Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 6, pp. 5205–5217, DOI 10.12694/scpe.v25i6.3280

RESEARCH ON THE MARKOV-CHAIN STATE INTERVAL DIVISION BASED ON PREDICTED DATA CORRECTION

LIXIN PENG[*], XIN ZHANG[†], JUNJIE LI[‡], WU BO[§], FUHAO YANG[¶], AND XU GONG[‖]

Abstract. By 2022, the total length of roads in Tibet Autonomous Region reached 121,447 kilometers. Due to the unique geological conditions in Tibet, various natural disasters such as earthquakes, mudslides, landslides, avalanches, and strong winds frequently occur. Along the Sichuan-Tibet Highway alone, over 300 disasters happen each year, significantly impacting the region's economic development. This study focuses on the complexity and randomness of natural disaster mechanisms and combines Markov chain theory to improve the accuracy of prediction data for mudslides, landslides, and earth subsidence etc. The main method is to modify the state interval of the prediction model parameter-Markov chain based on the distribution of discrete points on the number axis.

The following state interval division methods are proposed:(1) If the relative error of the predicted value exceeds 50%, adjust the prediction model. (2) Obtain the lower bound of state E1 by taking the floor value downward. (3) The width of each interval does not need to be uniform. (4) Arrange continuous, dense, and close points on the number line in the same side in batches, and represent a state continuously, dividing it into one suitable interval or batches. Using this method, an improved RMSE of 0.28mm and MAPE of 0.87% were obtained for engineering examples, outperforming other models such as GM(1,1), Verhulst, DGM(2,1) with corresponding RMSE values of 0.86 mm, 0.69 mm, and 1.38 mm, and MAPE values of 2.75%, 2.53%, and 5.99%. The combined prediction results for five sets of data yielded an RMSE of 0.14 mm and MAPE of 0.56%, which are quite close to the results obtained using Markov selection correction with an RMSE of 0.37 mm and MAPE of 1.01%. Furthermore, comparing the four sets of case, the average reduction in RMSE and MAPE is 3.56mm and 1.72%, respectively, demonstrating that this method can further improve the performance of Markov chain prediction.

Key words: Markov Chain, State Interval, Prediction, Correction

1. Introduction and examples. Forecasting, forecasting and early warning of natural disasters are important tools for disaster prevention and mitigation agencies and university researchers to combat natural disasters in a scientific, economic and rational way. The allocation of protection works in the coming years, the rational arrangement and use of human and material resources, and the promotion of economic development are of great importance to relevant departments. The prediction of natural disasters in the short and medium term involves numerous and complex factors, so the lack of a reasonable and scientific correction model can be fatal to such predictions [1, 2].

There are various prediction models involved in slope displacement prediction, ground settlement prediction, road disease prediction, rainfall prediction and lake area prediction, such as GM (1,1) [3, 4], deep learning [5, 6, 7, 8, 9], and neural network [10, 11, 12, 13], which all have their own advantages. GM(1,1) is more effective for predicting structured sample with less data. Deep learning is more prominent for predicting unstructured data. Neural network has outstanding regression prediction ability for structured data. For prediction models, there are advantages and disadvantages, and different models are selected according to different needs. However, for correction models, it is particularly important to study the commonalities among them [18, 19, 20].

This research employs statistical concepts.Statistical analysis is a crucial step in the five stages of statistical work: statistical design, data collection, sorting and summarization, statistical analysis, and information feedback.The use of statistical analysis methods in research is a high-level investigative requirement. The application of statistical analysis methods in scientific research has the following basic characteristics.

[*]Institute of Technology of Tibet University, Tibet, China
[†]Institute of Technology of Tibet University, Tibet, China
[‡]Institute of Technology of Tibet University, Tibet, China (Corresponding author, lijunjie@hhu.edu.cn)
[§]Institute of Technology of Tibet University, Tibet, China
[¶]Xizang Autonomous Region Sports Industry and Facilities Development Management Center Tibet, China
[‖]Institute of Technology of Tibet University, Tibet, China

1. Scientific. They are based on mathematics and have a strict structure. Specific procedures and specifications must be followed, from confirming topic selection and proposing hypotheses to sampling, specific implementation, analyzing, and interpreting data. Specific procedures and specifications must be followed, from confirming topic selection and proposing hypotheses to sampling, specific implementation, analyzing, and interpreting data. Specific procedures and specifications must be followed, from confirming topic selection and proposing hypotheses to sampling, specific implementation, analyzing, and interpreting data. The conclusions drawn must meet certain requirements of logic and standards.

2. Appreciation. It is important to appreciate that the real world is complex and diverse, and its essence and laws are difficult to grasp directly. Statistical analysis methods are used to collect data from real scenes and quantify them through steps, frequencies, and concise chart representations. Processing this data allows for research and exploration of the world, leading to insights into the inherent laws of the real world.

3. Repeatability. Reproducibility is a measurement index of the quality and level of current research. Research conducted using statistical analysis methods is reproducible. All aspects of the research, from the number of topics to the design of pollutants, as well as the collection and processing of data, can be repeated under the same conditions, allowing for verification of the research results.

The fundamental concept of statistics is to solve practical problems by:

1. Identifying practical issues related to statistics; this article addresses the issue of designing reasonable state interval division standards.
2. Establishing an effective index system; this article uses MAPE and RMSE as evaluation indicators.
3. Collecting data; this article presents 4 representative cases.
4. Selecting or creating effective statistical methods to process and display the characteristics of the collected data; this article lists them.
5. Make reasonable inferences about the overall characteristics based on the collected data, combined with qualitative and quantitative knowledge. Provide suggestions for better decision-making based on these inferences. This article presents a solution based on this approach.This is an idea. There are multiple ways to approach it. This process is typically referred to as the hypothesis testing method when added to a hypothetical solution [21-34].

2. Methods.

2.1. A.Principle of Markov Model.

Markov forecasting approach, proposed by Russian mathematician Markov in 1907, regards time series as a random process, in which the probability of a given event occurring is determined by the previous event, so as to determine the development of future states. If the event has K states $E_1 \sim E_k$, only one state can exist at a time, and each state can have K transition directions. The upper and lower bounds, $E_i \in [a_{1i}, a_{2i}], i = 1, 2, 3, \ldots, k$, are determined by relative error for each state.

$$a_i = (Y(i) - \hat{Y}(i))(Y(i))^{-1} * 100\%, \tag{2.1}$$

$Y(i)$ is the ith monitoring data, and $\hat{Y}(i)$ is the ith predicted data. a_i represents the relative error corresponding to the ith data.

Let $P_{ij} = m_{ij}M_i{}^{-1}$, indicating the probability of the state E_i transitioned to the state E_j by one step, where m_{ij} represents the number of times for the state E_i transitioned to the state E_j by one step, and M_i represents the number of the E_i occurrences. The matrix P composed of all one-step transition probabilities is called the state transition matrix, as follows:

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1k} \\ \vdots & \ddots & \vdots \\ P_{k1} & \cdots & P_{kk} \end{pmatrix}. \tag{2.2}$$

2.2. Markov-Chain Improvement Model Steps.

To begin with, the initial step is to identify the randomness between the predicted and measured values of the model. It is assumed that these values follow a random process. The relative error states are then divided through the Markov chain. Based on the theory, the probability of a given event occurring is determined from the previous event, allowing for the prediction of similar states. Corrections are made accordingly to obtain the latest corrected value. The specific steps involved in this process are as follows:

1. The prediction model is used to get the corresponding predicted value $\hat{Y}(i)$.
2. The state interval E_i is determined based on the magnitude of the predicted relative error.
3. The next state is predicted from the latest state. If there is a non-unique maximum probability value, a two-step transition or an n-step transition is performed until a unique state is predicted according to the maximum probability criterion.
4. According to the state interval, the predicted value is corrected using equation 2.3.

$$Y(t) = \hat{Y}(t)[1 + 0.005(a_{1i} + a_{2i})]. \tag{2.3}$$

In this equation, $Y(t)$ represents the corrected predicted value, while $\hat{Y}(t)$ represents the predicted value. Additionally, a_{1i} and a_{2i} represent the lower and upper bounds of the predicted value, respectively.

5. The data prediction correction process involves replacing the latest predicted data with the set of data farthest from the predicted data, and then repeating steps $1 \sim 4$ until the correction is completed [14].

(1) Some formulas are:

$$B_i(t) = \frac{1}{2}(\overline{\otimes_{1i}} + \overline{\otimes_{2i}}) - \hat{y}(t) = \frac{1}{2}(C_{iup} + C_{idown}). \tag{2.4}$$

The horse chain characteristics prediction curve is based on a prediction value of $\hat{y}(t) = \hat{x}^0(t)$. The upper and lower sides of the curve define the state, with each adjacent pair of curves representing the state. The prediction sequence is divided into intervals and denoted as $\otimes_1, \otimes_{2i}, \otimes_m$. Based on the determination of the transition state of system , the predicted value of the future moment random $B_i(t)$ is most likely taken at the midpoint of the interval $(\overline{\otimes_{1i}}, \overline{\otimes_{2i}})$ [35].

The method's disadvantage is that it evenly divides the state, which is excellent but requires a significant amount of computation. In some cases, this level of accuracy may not be necessary. Proportional revisions are more aligned with the general public.

(2) Some formulas are:

$$\hat{x}^0(k) = \hat{y}(k) + \frac{A_i + B_i}{2}. \tag{2.5}$$

The system is in state K, where the original state is $(\otimes_{1i}, \otimes_{2i})$,$\otimes_{1i} = \hat{y}(k) + A_i$,$\otimes_{2i} = \hat{y}(k) + B_i$ and A_i, B_i changes with k.

The addition of the corresponding error mean directly is a simple and crude method. However, it is important to note that this approach may not be sufficient to meet the logical evaluation criteria.

The $Y(t) = \hat{Y}(t)[1 + 0.005(a_{1i} + a_{2i})]$ It is scientific, rational, and objective in its value.

2.3. State Interval Division Conjecture. The effectiveness of state intervals is determined by their reasonable and scientific division. To achieve this, the following hypothesis is proposed: the final correction effect is dependent on the division of state intervals.

1. According to equation 2.1, the model should be reconsidered if 100 a_i exceeds 50.
2. If the relative errors are similar, they can be expressed as a range.
3. According to equation 2.3, it is recommended to place the upper and lower bounds of the same interval on the same side of the number axis, and subdivide them as much as possible.
4. According to our criteria, a prediction is considered accurate if the relative error is within 1%. In cases where at least 50% of the data has a relative error of 1%, we define an interval of [-1%, 1%].
5. The lower bound of the initial state is determined based on the relative error of the corresponding predicted value. The lower bound is selected using the down-integer operation.
6. The appropriate interval length is determined starting from the lower bound and increasing upwards. This ensures that there is at least one data point in all state intervals, without the requirement of a consistent interval width.

Convert to Mathematical Language:

If $\exists\{[Y(i) - \hat{Y}(i)](Y(i))^{-1}\} \in E_i[a_{1i}, a_{2i}]$ and $\lim_{n \to i}(\hat{Y}(n) - Y(i)) \sim 0$

If and only if $\{[Y(i) - \hat{Y}(i)](a_{1i} + a_{2i})\} \geq 0$

Table 4.1: Case 1 Original data.

Observed Phase	Measured Value/mm	MFF Fitted Value/mm	Error of Fitting/mm	Relative Error	State
1	30.2	23.904	6.296	20.85	4
2	40.3	32.07	8.23	20.42	4
3	70.3	44.639	25.661	36.5	4
4	80.4	61.097	19.303	24.01	4
5	90.2	80.954	9.246	10.25	4
6	120.4	128.921	-8.521	-7.08	1
7	160.5	184.767	-24.267	-15.12	1
8	200.1	245.09	-44.99	-22.48	1
9	300.3	307.061	-6.761	-2.25	2
10	420.5	427.873	-7.373	-1.75	2
11	500.7	484.161	16.539	3.3	3
12	540.8	536.795	4.005	0.74	3
13	640.6	608.393	32.207	5.03	3
14	690.5	671.237	19.263	2.79	3
15	730.3	725.926	4.374	0.6	3
16	760.3	773.323	-13.023	-1.71	2
17	790.7	814.349	-23.649	-2.99	2

Original paper [14] data.

Table 4.2: Comparison of models and their relative errors.

Observed Phase	Measured Value/mm	MFF Fitted Value/mm	Relative Error/%	Markov Improved MFF Predicted Value	Relative Error/%
18	810.8	849.889	-4.82	828.642	-2.2
19	830.4	880.739	-6.06	858.721	-3.41
20	840.2	907.595	-8.02	884.905	-5.32
21	842.3	931.051	-10.54	800.704	4.94

Original paper [14] data.

3. Accuracy Evaluation Method. The evaluation of prediction accuracy cannot be solely based on a single predicted value. This study employs root mean square error (RMSE/mm) and mean absolute percentage error (MAPE/%) as metrics to assess the accuracy.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y(i) - \hat{Y}(i))^2}{n-1}} \tag{3.1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{Y(i) - \hat{Y}(i)}{Y(i)}| \tag{3.2}$$

The variable $Y(i)$ represents the measured value, while $\hat{Y}(t)$ represents the corrected predicted value.

4. Experimental. The literature [14] divides the state intervals into $E_1[-23\%, -5\%]$, $E_2[-5\%, 0\%]$, $E_3[0\%, 10\%]$ and $E_4[10\%, 40\%]$ as shown in Table 4.1 and Table 4.2.

Based on the conjecture presented above, the state intervals have been redivided into $E1[-23\%, -16\%)$, $E2[-16\%, -8\%), E3[-8\%, 0), E4[0, 10\%)$, and $E5[10\%, 37\%]$. The initial state transfer matrix P is shown below:

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} & 0 \\ 0 & 0 & \frac{1}{5} & 0 & \frac{4}{5} \end{pmatrix}. \tag{4.1}$$

The $18th$ phase data is predicted from the $17th$ phase data. The state of the $17th$ phase is corrected to $E3$ and the matrix for the $18th$ phase data is calculated as $[0, \frac{1}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5}]$. According to the maximum likelihood criterion, the corrected value of the 18_{th} phase data is $849.899*(1+0.005(-8+0)) = 815.903$, and the predicted relative error is -0.63%. The data from phases $2nd$ to $18th$ are added to the original data to reset the state transition matrix P:

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}. \tag{4.2}$$

The process of calculating the $18th$ phase data is repeated. According to the maximum probability criterion, the state of the $19th$ phase data is obtained as $E3$, with the corrected value $880.739*(1+0.005(-8+0)) = 845.509$ and the predicted relative error -1.8%. The relative error is predicted using the $3rd$ to $19t$h phase data to calculate the state transition matrix P:

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{7} & \frac{4}{7} & \frac{1}{7} & \frac{1}{7} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}. \tag{4.3}$$

The data for the $20th$ phase is $907.595 * (1 + 0.005 * (-8 + 0)) = 871.291$, with a predicted relative error of -3.7%. Similarly, the state transition matrix P for the data from the $4th$ to $20th$ phases is shown below:

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{8} & \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \tag{4.4}$$

In Table 4.3, the corrected data for the $21st$ phase in state $E3$ is $931.051 * (1 + 0.005(-8 + 0)) = 893.809$, with a predicted relative error of -6.1%.

The state intervals, including $E1[-10\%, -5\%), E2[-5\%, 0), E3[0.5\%)$ and $E4[5\%, 10\%)$, have been divided in the literature [15]. Since the relative error equation a_i in the literature [15] and the relative error equation 2.1 defined in this paper are inverse to each other, the opposite number replacement operation is carried out to replace it with the format in this paper. The revised equation recalculates the $9th$ phase data according to 2.3, and the state is predicted as $E1$ after four transition steps, with the corrected value $14.95*[1+0.005(-10-5)] = 13.83$ and the relative error 5.47%. The $10th$ phase state is predicted as E1 after two transition steps, with a corrected value of $16.22 * [1 + 0.005(-10 - 5)] = 15.0$ and a relative error of -1.01%. The $11th$ phase state is predicted as $E4$ after two transition steps, with a corrected value of $17.63 * [1 + 0.005(10 + 5)] = 18.95$ and a relative error of -26.76%. The original data are shown in Table 4.4 and Table 4.5.

According to the given conjecture, the state intervals have been redivided into $E1[-10\%, -0.1\%]$ and $E2[-0.1\%, 10\%]$. The corresponding state transition matrix P, as per the theory, is as follows:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{4.5}$$

Table 4.3: Comparison of models and their relative errors.

Observed Phase	Value1 /mm	Value2 /mm	Relative Error	Value 3	Relative Error	Value 4	Relative Error
18	810.8	849.889	-4.82	828.642	-2.20	815.903	-0.63
19	830.4	880.739	-6.06	858.721	-3.41	845.509	-1.82
20	840.2	907.595	-8.02	884.905	-5.32	871.291	-3.70
21	842.3	931.051	-10.54	800.704	4.94	893.809	-6.12
RMSE/mm		74.119		40.204		35.936	
MAPE/%		7.360		3.968		3.068	

Corrected Predicted Value.
Value1, 2, 3, 4 refers to Measured Value, MFF Fitted Value, Markov Improved MFF Predicted Value and Corrected Predicted Value of State Redivision, respectively.

Table 4.4: Case 2 Original data.

No.	Measured Value/mm	Predicted Value/mm	Relative Error/%	State
1	6.33	6.33	0	3
2	8.31	8.35	-0.48	2
3	8.56	9.07	-5.96	1
4	10.95	9.86	9.95	4
5	10.72	10.71	0.09	3
6	10.67	11.64	-9.09	1
7	13.02	13.05	-0.23	2
8	13.74	13.75	-0.07	2

Table 4.5: Comparison between the measured values and the predicted values.

No.	Measured Value/mm	Predicted Value/mm	Corrected Value/mm	Relative Error/%
9	14.63	14.95	13.83	5.47
10	14.85	16.22	15.00	-1.01
11	14.95	17.63	18.95	-26.76

According to the $8th$ phase data, the initial vector $\varepsilon_0 = [0, 1]$, the product of the initial vector and the two-step state transition matrix for the $9th$ phase data is $\varepsilon_0 = [1, 0]$. The $9th$ phase state is predicted as $E1$, with the corrected data $14.95 * [1 + 0.005(-10 - 0.1)] = 14.20$ and the relative error 2.94%. Then, we remove the $1st$ phase data, and bring the $9th$ phase state $E2$ into the $2nd - 9th$ phases for predicting the state of the $10th$ phase data. The state transition matrix remains P. After two steps of transition, the state transition matrix is P_1:

$$P_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{4.6}$$

Therefore, in Table 4.6, the $10th$ phase is in state $E1$ with corrected data of 15.4 (-3.7% relative error) and the $11th$ phase data is revised to 16.74 (-11.97% relative error) by repeating the above steps.

Figure4.1 displays the scatter distribution of relative errors calculated based on the original model. The red scatter points represent the same distribution of relative errors as the original model. The blue horizontal line represents the new interval selected according to the state interval division standard proposed in this article, and the green horizontal line represents the state interval divided in the original model. Two adjacent lines of the same color represent a state interval, while the blue-green dotted line represents a common interval. The upper and lower limits of the respective status intervals can be found on the left.

Table 4.6: Comparison between the original prediction model and the state interval redivision.

No.	Measured Value/mm	Corrected Value/mm	Relative Error/%	Corrected Value of Redivision/mm	RE[1]/%
9	14.63	13.83	5.47	14.20	2.94
10	14.85	15.00	-1.01	15.40	-3.70
11	14.95	18.95	-26.76	16.74	-11.97
RMSE/mm		2.886		1.109	
MAPE/%		11.080		6.203	

1 refers to Relative Error of Corrected Value of State Redivision.

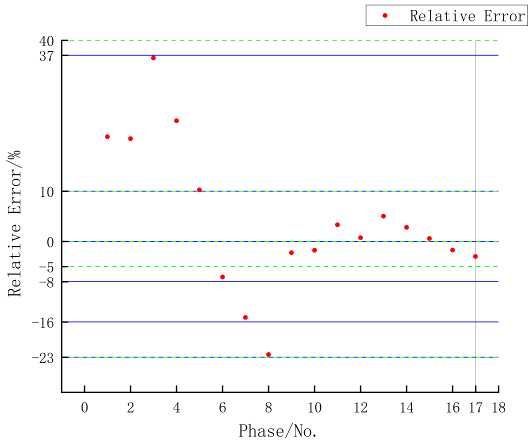

Fig. 4.1: Relative phase distribution scatter diagram and state interval distribution diagram for case 1.

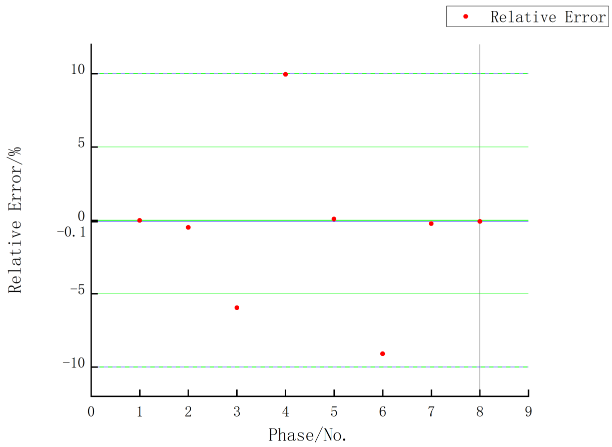

Fig. 4.2: Relative phase distribution scatter diagram and state interval distribution diagram for case 2.

Figure4.2 displays the scatter distribution of relative errors for the second example. The red scatter points represent the same relative error distribution as the model of the second example. The blue horizontal line

Table 4.7: Interval sorting.

state	1	3	3	3

Table 4.8: Case 3 Original data.

Initial Data	GM(1,1)	Verhulst	DGM(2,1)	Combined Prediction
(Original values and grey predicted values of the 25th-46th phases;) Unit: mm				
16.4	15.97	16.304	15.141	-
17.2	16.599	17.015	15.794	-
17.6	17.254	17.768	16.482	-
18.2	17.934	18.564	17.207	18.278
19.0	18.64	19.408	17.972	19.126
19.2	19.375	20.304	18.778	19.152
20.0	20.139	21.255	19.628	20.232
23.0	20.933	22.668	20.525	22.934

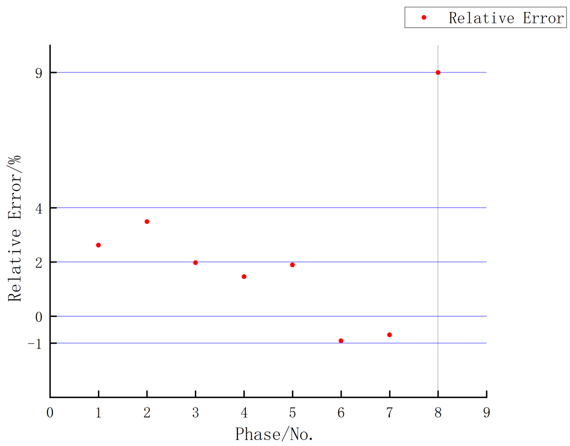

Fig. 4.3: Relative phase distribution scatter diagram and state interval distribution diagram for case 3.

represents the new interval selected according to the state interval division standard proposed in this article, and the green horizontal line represents the state interval divided in the original model. Two adjacent lines of the same color represent a state interval, while the blue-green dotted line represents a common interval. The upper and lower limits of the respective status intervals can be found on the left.Figure4.3 and Figure4.4 convey the same meaning, but with different cases and models.

Based on the above two examples, a preliminary conclusion of conjecture 2 and 3 (as described in Section 2.3) is as follows: If the upper and lower bounds of the same interval are continuously dense and close on the same side of the number axis, they can be included in a suitable interval or batch, and the continuously dense and close relative error values can be represented as a continuous state as much as possible, for example Table 4.7.

According to the data provided in literature [16], Table 4.8 shows Case 3 data.

The above GM(1,1) data is corrected on the basis of the Markov-chain improvement steps. According to Figure 4.3 and Table 4.9, the theoretically reasonable state intervals are $E1[-1\%, 0], E2[0, 2\%), E3[2\%, 4\%)$, and $E4[4\%, 9\%]$.

Table 4.9: case 3 State interval division of prediction model.

Phase	Initial Data	GM(1,1)	Relative Error	State
1	16.4	15.97	2.62	3
2	17.2	16.599	3.49	3
3	17.6	17.254	1.97	2
4	18.2	17.934	1.46	2
5	19.0	18.64	1.89	2
6	19.2	19.375	-0.91	1
7	20.0	20.139	-0.69	1
8	23.0	20.933	8.99	4

Table 4.10: Case 3 State interval division of prediction model.

Phase	Initial Data	GM(1,1)	Corrected Value /mm	Relative Error/%
1	16.4	15.97	16.449	-0.30
2	17.2	16.599	17.097	0.60
3	17.6	17.254	17.427	0.98
4	18.2	17.934	18.113	0.48
5	19.0	18.64	18.826	0.92
6	19.2	19.375	19.278	-0.41
7	20.0	20.139	20.038	-0.19
8	23.0	20.933	22.294	3.07

Table 4.11: Case 3 Comparison of accuracy of various models.

Initial Data	GM(1,1)	Verhulst	DGM (2,1)	Combined Prediction	Markov Correction	Selected Comparison
16.4	15.97	16.304	15.141	-	16.449	-
17.2	16.599	17.015	15.794	-	17.097	-
17.6	17.254	17.768	16.482	-	17.427	-
18.2	17.934	18.564	17.207	18.278	18.113	18.113
19.0	18.64	19.408	17.972	19.126	18.826	18.826
19.2	19.375	20.304	18.778	19.152	19.278	19.278
20.0	20.139	21.255	19.628	20.232	20.038	20.038
23.0	20.933	22.668	20.525	22.934	22.294	22.294
RMSE/mm	0.86	0.69	1.38	0.14	0.28	0.37
MAPE/%	2.75	2.53	5.99	0.56	0.87	1.01

The predicted data is revised according to 2.3, with the results as follows Table 4.10.

It can be seen from Table 4.11 that the accuracy of the Markov correction model is higher than that of other models, indicating that the division of state intervals is very reasonable and successful.

According to the data provided in reference [17] Table 4.12.

According to the division of state intervals conjecture, the original state interval $E1[-14.242\%, -8.323\%), E2[-8.323\%, -2$ is divided into $E1[-15\%, -5\%), E2[-5\%, -3\%), E3[-3\%, 0\%), E4[0\%, 1\%), E5[1\%, 4\%]$ As shown in the following picture Figure 4.4.

According to the hypothesis, the state interval is divided, and the initial state transition matrix from 2011

Table 4.12: Case 4 Original data.

Year	Actual Value	Estimated Value	Residual	Relative Residual
2011	80.2	80.2	0	0
2012	103.9	118.6985	-14.7985	-0.14243022
2013	135.2	138.4021	-3.2021	-0.02368417
2014	159.1	165.8975	-6.7975	-0.04272470
2015	198.4	194.3022	4.0978	0.02065423
2016	229.4	232.1136	-2.7136	-0.01182911
2017	280.2	274.1254	6.0746	0.02167951
2018	336.4	324.5689	11.8311	0.035169738
2019	387.5	384.2121	3.2879	0.008484903
2020	436.6	452.5011	-15.9011	-0.03642029

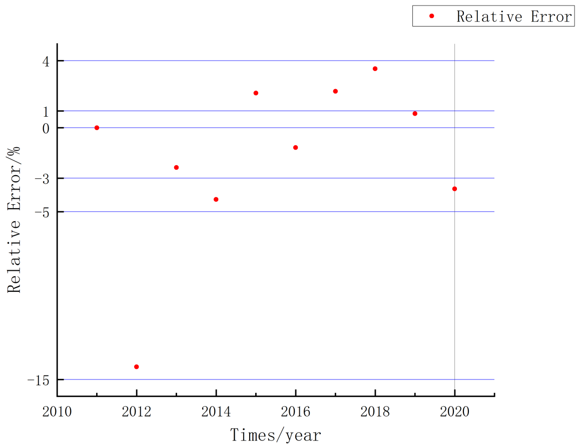

Fig. 4.4: Relative phase distribution scatter diagram and state interval distribution diagram for case 4.

to 2020 is recalculated according to the revised model:

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}. \tag{4.7}$$

In 2020, the state is classified as $E2$, with $\varepsilon_0 = (0, 1, 0, 0, 0)$. After one transition step, multiplied by the initial state transition matrix P, it becomes $\varepsilon_1 = (0, 0, 0, 0.5, 0.5)$. According to the maximum probability criterion, it is impossible to determine the state in 2021. However, after analyzing and multiplying with the ten-step state transition matrix, it can be determined that the state in 2021 is $E2$, with $\varepsilon_9 = (0, 1, 0, 0, 0)$ Therefore, the predicted correction value for 2021 is $Y(t) = 536.7538 * [1 + 0.005(-5 - 3)] = 515.2836$. Using the 'equal innovations' model and excluding 2011, re-modeling from 2012 to 2021 yields a predicted value of 622.2154 for 2022. The product of the two-step state transition matrix and the initial state matrix was recalculated to obtain the 2022 status, which is $E3$. The corrected value is $Y(t) = 622.2154 * [1 + 0.005(-3 - 0)] = 612.8822$.The comparison results are shown in Table 4.13.

Table 4.13: Case 4 Data comparison.

Years	Actual Value	The optimal estimated value of the original model	Re-estimation of the estimated value.
2021	503.5	525.5426	515.2836
2022	576.4	617.5202	612.8822
RMSE/mm	46.6556		38.3380
MAPE/%	5.76		4.33

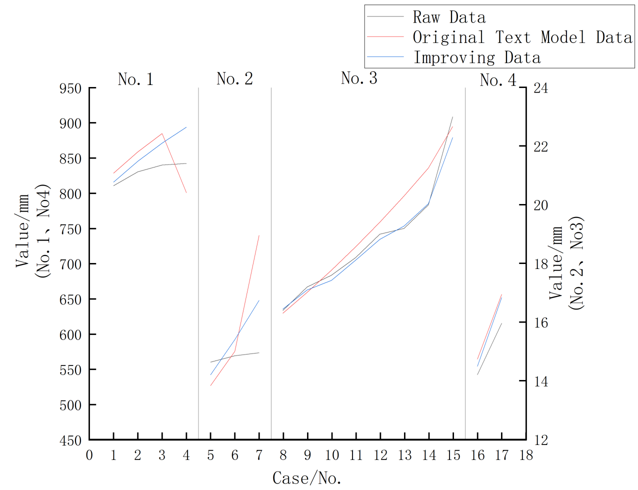

Fig. 5.1: Comparison of Numerical Fitting in Different Cases.

5. Discussion. In comparing the above tests to the original optimization model, the MAPE and RMSE indicators, as well as the level of fitting curves, have all significantly improved. This paper's rules were used to classify the results. Researchers in the fields of landslide displacement prediction and ground subsidence prediction can use this method to divide the state intervals of Markov chains. See Figure 5.1 for a visual representation. Figure 5.1 displays four distinct case models, each separated by a vertical black line and labeled with its corresponding case number. The black curve represents the original data, the red curve represents the curve of the original model data fitting, and the blue curve represents optimization. The final curve should be as close as possible to the black curve for better prediction accuracy. In the figure, the blue curve closely resembles the black curve, indicating a good fitting effect. This suggests that the state interval division method used in this paper is both superior and more scientific.The values of Case 2 and Case 3 correspond to the right coordinate value, while the values of Case 1 and Case 4 correspond to the left coordinate axis value.

6. Conclusions and Results. This paper addresses the scientific division of Markov chain state intervals using the hypothesis testing method. In the geological prediction problem, the MAPE and RMSE indicators are used as references, and this study has significantly optimized the data for these two indicators.

The RMSE for the Markov-MFF model has decreased by 4.268mm, and the MAPE has reduced by 0.9%. RMSE for the Grey-Markov model has decreased by 1.778 mm, and the MAPE has reduced by 4.877%. RMSE for the GNN model has decreased by -0.14 mm, and the MAPE has reduced by -0. 31%. RMSE for the GMM has decreased by 8.3176 mm, and the MAPE has reduced by 1.43%.

No previous scholarly research has been conducted on the reasonable division of state intervals. This article establishes the foundation for such research and provides the possibility for improved data optimization. However, the processing of complex relative error scatter distributions for large sample data is time-consuming, which presents a significant challenge. The future of this study will likely involve a discussion on the appropriate

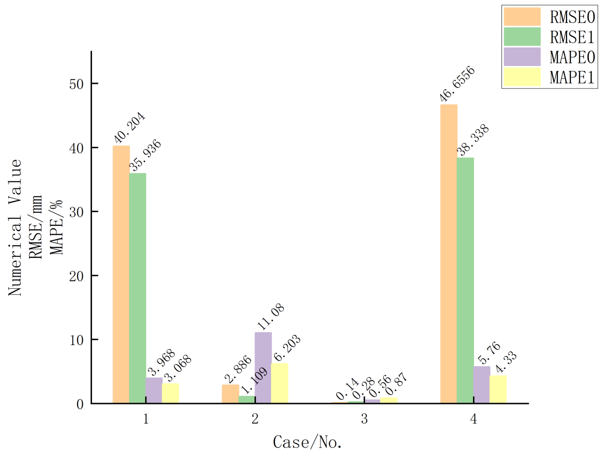

Fig. 6.1: Comparison of RMSE and MAPE optimization index results.

mathematical model for solving the problem of scatter distribution classification. The rules presented in this paper for the division of Markov chain state intervals are adequate. Figure 6.1 displays the results of the case comparison. Figure 6.1 compares the RMSE and MAPE indicators of four different models before and after improvement. Smaller values are better for both indicators. In Case1, RMSE0 represents the original model's RMSE index, while RMSE1 represents the improved index based on the rules proposed in this paper. Similarly, MAPE0 represents the original model's MAPE index, while MAPE1 represents the index after applying the improved rules proposed in this paper. The remaining values are analogous. Figure 6.1 shows that the method proposed in this article has significant optimization effects.

Acknowledgments. Thanks to the Science and Technology Support Project for the Construction of Sichuan-Tibet Railway (XZ202101ZD0001G) of Tibet Autonomous Region Science and Technology Agency project for funding.

REFERENCES

[1] Junjie Li, Guoyang Liu, Tangjin Ye, Wu Bo and Penghui Zhao. *Research on rock slope failure of Tibetan plateau based on three-dimensional discontinuous deformation analysis*, Plateau Science Research, 2018, 2(01): 1-13.

[2] Penghui Zhao. *Research on Machine Vision Identification and Monitoring of Landslide in Southeast Tibet*, Dalian University of Technology, 2022.

[3] Yuehua Huang and Xiaolong Chen. *The port throughput prediction based on optimized GM(1,1) model*, Navigation of China, 2019, 42(04): 136-40.

[4] Guangyuan Li, Xianghong Hua, Haoran Han and Dong Xu. *The prediction model for high-speed railway settlement with a variable coefficient combined ARIMA and GM(1,1)*, Journal of Geomatics, 2019, 44(06): 110-3.

[5] Yao Zhang, Yiquan Wu and Huixian Chen. *Research progress of visual simultaneous localization and mapping based on deep learning*, Chinese Journal of Scientific Instrument: 1-31.

[6] Darui Li, Wanjun Hu, Guangyao Liu, Tiejun G, Laiyang M and Jing Z. *Prediction of preoperative grading of glioma by 3D-ResNet101 deep learning model based on multi-center MRI*, Chinese Journal of Magnetic Resonance Imaging, 2023, 14(05): 25-30.

[7] Ziqi Li, Yuxuan Su, Jun Sun, Yonghong Z, Qingfeng X and Hefeng Y. *Advances in multi-focus image fusion method based on deep learning*, Journal of Frontiers of Computer Science & Technology: 1-22.

[8] Yiquan Wu, Huixian Chen and Yao Zhang. *Research progress of 3D point cloud processing based on deep learning*, Chinese Journal of Lasers: 1-41.

[9] Jun Xie and Xianqiong Chen. *Application of deep learning in seismic location*, China Earthquake Engineering Journal, 2023, (01): 235-43.

[10] Jiahao Zhu, Wei Zheng, Fengyu Y, Xin F and Peng X. *Software quality prediction based on ant colony optimization backpropagation neural network*, Journal of Computer Applications: 1-8.

[11] Ming Yuan, Shuo Wang, Donghuang Yan, Jun Liu and Lian Huang. *Study on damage prediction of concrete bridges based on acoustic emission and convolutional neural network*, China and Foreign Highway, 2022, 42(04): 69-75.
[12] Ting Jiang, Zhenzhong Shen, Liqun Xu, Chong Liu and Jiacheng Tan. *Slope displacement time series prediction model based on SVM-WNN*, Engineering Journal of Wuhan University, 2017, 50(02): 174-81.
[13] Xinhua Xue and Xiaodong Yao. *Fuzzy neural network model for slope stability prediction*, Journal of Engineering Geology, 2007, (01): 77-82.
[14] Yahong Zhao, Weina Wang, Peihua Jiang and Jiahui Li. *MMF settlement prediction model with improved Markov chain and its application*, Bulletin of Surveying and Mapping, 2022, (01): 79-83.
[15] Zhijian Weng, Chenjie Qiu, Fuxiang Qiu, Yunhong Yang, Rufa Lu and Shenglong He. *Grey GM (1,1) settlement prediction model based on Markov optimization and its application*, Science Technology and Engineering, 2020, 20(29): 12065-70.
[16] Yongbo Yang, Minggui Liu, Xianghong Yue and Qi Li. *Slope displacement prediction based on grey theory and neural network*, Journal of Natural Disasters, 2008, (02): 138-43.
[17] Hebing Zhang *Mine Inflow Forecast Based on the Equivalent Novelty Grey Markov*, Mathematics In Practice And Theory,2023,53(08):139-145.
[18] Dabao Yuan, Chengxing Geng, Zhang Ling, Zhenchao Zhang. *Application of gray-markov model to land subsidence monitoring of a mining area*, IEEE Access,2021,9(1): 2169-3536.
[19] Weili Wu. *An intelligent gray prediction model based on fuzzy theory*, International Transactions on Electrical Energy Systems,2022,2022(6):1-9.
[20] Joseph Lopez, Ilona Juan, Adela Wu,Georges Samaha,Brian Cho, J D Luck, Ashwin Soni, Jacqueline Milton, James W May Jr, Anthony P Tufaro and Amir H Dorafshar. *The impact of financial conflicts of interest in plastic surgery: Are they all created equal*, Annals of Plastic Surgery, 2016,77(2): 226-230.
[21] WANG Q H ,JING B Y. *Empirical likelihood for a class of functionals of survival distribution with censored data*, Annals of the Institute of Statistical Mathematics, 2001,53(3):517-527.
[22] LI G, WANG Q H. *Empirical likelihood regression analysis for right censored data*, Statistica Sinica,2003,13(1): 51-68.
[23] WU T T,LI G,TANG C Y. *Empirical likelihood for censored linear regression and variable selection*, Scandinavian Journal of Statistics,2015,42(3):798-812.
[24] ZHOU M,YANG Y F. *A recursive formula for the Kaplan-Meier estimator with mean constraints and its application to empirical likelihood*, Computational Statistics,2015,30(4):1097-1109.
[25] HE S Y,LIANG W. *Empirical likelihood for right censored data with covariables*, Science China Mathematics, 2014,57(6):1275-1286.
[26] HE S Y,LIANG W,SHEN J S,et al. *Empirical likelihood for right censored lifetime data*, Journal of the American Statistical Association,2016,111(514):646-655.
[27] REN J J. *Weighted empirical likelihood ratio confidence intervals for the mean with censored data*, Annals of the Institute of Statistical Mathematics,2001,53(3): 498-516.
[28] SHEN J S,YUEN K C,LIU C L. *Empirical likelihood confidence regions for one-or two-samples with doubly censored data*, Computational Statistics & Data Analysis,2016,93:285-293.
[29] XUE L G. *Empirical likelihood for linear models with missing responses*, Journal of Multivariate Analysis, 2009,100(7):1353-1366.
[30] WANG D,CHEN S X. *Empirical likelihood for estimating equations with missing values*, The Annals of Statistics, 2009,37(1):490-517.
[31] CHEN J H, VARIYATH A M, ABRAHAM B. *Adjusted empirical likelihood and its properties*, Journal of Computational and Graphical Statistics,2008,17(2): 426-443.
[32] TSAO M,WU F. *Empirical likelihood on the full parameter space*, The Annals of Statistics,2013,41 (4):2176-2196.
[33] LIANG W,DAI H S, HE S Y. *Mean empirical likelihood*, Computational Statistics & Data Analysis,2019, 138:155-169.
[34] LIANG W, DAI H S. *Empirical likelihood based on synthetic right censored data*, Statistics & Probability Letters,2021,169:108962.
[35] Wei G. *Research on improved grey GM(1,1) prediction model based on Markov theory*, Computer Engineering and Science,2011,33(02):159-163.
[36] Yi Liu, Lin Zhou. *Research on equipment failure interval prediction based on grey combination model*,Electro-Optics and Control,2010,17(11):61-64.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: Jan 31, 2024
Accepted: Apr 25, 2024