be smoothed and homogenized; y_t is the smooth series after processing; in this paper, we use the equidimensional dynamic gray model to fit d_t, and use the higher-order $AR(n)$ to fit y_t.

3 Dynamic time series prediction simulation experiment

In the previous chapter, this paper proposed a dynamic time series model based on gray forecasting, in which gray forecasting is combined with time series in order to achieve tapping time series modeling and improve the forecasting accuracy of time series. In this chapter, simulation experiments of dynamic time series prediction will be conducted to verify the prediction performance performance of the dynamic time series model in this paper.

3.1 Data pre-processing

The data used in this simulation experiment are, the deformation simulation data of pit construction at monitoring point A of a subway station during the period of August 2024-September 2024, named Data1 dataset. The simulation experiment will use the dynamic time series combination model based on gray prediction proposed in this paper to carry out dynamic time series modeling based on the original data of the 62nd-91st period of the monitoring point A, and predict the cumulative settlement in the next 5 periods through the established model.The original data of the 62nd-91st period of the cumulative settlement of the monitoring point A are shown in Table 1.

Table 1 Original Cumulative settlement

Period	Cumulative settlement(mm)	Period	Cumulative settlement(mm)
62	2.4	77	10.03
63	3.75	78	9.98
64	4.84	79	10.07
65	5.2	80	11.07
66	5.22	81	11.25
67	6.18	82	9.88
68	7.57	83	10.01
69	7.94	84	10.4
70	8.02	85	10.7
71	8.97	86	11.4
72	8.51	87	12.61
73	8.57	88	12.78
74	8.63	89	11.96
75	8.83	90	12.65

76	9.07	91	13.23

The dynamic time series model in this paper was utilized to fit the data for the first 30 periods to generate the residual series specifically shown in Table 2.

Table 2 Cumulative settlement residuals

Period	Residual error(mm)	Period	Residual error(mm)
62	0	77	0.79
63	-2.3	78	0.55
64	-1.32	79	0.47
65	-1.27	80	1.3
66	-1.29	81	1.03
67	-0.58	82	-0.6
68	0.66	83	-0.87
69	0.71	84	-0.62
70	0.67	85	-0.69
71	1.37	86	-0.3
72	0.6	87	0.59
73	0.55	88	0.17
74	0.3	89	-0.86
75	0.21	90	-0.41
76	0.22	91	-0.42

3.2 Analysis of residual series

As the dynamic time series analysis model in this paper can fully explore the continuity and correlation of the time series data, and then understand the influence of the historical information within the data on the current information contained in the data. Therefore, the dynamic time series model of the data residual series can be obtained by modeling the random part of the original data through time series analysis, and now the random part of the original data (i.e., the residuals generated when extracting the trend term using the gray prediction) is tested for smoothness and zero-meanness. The plot of the residual series is specifically shown in Fig. 1. From the figure, it can be seen that after extracting the trend term of the original observation series through the dynamic time series analysis model in this paper, the residual series appears to be smoother compared to the original data.

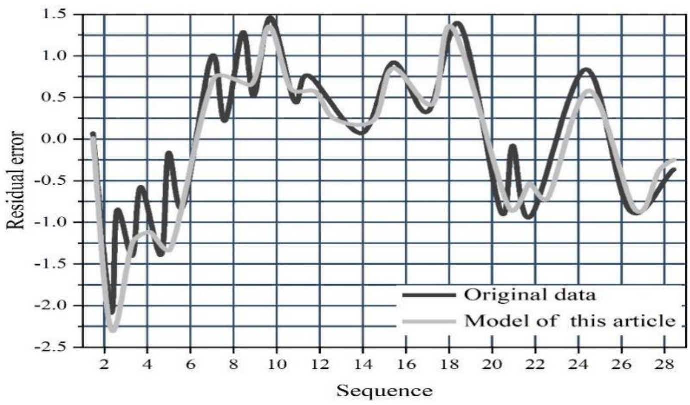


Figure 1 Residual sequence diagram

Continue to do unit root test on the residual series to finalize its smoothness, as shown in Table 3. After the unit root test, according to the P-value and T-statistic, the residual series after differencing is confirmed to be smooth, with a P-value of 0.0016<0.01, showing a significant difference, which confirms that the residual series after differencing is smooth and meets the conditions of dynamic time series modeling.

Table 3 Unit root test

-		T	P
Augmented Dickey-Fuller test statistic		-4.51388	0.0016
Test critical values	1%level	-3.66887	-
	5%level	-2.96235	-
	10%level	-2.62653	-

3.3 Analysis of dynamic time series prediction results

The dynamic time series model of this paper is used to analyze and forecast the data of period 62-91, and obtain the predicted values of the last 5 periods of 92-96. The prediction results of this paper's model are compared with those of the traditional GM(1,1) model and the time series analysis model to explore the prediction effect of each model. The prediction results of each model are specifically shown in Table 4. From the table, it can be seen that the difference between the predicted and measured cumulative settlement of this paper's model in period 92-96 is 0.1, 0.1, 0.11, 0.06, 0.11, respectively, and the difference is lower than that of the traditional GM(1,1) model and time series analysis model, and the best prediction results are obtained.

Table 4 Fitting results of the composite pattern

Period	Measured value	GM(1,1)model		Time series analyzes model		Model of this article	
		Predicted value	Residual error	Predicted value	Residual error	Predicted value	Residual error
92	13.71	13.99	0.25	13.1	0.27	13.61	0.35
93	14	14.42	0.44	13.72	0.22	14.1	0.15

94	14.76	14.89	0.11	14.08	0.7	14.65	0.41
95	14.05	15.29	1.17	14.35	0.29	14.11	41.03
96	14.65	15.74	1.08	14.63	-0.02	14.76	0.19

The prediction accuracy of the three models is further compared by calculating the residual sum of squares, residual median error, and average relative error, and the prediction results of each model are shown in Table 5. The residual sum of squares of this paper's model is 0.475, which is less than the traditional GM(1,1) model and the time series analysis model 2.4401 and 0.2462, respectively. In the residual medium error, this paper's model reaches 0.3079, which is lower than the traditional GM(1,1) model and the time series analysis model 0.4604 and 0.0675. In the average relative error, the difference between this paper's model and the traditional time series analysis model is smaller, which is only 0.01 less than that of the traditional GM(1,1) model, the traditional GM(1,1) model, and the traditional GM(1,1) model, and the traditional GM(1,1) model. In terms of average relative error, the difference between this model and the traditional time series analysis model is small, only 0.01 less than that of the traditional GM(1,1) model, and 0.329 less than that of the traditional GM(1,1) model, and it is obvious that the model of this paper has an excellent prediction effect in the dynamic time series prediction.

Table 5 Comparison of several models

Model	Sum of residual squares	Residual mean square error	Average relative error
GM(1,1)model	2.9151	0.7683	0.616
Time series analyzes model	0.7212	0.3754	0.307
Model of this article	0.475	0.3079	0.297

4 Algorithms for recovering missing data from dynamic time series

In the process of dynamic time series modeling, the problem of missing data in the original time series caused by some uncontrollable factors at the time of data collection often hinders the construction of dynamic time series models and affects the dynamic time series forecasting work. In this paper, we will also design an efficient recovery algorithm based on sparse Bayesian learning to effectively predict the missing data in multiple dynamic time series and solve the modeling problem of the dynamic time series model with gray prediction constructed in this paper in the face of missing data.

4.1 Design of sparse representation bases

Facing the common problem of missing data in dynamic time series modeling, a dynamic time series with missing data is listed. $S=\{s_1,s_2,s_3,s_4,\ldots,s_K\}$ denotes the multi-source time series, $s_j \in R^N$ denotes the data collected from the j rd data source, t_i denotes the i th sampling moment, s_{ij} denotes the sampling value at the j th data source at the i th sampling moment, and "?" denotes missing data. In addition, we define a matrix $W \in R^{N \times K}$ to indicate whether the data in S is missing or not.

$$W_{\bar{v}} = \begin{cases} 0, & \text{If } s_{ij} \text{ is absent} \\ 1, & \text{Or else} \end{cases} \tag{22}$$

The aim of this work is to use all or some of the observations to predict all missing data.

Compressed perception theory suggests that if a signal $s \in R^{N}$ is sparse, i.e., $\| s \|_0 \ll N$, it can be sampled at a rate lower than Nyquist's law according to the observation matrix $\psi \in R^{M \times N}$ and the original signal can be recovered with high probability from the observations $y_{M \times 1} = \psi s$. Many signals in practice are not sparse per se, but can be sparsely represented under some sparse representation basis φ, i.e., $s = \varphi x, \| x \|_0 \ll N$, and can likewise be undersampled according to the observation matrix ψ and the original signal recovered with high probability via observation value $y = \psi \varphi x$ [21]. In fact, most time-series signals have natural time-domain smoothing, e.g., the temperature of a room, the energy consumption of a city, the price of a commodity, etc., all of which change significantly only at a few moments. Therefore, only a small number of values in the difference between two neighboring sampled values of signal s should be large, while most of the others can be ignored. Therefore, the matrix shown in equation (2) is designed:

$$\Omega_1 = \begin{bmatrix} 1 & -1 & 0 & \cdots \\ 0 & 1 & -1 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{23}$$

Let the data collected from j data source can be represented as: $s_j = \{ s_{1j}, s_{2j}, s_{3j}, \cdots, s_{Nj} \}$, s_j The projection vector under matrix Ω_1 is:

$$x_{j1} = \begin{bmatrix} 1 & -1 & 0 & \cdots \\ 0 & 1 & -1 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} s_{1j} \\ s_{2j} \\ \vdots \\ s_{Nj} \end{bmatrix} = \begin{bmatrix} s_{1j} - s_{2j} \\ s_{2j} - s_{3j} \\ \vdots \\ s_{Nj} - s_{1j} \end{bmatrix} \tag{24}$$

Element $s_i - s_{(i+1)j}$ in x_{j1} represents the difference between two neighboring sampled values of time series s_j. Therefore, only a small number of elements in x_{j1} are larger and most of the other elements can be ignored. Other matrices commonly used to represent time-domain smoothness are second-order difference equations, as shown in Eq. (25):

$$\Omega_2 = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots \\ -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{25}$$

The projection vector of s_j under matrix Ω_2 is:

$$\begin{aligned} x_{j2} &= \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots \\ -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} s_{1j} \\ s_{2j} \\ \vdots \\ s_{Nj} \end{bmatrix} \\ &= \begin{bmatrix} 2s_{1j} - s_{2j} \\ 2s_{2j} - s_{1j} - s_{3j} \\ \vdots \\ 2s_{Nj} - s_{(N-1)j} \end{bmatrix} \end{aligned} \tag{26}$$

Element $(s_{ij} - s_{(i-1)j}) - (s_{(i+1)j} - s_{ij})$ in x_{j2} approximates the magnitude of the acceleration of the change in time series s_j. Thus, only a small number of elements in x_{j2} are also larger, and most of the other elements can be ignored. Let $\psi_1 = \Omega_1^{-1}, \psi_2 = \Omega_2^{-1}$, collectively Ω_1 and Ω_2 be Ω, x_{j1} and x_{j2} be x_j, and ψ_1 and ψ_2 be ψ, then the dynamic time series can be expressed as $s_j = \psi x_j (j = 1, 2, 3, \cdots, K)$.

4.2 Design of the observation matrix

As mentioned above, column j of matrix W indicates whether the data in time series s_j are missing or not. In this paper, we will utilize those non-missing data as measurements to recover the original time series. Therefore, the observation matrix should be designed to correspond to the positions of the non-missing data one by one [22]. Based on this, the observation matrix shown in equation (27) is designed:

$$\varphi_j = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{27}$$

If the value at (m,n) in matrix $\psi_j \in R^{m_j \times N}$ is 1, it indicates that the m rd measurement was obtained at the n th sampling moment. Let $y_j \in R^m$, denote the set of unmissing data in s_j, then there is:

$$y_j = \psi_j \varphi x_j = A_j x_j \tag{28}$$

The problem to be solved now is to find x_j by y_j and A_j, while ideally x_j is sparse. Therefore, in this paper, the missing data prediction problem in coevolutionary time series is modeled as a multiple sparse vector recovery problem by designing the corresponding sparse representation bases and observation matrices.

4.3 Sparse Bayesian learning

Previous joint recovery algorithms for multiple sparse vectors require the common support information of multiple sparse vectors to be known in advance, which is more difficult to realize in practical applications, especially in the presence of a large amount of missing data. Based on this, a sparse Bayesian learning-based recovery algorithm is designed, which can learn to obtain partial support information to recover multiple sparse vectors simultaneously [23].

First, assume that $p(y_j \mid x_j)$ satisfies a Gaussian distribution with variance σ^2 : $p(y_j \mid x_j; \sigma^2) = \mathrm{N}(A_j x_j, \sigma^2 I_{m_j})$. where I_{m_j} is a discrimination matrix. Let $X = (x_1, x_2, x_3, \cdots, x_K)$, for row i of matrix X, consider the following two Gaussian prior models:

$$p(x_{ij} \mid M_i = 1) \sim N(0, \gamma_i^b)(\text{Line sparsity}) \tag{29}$$

$$p(x_{ij} \mid M_i = 0) \sim N(0, \gamma_{ij}^s)(\text{Element sparsity}) \tag{30}$$

The binary sequence M_i represents the model labels in row i. γ_i^b and γ_{ij}^s are unknown variance parameters. The row sparse model has the same variance for all elements of the row, while the element sparse model has its corresponding variance for each element. Let $M = \{M_1, M_2, M_3, \cdots, M_N\}$ be the set of all row model labels. Combined with the characteristics of dynamic time series, it is known that at least part of M is 1. Each time series s_j defines a diagonal matrix $\Gamma_j \in R^{N \times N}$ whose i th diagonal element is γ_i^b or γ_{ij}^s determined by the model label of that row. Let $\theta = \{\sigma^2, M, \Gamma_1, \Gamma_2, \cdots, \Gamma_K\}, Y = \{y_1, y_2, \cdots, y_K\}$, the goal of this paper is to find the θ that maximizes the probability $p(Y; \theta)$ Since Y is known, once θ it is determined, we can obtain the value of X it by solving for its maximum a posteriori $\arg\max p(X \mid Y; \theta)$

probability. However, it is very difficult to compute the maximum value of $p(Y;\theta)=\int p(Y \mid X;\theta)p(X;\theta)dX$ directly, since picking the most appropriate model from M requires solving 2^N different non-convex optimization problems. Sen can use variational Bayesian theory to convert $p(Y;\theta)$ the computation into the form of Eq. (31):

$$\ln p(Y;\theta)=KL(q(X)\|p(X \mid Y;\theta))+F(q(X),\theta) \tag{31}$$

where the variational distribution $q(X)$ is an approximation of the posterior probability distribution $p(X \mid Y;\theta)$. Since the KL scatter is non-negative, there is $\ln p(Y;\theta)\geq F(q(X),\theta)$ and the equation holds if and only if $q(X)=p(X \mid Y;\theta)$. Similar to the EM algorithm, one can maximize the value of $\ln p(Y;\theta)$ by iterating $q(X)$ and θ.

1) Step E: Let the KL scatter be zero, then the variational distribution $q(x_j)$ can be updated according to equation (32):

$$q(x_j)=p(x_j \mid y_j;\Gamma_j,\sigma^2,M)=N\sim(\mu_j,\Sigma_j) \tag{32}$$

μ_j and Σ_j were calculated according to equations (33) and (34), respectively:

$$\mu_j=\Gamma_j A_j^{\mathrm{T}}(\Sigma_j^y)^{-1}y_j \tag{33}$$

$$\Sigma_j=\Gamma_j-\Gamma_j A_j^{\mathrm{T}}(\Sigma_j^y)^{-1}A_j\Gamma_j \tag{34}$$

where $\Sigma_j^y=\sigma^2 I+A_j\Gamma_j A_j^{\mathrm{T}}$.

2) Step M: Substituting $q(X)$ into $F(q(X),\theta)$ gives θ as:

$$\begin{aligned}\theta&=\arg\max\int q(X)\ln p(Y,X;\theta)\\&=\{\arg\max_{M,\Gamma_1,\cdots,\Gamma_K}\int q(X)\ln p(X;M,\Gamma_1,\cdots,\Gamma_K)dX\\&=\arg\max_{\sigma^2}\int q(X)\ln p(Y \mid X;\sigma^2)dX\}\end{aligned} \tag{35}$$

Once M is determined, then parameter Γ_j can be updated to:

$$\langle\Gamma_j^{opt}\rangle_{j-1}^K=\arg\max_{\langle\Gamma_j\rangle_{i=1}^K}\int q(X)\ln p(X;M,\Gamma_1,\Gamma_2,\cdots,\Gamma_K)dX \tag{36}$$

Then there is:

$$\gamma_i^b = \begin{cases} \frac{1}{K}\sum_{j=1}^{K}((\Sigma_j)_{i,i} + \mu_{i,j}^2) & M_i = 1 \\ \gamma_{ij}^s = (\Sigma_j)_{i,i} + \mu_{i,j}^2 & M_i = 0 \end{cases} \tag{37}$$

where μ_i, j denotes the i rd element of μ_j and $(\Sigma_j)_i$, i denote the i th diagonal element of Σ_j. However, it is necessary to find M the correct way of updating, otherwise it is necessary to compare the values of 2^N times $\int q(X)\ln p(X;M,\{\Gamma f^{\rho t}\}_{j-1}^{K})dX$ to find the best M. To easily derive an efficient model selection scheme, an upper bound on $\int q(X)\ln p(X;M,\{\Gamma_J^{\varphi t}\}_{j-1}^{K})dX$ can be considered:

$$\int q(X)\ln p(X;M)dX \le \ln\int q(X)p(X;M)dX \tag{38}$$

Argument $\left\{\Gamma_J^{\varphi t}\right\}_{j-1}^{K}$ is omitted here to maintain flushness. This alternative objective function allows us to decide independently whether each row is row sparse or element sparse. Considering X as an “observation”, $p(X;M)$ becomes the evidence for the model. Applying BIC theory, the approximation of the evidence for the model in row i of X for x_i can be computed as:

$$\begin{aligned} \ln p(x_i \mid M_i = 1) &\approx \max \ln p(x_i \mid \gamma_i^b) - \frac{1}{2}\ln K + C_0 \\ &= -\frac{K}{2}\ln\frac{\sum_{j=1}^{K} x_{ij}^2}{K} - \frac{K}{2}(1 + \ln 2\pi) \\ &-\frac{1}{2}\ln K + C_0 \end{aligned} \tag{39}$$

$$\begin{aligned} \ln p(x_i \mid M_i = 0) &\approx \max_{\gamma_i^s,\cdots,\gamma_{iK}^s} \ln p(x_i \mid \gamma_{i1}^s,\cdots,\gamma_{iK}^s) - \frac{K}{2}\ln K + C_0 \\ &= -\frac{1}{2}\sum_{j=1}^{K}\ln x_{ij}^2 - \frac{K}{2}(1+\ln 2\pi) - \frac{K}{2}\ln K + C_0 \end{aligned} \tag{40}$$

where C_0 is a constant:

$$\ln\int q(x_i)\, p(x_i;M_i = 1)\, dx_i = -\frac{K}{2}\ln\frac{2\pi\sum_{j=1}^{K}\mu_{i,j}^2 + (\Sigma_j)_{i,i}}{K} - \frac{K}{2} - \frac{\ln K}{2} + C_0 \tag{41}$$

$$\ln\int q(x_i)p(x_i;M_i=0)\mathrm{d}x_i=-\frac{1}{2}\sum_{j=1}^{K}\ln(\mu_{i,j}^2+(\Sigma_j)_{i,i}-\frac{K}{2}(1+\ln 2K\pi)+C_0 \tag{42}$$

Finally, the noise variance can be updated by maximizing the first term of equation (43):

$$(\sigma^2)^{new}=\arg\max_{\sigma^2}\int q(X)\ln p(Y\mid X;\sigma^2)dX$$
$$=\frac{\sum_{i=1}^{K}\| y_j-A_{\mu_j}\|_F^2+\sigma^2\sum_{j=1}^{K}\sum_{i=1}^{N}(1-(\Gamma_j)_{i,i}^{-1}(\Sigma_j)_{i,i})}{\sum_{j=1}^{K}m_j} \tag{43}$$

5 Dynamic time series missing data recovery simulation experiments

This paper proposes a time series missing data recovery algorithm based on sparse Bayesian learning to solve the problem of missing data that may exist in the dynamic time series model in the process of data collection and to ensure the function of dynamic time series prediction of the model in this paper. In this chapter, simulation experiments of dynamic time series missing data recovery will be carried out to test the performance of this algorithm in the face of dynamic time series data recovery under the situation of missing data.

The dataset used in this simulation experiment is the time series data collected from 54 sensors deployed by Intel Berkeley Research Laboratory over a period of one month, which is named as GAS dataset. According to different data missing rates, some data are randomly deleted from the complete GAS dataset to simulate the missing data. The missing rate is defined as the ratio of the number of missing data to the total amount of data. In this paper, Root Mean Square Error (RMSE) and Average Running Time (ART) are used as performance evaluation criteria.

Root Mean Square Error: i.e. RMSE. The smaller the RMSE value, the better the prediction of the model.

Average Running Time: i.e. ART. In order to analyze the computational complexity of different methods, each method was repeated 100 times to calculate its average running time in seconds.

5.1 Performance Comparison with Different Data Missing Rates

The SI, RNN, KPPCA and TDMF algorithms are selected for comparison in the simulation experiments in this chapter. The simulation results of each algorithm on the GAS dataset are compared using RMSE as the criterion. The performance of each algorithm under different data missing rates is specifically shown in Table 6. From the table, it can be seen that compared with other algorithms, this paper's algorithm has a smaller RMSE at any data missing rate. When the data missing rate reaches the highest 95%, the RSME value of this paper's algorithm is 0.1589, which is always lower than the level of the RSME of less than 0.2, which is consistent with the other cases of data missing rate. The simulation results show that the algorithm in this paper is applicable and very effective in solving the problem of missing data in time series when there is more missing data.

Table 6 Performance comparison under different data missing rates

Algorithm	Data missing rate(%)

	20	50	80	90	95
Algorithms in this article	0.0186	0.0145	0.092	0.1031	0.1589
TDMF	0.0367	0.0921	0.1932	0.2524	0.3784
SI	0.0932	0.1624	1.0202	1.7712	3.5164
RNN	0.0472	0.1021	0.8835	0.9114	1.0629
KPPCA	0.0378	0.0648	0.1876	0.2819	0.4476

5.2 Comparison of computational complexity

The computational complexity of each algorithm is compared at different data true rates using ART as a criterion. The comparison of average runtime (ART) of different algorithms is specifically shown in Table 7. From the table, it can be seen that the SI algorithm requires the shortest operation time, with the lowest ART value of 0.7216 s. This is because the SI algorithm only carries out simple interpolation operations on the data, so the operation time is also the least. In this paper, the algorithm needs to extend the data to multi-dimensional operation, so the operation time is relatively long. However, during the whole algorithm operation, only one modeling is performed and it is a simple dimension expansion operation, and the algorithm only solves a 1-dimensional sparse vector recovery problem, so the operation time is still very considerable.

Table 7 ART of different algorithms

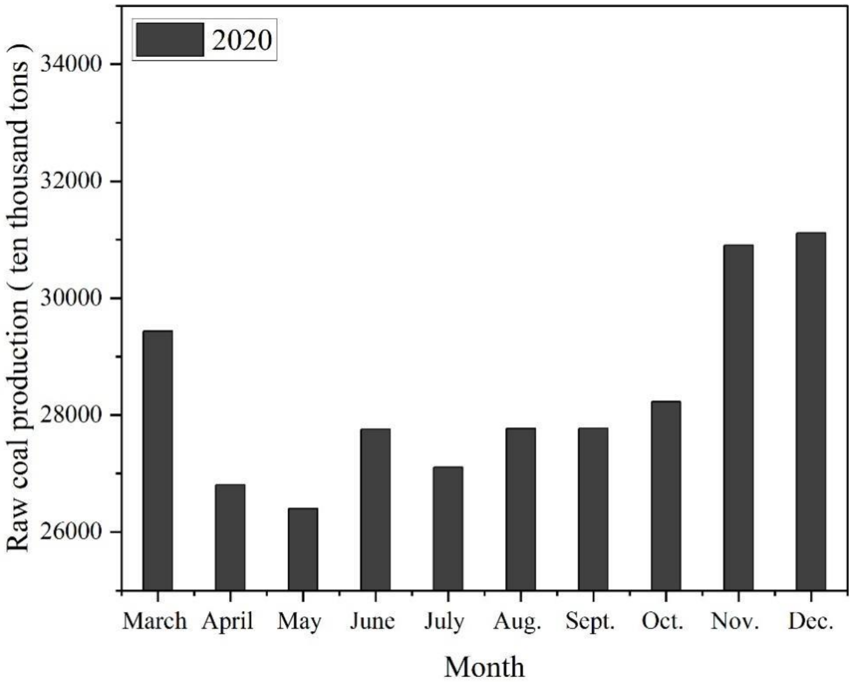
Data missing rate(%)	SI	RNN	KPPCA	TDMF	Algorithms in this article
20	0.7216	2.5216	10.8683	8.4362	2.9032
50	0.8942	3.1832	11.5781	8.9381	3.0802
80	1.2647	5.161	14.944	14.411	3.8317
90	3.4823	7.032	20.445	20.4093	4.923
95	3.6491	10.491	24.8685	22.699	5.1441

6 Example Application of Dynamic Time Series Forecasting of Raw Coal Production

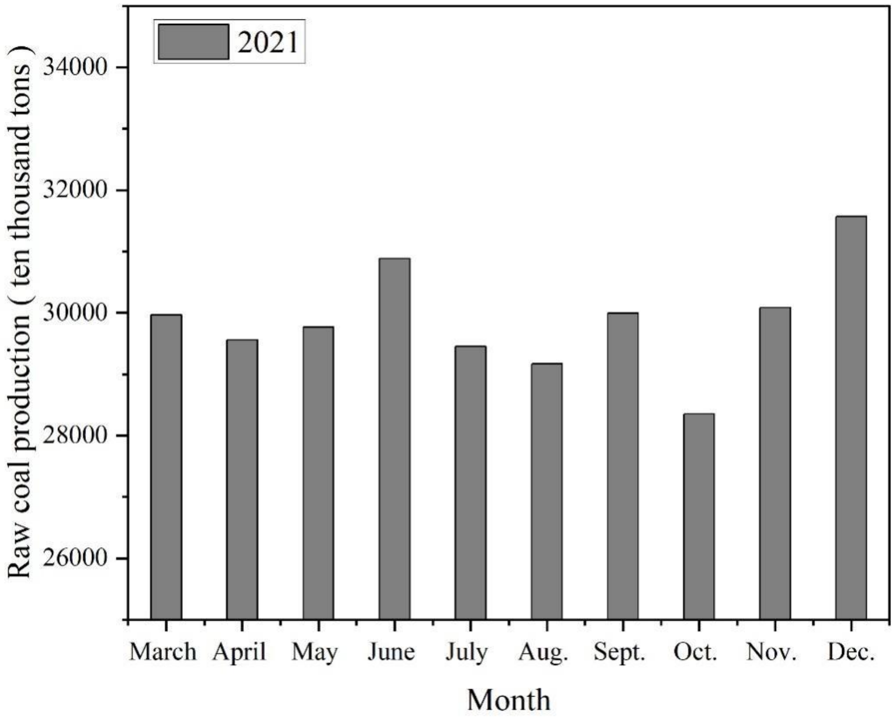
This chapter applies the dynamic time series model based on gray prediction constructed in this paper to the prediction of raw coal production in reality, and at the same time applies the dynamic time series missing data recovery algorithm based on sparse Bayesian learning to solve the problem of missing data in the prediction of raw coal production, and explores the practical utility of the model and the missing data recovery algorithm in this paper.

6.1 Data sources and processing

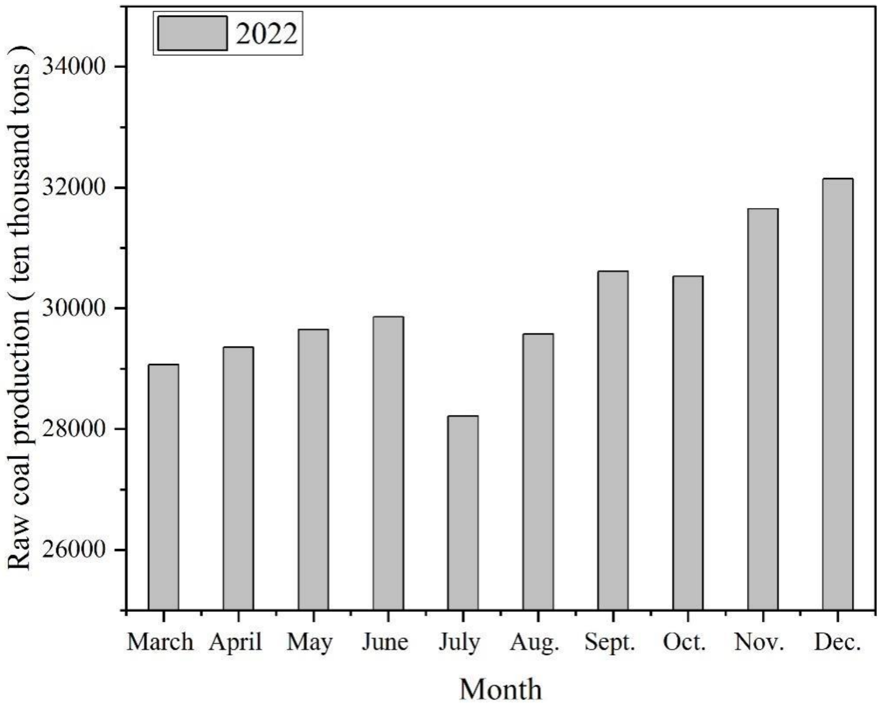
The data used in this chapter comes from the National Bureau of Statistics, and the monthly data of China's raw coal production in 2019-2023 are selected to train the model with the monthly data in 2020-2023 to predict the future changes with its trend changes. The data of January and February of each year are missing.The raw coal production of each month in the period of 2020-2023 is specifically shown in Figure 2. Figures (a) to (d) represent the raw coal production in 2020, 2021, 2022 and 2023 in turn. From the figure, it can be seen that the raw coal production in each month shows an increasing trend. Meanwhile, the raw coal production fluctuates more in 2020 and 2023, and the raw coal production fluctuates less in 2020 and 2021. Next, the dynamic time series model based on Bayesian learning with gray forecasting proposed in this paper will be used to forecast the raw coal production in 2024.



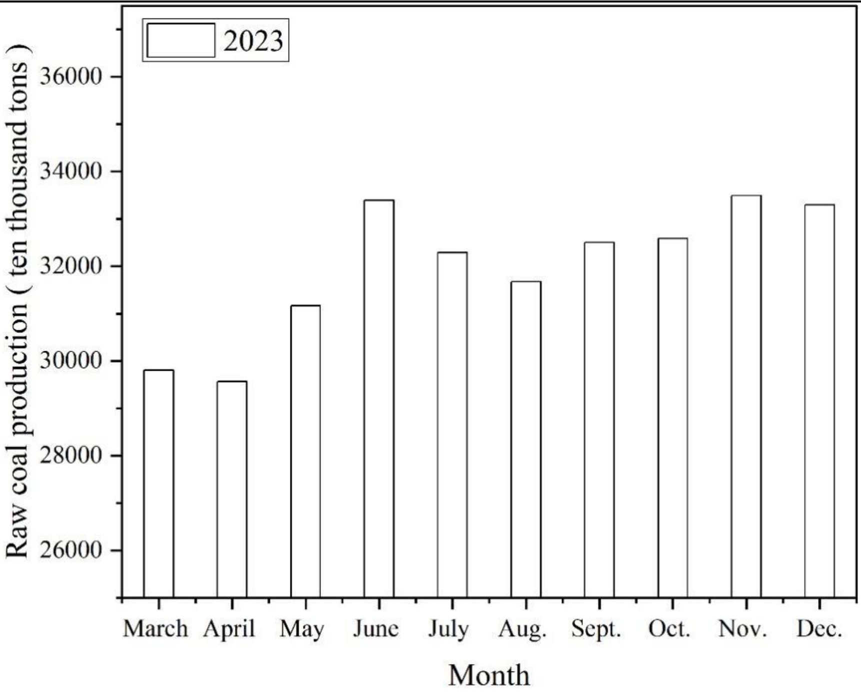
(a)Raw coal production in 2020



(b)Raw coal production in 2021



(c)Raw coal production in 2022



(d)Raw coal production in 2023

Figure 2 Raw coal production in 2020-2023

6.2 Missing value interpolation

In this section, the monthly data of raw coal production from 2020 to 2023 will first be tested for normality, after which 10% of the number of missing data will be randomly generated, and the sparse Bayesian-based time series missing data recovery algorithm proposed in this paper will be utilized for iteration to recover the raw coal production data of January and February. The EM algorithm is selected as a comparison object, and the time spent by the EM algorithm of this paper and the algorithm of this paper in filling the missing data is compared, as shown in Figure 3. It can be seen that the time consumed by this paper's algorithm to recover data basically fluctuates around 4.5s, and the fluctuation is relatively smooth. In contrast, the EM algorithm's data recovery time is also as low as 10.27s, which is also greater than 10s, and the fluctuation is more intense. At the same time, using the ADF test of this paper's algorithm and the EM algorithm's missing value interpolation results, both of which show better smoothness. In general, this paper's algorithm performs better in the recovery of missing data of raw coal production.

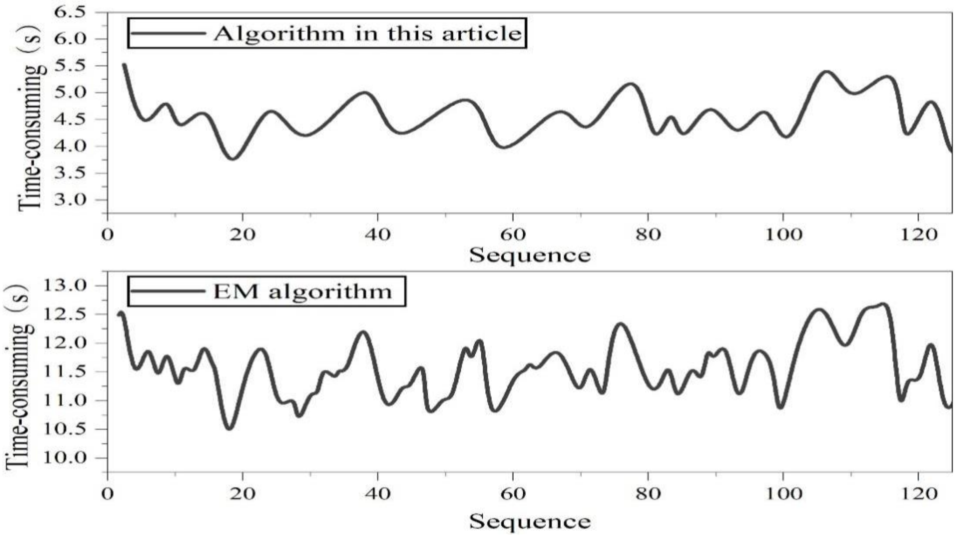


Figure 3 Time consuming of rejecting missing data

6.3 Dynamic time series forecasting analysis

The traditional GM (1,1) model is selected as a comparison model, using the dynamic time series model based on gray prediction proposed in this paper and the traditional GM (1,1) model to predict the raw coal production in 2024 The comparison of the prediction results of the raw coal production in 2024 with the actual value is shown in Table 8.The actual value of the raw coal production in 2024 is subtracted from the predicted production value to the prediction error. It can be seen that the traditional GM (1,1) model yield forecast value is generally higher than the actual value of raw coal production, in 2024, January to July, October yield forecast value is higher than the actual value of raw coal production of more than 1,000 levels. In the remaining months of August, September, November, and December, the production prediction values obtained from the traditional GM(1,1) model are lower than the actual value of raw coal production, with the differences of 1312.07, 1044.4, 1418.84, and 1599.73, respectively, which are also greater than the 1000 level. In contrast, the predicted value of raw coal production of this paper's model in January, May, June, October and November is higher than the actual value of raw coal production of 492.32, 717.02, 314.32, 390.58, 382.11, respectively, and the predicted value of raw coal production of the remaining other months is lower than the actual value of raw coal production, but the difference is still maintained at the level of less than 1000. Obviously, the predicted values of the dynamic time series model proposed in this paper are closer to the actual production values and perform better in the real dynamic time series forecasting work.

Table 8 Comparison of prediction results

Month	Actual raw coal output	Prediction of the model of this article	Error	Prediction of GM(1,1)model	Error
2024-01	33005.13	33497.45	-492.32	34569.16	-1564.03
2024-02	33023.85	32483.61	540.24	34455.15	-1431.3

2024-03	33726.17	32955.75	770.42	35720.47	-1994.3
2024-04	32212.16	31729.61	482.55	33829.36	-1617.2
2024-05	31884.43	32601.45	-717.02	32938.52	-1054.09
2024-06	33427.68	33742	-314.32	35048.31	-1620.63
2024-07	31794.15	31593.69	200.46	33158.47	-1364.32
2024-08	32580.96	32312.18	268.78	31268.89	1312.07
2024-09	33742.47	33084.55	657.92	32698.07	1044.4
2024-10	31593.88	31984.46	-390.58	32795.99	-1202.11
2024-11	32312.74	32694.85	-382.11	30893.9	1418.84
2024-12	32601.54	32154.69	446.85	31001.81	1599.73

7 Conclusion

This paper constructs a dynamic time series model based on Bayesian learning and gray prediction, which provides a solution to the missing data problem often encountered in the dynamic time series modeling process while improving the prediction accuracy of dynamic time series.

In the simulation experiment of dynamic time series prediction, based on the deformation simulation data of pit construction at monitoring point A of a subway station, the residual series of this paper's model after extracting the trend term of the original observation series is smoother than that of the original data, and the P-value is 0.0016<0.01, which meets the conditions of dynamic time series modeling. The simulation prediction of the cumulative sedimentation in the five periods after 92-96, the difference between the predicted cumulative sedimentation in the 92-96 period of this paper's model and the measured value is only 0.1, 0.1, 0.11, 0.06, 0.11, and the difference is much lower than that of the traditional GM (1,1) model and the time-sequence analysis model as a comparison, and the prediction result is the best. In terms of the residual sum of squares, residual medium error, and average relative error, the model in this paper can obtain 0.616, 0.307, and 0.297 respectively by calculation, which are still less than the other comparative models, and show excellent prediction results.

In order to test the data recovery performance of the dynamic time series missing data recovery algorithm proposed in the text model, further data recovery simulation experiments are carried out. Compared with SI, RNN, KPPCA and TDMF algorithms, this paper's algorithm has smaller RMSE at any data missing rate, and the RSME value is only 0.1589 when the data missing rate reaches the highest of 95%.In terms of the computational complexity at different data real rates, the SI algorithm has the lowest ART value of 0.7216s, and the computation time of this paper's algorithm is comparatively longer and only slower than the SI algorithm, which is still very objective.

Finally, the dynamic time series prediction of raw coal production is carried out, and the model of this paper is applied to the real raw coal production prediction work to analyze the realistic utility of the model of this paper. Recovering the missing raw coal production data of January and February in the period of 2020-2023, the recovery time of this paper's dynamic time series missing data recovery algorithm basically fluctuates around 4.5s, and presents a better smoothness, which is better than the comparative EM algorithm. The traditional GM(1,1) model is selected as the comparison model to forecast the raw coal production in 2024. The difference between the predicted value and the actual predicted value of the traditional GM(1,1) model always stays above the level of 1000, while the predicted difference of this paper's model is always lower than 1000, and the prediction effect is better.

In conclusion, the dynamic time series model based on Bayesian learning and gray prediction proposed in this paper has excellent prediction performance and good data recovery ability in the face of missing time series data, which is outstanding in the prediction of dynamic time series data in reality.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

[1] Feyrer, J. (2019). Trade and income—exploiting time series in geography. American Economic Journal: Applied Economics, 11(4), 1-35.

[2] Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. Neural computing and applications, 32, 17351-17360.

[3] Idrees, S. M., Alam, M. A., & Agarwal, P. (2019). A prediction approach for stock market volatility based on time series data. IEEE Access, 7, 17287-17298.

[4] Fulcher, B. D. (2018). Feature-based time-series analysis. In Feature engineering for machine learning and data analytics (pp. 87-116). CRC press.

[5] Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. Knowledge and information systems, 51(2), 339-367.

[6] Ghashghaie, M., Eslami, H., & Ostad-Ali-Askari, K. (2022). Applications of time series analysis to investigate components of Madiyan-rood river water quality. Applied Water Science, 12(8), 202.

[7] Zheng, J., & Huang, M. (2020). Traffic flow forecast through time series analysis based on deep learning. IEEE Access, 8, 82562-82570.

[8] Jermsittiparsert, K., & Chankoson, T. (2019). Behavior of tourism industry under the situation of environmental threats and carbon emission: Time series analysis from Thailand. International Journal of Energy Economics and Policy, 9(6), 366-372.

[9] Pole, A., West, M., & Harrison, J. (2018). Applied Bayesian forecasting and time series analysis. Chapman and Hall/CRC.

[10] Chen, X., & Sun, L. (2021). Bayesian temporal factorization for multidimensional time series prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(9), 4659-4673.

[11] Menchetti, F., & Bojinov, I. (2020). Estimating causal effects in the presence of partial interference using multivariate Bayesian structural time series models. Harvard Business School Technology & Operations Mgt. Unit Working Paper, (21-048).

[12] Sanchez-Castillo, M., Blanco, D., Tienda-Luna, I. M., Carrion, M. C., & Huang, Y. (2018). A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. Bioinformatics, 34(6), 964-970.

[13] Li, T., Liang, H., Xiao, B., Pan, Q., & He, Y. (2023). Finite mixture modeling in time series: A survey of Bayesian filters and fusion approaches. Information Fusion, 98, 101827.

[14] Wan, K., Li, B., Zhou, W., Zhu, H., & Ding, S. (2021). A novel time-power based grey model for nonlinear time series forecasting. Engineering Applications of Artificial Intelligence, 105, 104441.

[15] Zhou, W., Jiang, R., Ding, S., Cheng, Y., Li, Y., & Tao, H. (2021). A novel grey prediction model for seasonal time series. Knowledge-Based Systems, 229, 107363.

[16] Dang, Y., Zhang, Y., & Wang, J. (2023). A novel multivariate grey model for forecasting periodic oscillation time series. Expert Systems with Applications, 211, 118556.

[17] Han, M., Zhang, R., Qiu, T., Xu, M., & Ren, W. (2017). Multivariate chaotic time series prediction based on improved grey relational analysis. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(10), 2144-2154.

[18] Cui, Z., Wu, J., Ding, Z., Duan, Q., Lian, W., Yang, Y., & Cao, T. (2021). A hybrid rolling grey framework for short time series modelling. Neural Computing and Applications, 1-15.

[19] Sajedeh Hedayatollahi Pour,Omid Solaymani Fard & Bo Zeng. (2025). A grey prediction model based on Von Bertalanffy equation and its application in energy prediction. Engineering Applications of Artificial Intelligence110012-110012.

[20] Yuying Man,Nian Li,Zhen Liu & Xiangyong Zeng. (2024). The explicit values of the UBCT, the LBCT and the DBCT of the inverse function. Finite Fields and Their Applications102508-102508.

[21] Shuping Wang,Haitao Wang,Xianming Yang,Dehai Huang,Kai Zheng & Jianbo Chu. (2025). Laser ultrasonic reconstruction model for additive manufacturing based on compressed sensing theory. Measurement116349-116349.

[22] Zhihua Cui,Chunmei Zhang,Yaru Zhao & Zhentao Shi. (2019). Adaptive Bat Algorithm Optimization Strategy for Observation Matrix. Applied Sciences(15).

[23] Lei Yao,Ruoyu Zhang,Changcheng Hu & Wen Wu. (2025). Off-grid DOA estimation for metasurface antenna systems using sparse Bayesian learning. AEUE - International Journal of Electronics and Communications155615-155615.