Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 6, pp. 5265–5275, DOI 10.12694/scpe.v25i6.3206

CONSTRUCTION OF CROSS ENERGY TYPE DATA MODEL BASED ON SPATIOTEMPORAL DATA MINING

BO PENG [*], YAODONG LI[†], XIANFU GONG[‡], GANYANG JIAN [§] AND GUO LI[¶]

Abstract. In order to ensure the accuracy of oilfield development dynamic data, the author starts from analyzing the characteristics of development dynamic data, and conducts in-depth research on the characteristics of development dynamic data, the algorithm set for accuracy detection of development dynamic data, and comprehensive analysis methods. Firstly, in response to the spatiotemporal heterogeneity in developing dynamic data, combined with the design concept of a multi detector combination algorithm based on spatiotemporal mixed patterns, the accuracy detection algorithm is evaluated and selected. Based on this, the author proposes a development dynamic data accuracy detection method that considers the influence of multiple factors (FAGTN); Secondly, ARIMA, MGLN, STGCN, and FAGTN algorithms were selected as the algorithm sets for developing dynamic data accuracy detection, in order to complete the data accuracy detection based on monthly oil well data as the research object; Then, a combined weighting based analysis method was proposed to comprehensively analyze the accuracy detection results of dynamic data development, and the results showed: The dynamic data accuracy detection method based on ARIMA has the worst performance, with detection accuracy below 70% in different detection attributes, which is relatively not high enough; The development of dynamic data accuracy detection method based on MGLN achieved an accuracy rate of 80.53% when detecting sleeve pressure, but the accuracy rate did not reach 80% when detecting oil pressure, dynamic liquid level, monthly oil and water production, and the detection effect was relatively unstable; The accuracy of developing dynamic data accuracy detection methods based on STGCN fluctuates around 80%; Realize comprehensive evaluation of detection results; Finally, the experiment and evaluation of the comprehensive detection method for developing dynamic data accuracy were completed using real sample data.

Key words: Spatiotemporal data mining, Cross energy types, Data model construction, Accuracy testing

1. Introduction. In the process of exploration and development, accuracy testing of the dynamic data of oilfield development that has undergone preliminary inspection is an important prerequisite for formulating oilfield development plans, in order to efficiently identify abnormal data [1]. In order to ensure the accuracy of dynamic data in oilfield development, researchers have become enthusiastic about researching methods for detecting the accuracy of dynamic data in development. At present, oilfield workers detect anomalies in dynamic oilfield development data by referring to historical data changes, making judgments based on manual experience, or using machine learning techniques. Due to the reliance on manual experience and lack of dynamism, this detection method has low detection efficiency and accuracy. The specific manifestation is that the professional knowledge and sensitivity to data of oilfield field workers vary, and the basic values for dividing the range of abnormal data based on expert knowledge are not precise enough, this will result in lower detection accuracy, and manual detection can only detect simple data filling errors and data format errors, while the detection ability for data with abnormal values is relatively weak. Therefore, relying on manual experience and expert knowledge to detect data accuracy has great limitations. Existing accuracy detection methods have not solved the applicability problem, and their intelligence is relatively weak when processing data with prominent spatiotemporal heterogeneity. At the same time, due to the influence of factors such as irregular spatial distribution of wells, complex connectivity between wells, well construction problems, changes in injection well indicators, and types of ternary composite flooding in the actual production process of developing dynamic data, existing accuracy detection methods lack applicability and dynamism, and there

[*]Grid Planning & Research Center, Guangdong Power Grid Company, Guangzhou, Guangdong, 510080, China (Corresponding author, pengbopb91@163.com)
[†]Grid Planning & Research Center, Guangdong Power Grid Company, Guangzhou, Guangdong, 510080, China
[‡]Grid Planning & Research Center, Guangdong Power Grid Company, Guangzhou, Guangdong, 510080, China
[§]China Southern Power Grid Research Technology Co., Ltd., Guangzhou, Guangdong, 510080, China
[¶]Electric Power Research Institute of China Southern Power Grid, Guangzhou, Guangdong, 510080, China

may be misjudgments, which will further affect oilfield decision-making. With the update and development of modern technology, data with spatiotemporal characteristics has gradually become a typical data type in the era of big data. Compared to non spatiotemporal data, spatiotemporal data has more complex data dimensions, which leads to an increase in the workload required to process spatiotemporal data. Spatiotemporal data mining is the process of extracting intrinsic, uncertain, and interfering knowledge with important information from a dataset with spatiotemporal characteristics. Its purpose is to explore the spatiotemporal patterns, features, and laws that users are interested in. Currently, domestic and foreign scholars are enthusiastic about researching spatiotemporal data mining techniques and have achieved numerous research results in multiple fields such as data mining and deep learning. In addition, spatiotemporal data mining technology has also been widely applied in fields such as mobile e-commerce, digital urban management maps, air quality prediction, crime detection, traffic management, risk prediction, public health and medical health, human movement trajectory prediction, and oil and gas development.

Data accuracy detection is a branch of data quality detection, and as an important indicator of data quality evaluation, research on it is becoming more and more in-depth with the development of data quality evaluation. Early research on data accuracy detection mainly focused on building a data quality framework and completing data quality detection from multiple dimensions. The data quality framework designed based on this approach can effectively solve the conventional measurement problem of data accuracy, but the definition of data accuracy evaluation indicators is slightly weak. After a period of development, scholars have begun to establish a data accuracy evaluation index system from the perspectives of differential analysis of data accuracy measurement, data lifecycle, and data completion.

Based on the literature on the accuracy detection of dynamic data in oilfield development both domestically and internationally, this study focuses on two main topics: Data spatiotemporal feature mining and accuracy detection research. Currently, many scholars have conducted extensive research on spatiotemporal data mining, data spatiotemporal feature extraction, and accuracy detection, through a comprehensive analysis of the current research status on accuracy detection of dynamic data in oilfield development at home and abroad, the following conclusion can be drawn: Currently, there are few accuracy detection methods for oilfield data with spatiotemporal heterogeneity, which not only have simple rules and low detection accuracy, but also lack intelligence. The existing accuracy detection methods for oilfield data mostly rely on expert experience and obtain abnormal data detection results through knowledge base inference. This method ignores the spatiotemporal heterogeneity of the data, resulting in a lack of rationality and scientificity in the detection results [2]. Therefore, this study investigates the spatiotemporal characteristics of dynamic data in oilfield development, which helps to explore the spatiotemporal correlation between data and improve the efficiency of data accuracy detection.

The spatiotemporal data mining technology, with its excellent spatiotemporal feature extraction method and comprehensive spatiotemporal feature analysis process, can replace manual problem-solving in some aspects. Domestic and foreign scholars have achieved fruitful results in using spatiotemporal data mining techniques to solve anomaly detection problems. At present, some people have applied spatiotemporal data mining to reservoir data processing and proposed a knowledge discovery framework, but there is a lack of analysis in the spatiotemporal characteristics of the data. Therefore, based on spatiotemporal data mining techniques, the author constructs a spatiotemporal data analysis model to achieve accuracy detection of data. Through research, it has been found that designing an accuracy detection method for dynamic data in oilfield development based on spatiotemporal data mining has high effectiveness and application value [3,4].

2. Methods.

2.1. Analysis and selection of accuracy testing methods. In a spatiotemporal heterogeneous environment, there may be significant differences in the changes of various indicators for developing dynamic data. Therefore, using only one accuracy detection method to obtain detection results has significant limitations. It is necessary to fully consider multiple aspects and select multiple algorithms for comparative evaluation in order to obtain more reasonable accuracy detection results. By analyzing and summarizing the characteristics of developing dynamic data, a suitable set of accuracy detection algorithms for developing dynamic data is selected.

(1) Data characteristics. By analyzing the storage structure and spatial distribution of dynamic data, it is concluded that the development of dynamic data has the following characteristics:

The spatiotemporal heterogeneity is prominent. Developing dynamic data changes over time and space. In terms of space, the spatial position of each well is independent, and data indicators such as oil pressure and casing pressure have dynamic differences with changes in the spatial position of the well. In terms of time, the development of dynamic data has significant temporal characteristics. Taking monthly data of oil production wells as an example, there may be significant differences between data from different months.

Dynamic changes in spatiotemporal correlations. The difference in spatial location of wells leads to different spatial correlations between wells, specifically manifested as: The spatial correlation between connected wells is greater than that between adjacent wells, and the spatial correlation between adjacent wells is greater than that of the remaining wells. At the same time, the correlation between adjacent or similar data at time points is greater than that between data with longer distance from time points. That is, the closer two time periods are, the more significant the corresponding data correlation is. Conversely, the less significant the correlation is.

Multiple factors have a significant impact. The output environment for developing dynamic data is complex, and the data is easily influenced by various external factors. For example, indicators related to injection wells, types of ternary composite flooding, well construction issues, and equipment conditions can all have uncertain impacts on the output of development dynamic data [5].

(2) Selection ideas for accuracy detection methods. Selection idea: Developing accuracy detection for dynamic data requires addressing both temporal and spatial processing issues, and using only one intelligent detection method may result in biased results. The core idea for selecting and developing dynamic data accuracy detection methods is to break down the target problem into multiple sub problems and adopt the most appropriate intelligent detection technology to solve different problems by drawing on the idea of "divide and conquer, complement each other's advantages". The basic idea of divide and conquer is to decompose complex problems into relatively independent and easily solvable subproblems, until solutions to all subproblems are obtained, and then merge them into the original solution to the problem. For the accuracy detection problem of developing dynamic data, the process is phased and each stage is relatively independent, so a divide and conquer approach can be adopted to solve it. The complementary advantages are reflected in the selection of detection algorithms, which focus on analyzing the problem-solving ability of different detection methods, evaluating their advantages and disadvantages, using advantages to compensate for disadvantages, and integrating multiple technologies and methods to obtain the best solution to the problem. The selection process is shown in Figure 2.1.

Based on the idea of "divide and conquer, complement each other's advantages", consider from both spatial and temporal dimensions. Graph Convolutional Neural Networks (GCNs) are gradually gaining recognition in dealing with spatial structures based on graph models. In the processing of temporal data, Time Convolutional Networks (TCN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (LSTM), and others are all popular methods. Multi detector combination based on spatiotemporal hybrid mode: The spatiotemporal hybrid mode is a hybrid mode framework that comprehensively considers business needs and satisfies the mining of spatiotemporal heterogeneous data patterns [6]. The spatiotemporal mixed pattern is divided into two parts: temporal pattern and spatial pattern, which represent different stages of pattern mining. It adds labels in the temporal and spatial dimensions, mainly explaining the changes of data objects in time and space. This classification method is not only applicable to the detection of dynamic data in oilfield development, but also to the quality inspection of other data. The difference lies in the differences in the quality inspection field and the difficulty of the business, as well as the different focuses and tendencies. A large number of examples and research results indicate that using "pattern analysis, individual detection, and merge analysis" for data detection in mixed mode is a good solution. Specific implementation methods include multi detector combination mode, tree pruning mode, etc. The detection mode that uses multiple detectors and scientifically combines them according to their respective applicable ranges is called the COMD combination mode (Combination of Multiple Detection). In the design of dynamic data accuracy detection methods, the COMD combination pattern is based on the expectation that "group capability is greater than member capability", and combines multiple detectors to form a comprehensive detection method to obtain the final detection results.

According to the data characteristics of developing dynamic data and the design concept of data accuracy

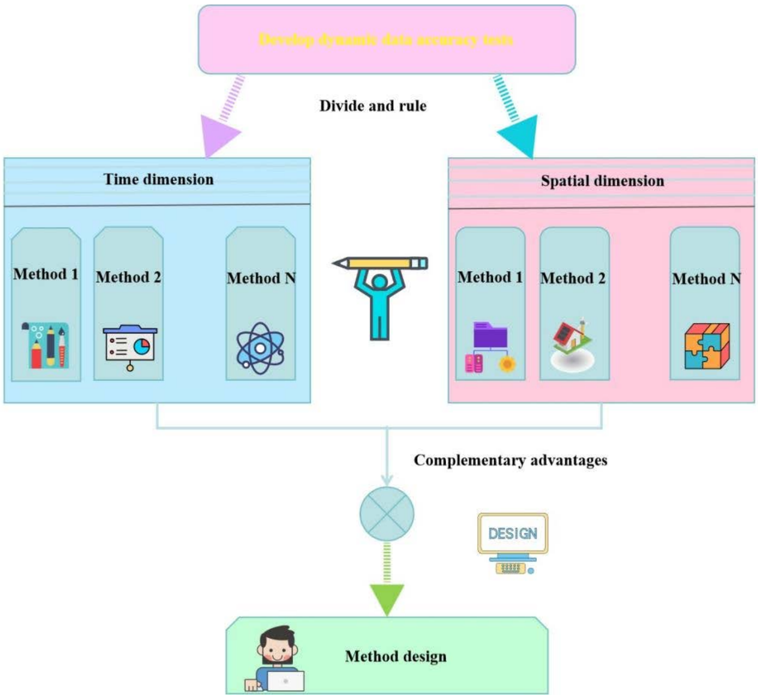

Fig. 2.1: Design ideas for developing dynamic data accuracy detection methods

detection methods, the spatiotemporal pattern mining process of developing dynamic data mainly includes two stages: Spatial feature mining based on well position coordinates and temporal feature mining based on monthly oil well data. Therefore, when testing the accuracy of dynamic data in development, different detection methods are used for different dimensions of data, and the detection results are ultimately combined for analysis.

(3) Development of Dynamic Data Accuracy Detection Algorithm Selection. After weighing various factors in the selection of accuracy detection algorithms for developing dynamic data, the author chose ARIMA, MGLN, and STGCN as the accuracy detection algorithms for developing dynamic data. In order to highlight the spatiotemporal heterogeneity of dynamic data development, the author selected the Autoregressive Mean Moving Model (ARIMA), which is mainly used for anomaly detection in time series data, as a single detection reference to compare with other methods that consider spatial factors. The reasons for selecting all methods are as follows:

The ARIMA model uses existing stable time series data to predict future values, that is, in order to obtain future data from existing stable time series data and complete data anomaly detection. The data form targeted by the model is similar to the indicator changes of development dynamic data, so the ARIMA model effectively utilizes the differences between development dynamic data of different time series lengths for detection [7].

The MGLN algorithm is based on the detection principle of mining the spatiotemporal correlation of data, extracting and analyzing features from both spatial and temporal dimensions. The spatial characteristics of developing dynamic data are reflected in the global or local correlation of data related indicators with changes in well spatial positions, and their temporal characteristics are reflected in the temporal nature of the data. The relationship between values in different time periods is complex. Therefore, the MGLN algorithm effectively utilizes the spatiotemporal heterogeneity of dynamic data development and has certain advantages in processing long time series data.

The implementation approach of the STGCN algorithm is similar to that of the MGLN algorithm, both analyzing from the dimensions of time and space. The difference lies in the different methods used by the algorithm to analyze the temporal features of the data. Therefore, the STGCN algorithm can also effectively

Table 2.1: Schematic diagram of well groups affected by water injection in some wells

Water injection well number	Affected well group
G34-32	G23-32, G23-S325,G24-S315,G24-S32
G34-33	G23-S33,G24-S325,G23-S335,G24-S33
...	...
G34-335	G23-S335,G23-S34,G24-S33,G24-S335

explore the spatiotemporal variation patterns of dynamic data in development, thereby completing accuracy detection tasks.

2.2. Improved Multi factor Development Dynamic Data Accuracy Detection Method.

(1) The overall design of the FAGTN method. This section introduces the overall structure of the Development Dynamic Data Accuracy Detection Method (FAGTN) based on GCN and TCN. Continuing from the data preprocessing methods in Chapter 3, the FAGTN method consists of two main parts: Data modeling and preprocessing, and method design and experimentation. In the data modeling and preprocessing section, unlike Chapter 2, this method requires processing of data related to injection well indicators, well construction issues, and types of ternary composite flooding to generate an external influencing factor matrix; In the network construction and experimental part, a dynamic data accuracy detection network model is constructed based on GCN and TCN. GCN is used for spatial feature mining of dynamic data, while TCN is used to discover the temporal correlation between dynamic data.

(2) Analysis of Factors Influencing the Development of Dynamic Data. The accuracy detection of dynamic data development is not only related to the spatiotemporal heterogeneity of the data itself, but also influenced by various external factors, such as inter well connectivity, injection well related indicators, ternary composite flooding types, well construction problems, etc. In the data preprocessing stage of this study, the inter well connectivity was transformed into a weight matrix through weight calculation, thereby enhancing the saliency of data space feature extraction. Therefore, this section analyzes and explains the impact of external factors from three aspects: injection well related indicators, well construction issues, and ternary composite flooding types.

Analysis of external influencing factors: injection well related indicators. Oilfield water injection plays a crucial role in the entire reservoir development process. Reasonable water injection can not only effectively maintain formation energy, but also improve the efficiency of oilfield development. In the actual process of oilfield water injection research, water injection utilization rate, water injection volume, water injection intensity, water drive index, underground deficit and other water injection related indicators are usually analyzed to evaluate the effectiveness of water injection development. The author uses injection well related indicators as external influencing factors for accuracy detection of development dynamic data, so only monthly injection water volume is selected as the representative influencing parameter of injection wells. Monthly water injection refers to the cumulative amount of water injected into the formation within each month, which can be expressed in cubic meters. It is an important indicator to characterize the water injection status of an oilfield. This study divides the range of water injection influence into well groups by analyzing the connectivity between oil wells and water wells. Table 2.1 shows a schematic diagram of the water injection impact range of some wells in the well group. (Note: The well numbers and other data in the following table have been processed accordingly).

There is a certain correlation between the changes in dynamic data of oilfield development and the monthly water injection volume of adjacent injection wells, and the correlation between the two is uncertain. This study is based on the grey correlation theory. By analyzing the correlation between the monthly water injection volume of injection wells and the monthly data of oil production wells, the impact coefficient of monthly water injection volume on the monthly data of oil production wells is calculated, and a reasonable evaluation of the impact of monthly water injection volume on development dynamic data is achieved. The grey correlation method analyzes whether the time-varying trends (such as direction, speed, and magnitude of changes) between data have similarities, in order to better explore the degree of correlation between each data. For example, for an injection well, there is a high similarity between the changes in the time series of the injection water volume

Table 2.2: Example of G34-32 Well Cluster Dataset

Well No.	Monthly water injection volume of injection well (m3)	Monthly water production of oil wells (m^3)			
time	G34-32	G23-32	G23-S325	G24-S315	G24-S32
T:l	1144	427	209	815	915
T:2	1190	329	393	543	879
T:3	1327	351	300	523	1024
T:4	1086	336	318	306	1097
T:5	1318	345	294	904	1217
T:6	1326	324	226	1124	1072
T:7	1365	964	237	816	1105
T:8	1332	1102	275	602	1094
T:9	1294	924	244	505	1106
T:10	1299	773	204	443	1094
T:11	1279	596	171	330	1032
T:12	1354	360	452	350	1175

Table 2.3: Calculation results of the correlation degree between injection wells and production wells

Water injection well	Production well	correlation
G34-32	G23-32	0.56
	G23-S325	0.701
	G24-S315	0.623
	G24-S32	0.905

and the monthly data time changes of a certain oil production well. The higher the coefficient of influence between the two, the greater the impact of the monthly output data of the oil production well on the injection well's monthly injection water volume, and vice versa. The specific calculation steps are as follows.

Step 1: Data preparation: As shown in Table 2.2, a dataset example of dynamic production data for G34-32 well group in a continuous time series is provided;

Step 2: Use the monthly injection water volume of the injection well as the parent sequence, and the monthly production water volume of the other wells as the subsequence;

Step 3: Use grey correlation analysis to calculate the correlation between this injection well and other production wells;

Step 4: Repeat Step 3 by sequentially taking the other parameters (oil pressure, casing pressure, dynamic liquid level, monthly oil production) of the remaining production wells in this group as subsequences;

Step 5: Calculate the mean correlation between this injection well and other wells. The larger the mean, the higher the correlation between the well and surrounding wells. According to the value of the influence coefficient, the correlation degree is divided into three levels: strong correlation (0.8-1.0), strong correlation (0.6-0.8), and weak correlation (0-0.6).

According to Table 2.2, the correlation degree between the injection wells and production wells in the well group is calculated as shown in Table 2.3.

The greater the correlation between water injection wells and oil production wells, the greater the impact of the monthly water injection volume of water injection wells on the monthly data of oil production wells. According to the calculation results shown in Table 3, G24-S32 is strongly correlated with G34-32 water injection wells, G23-S325, G24-S315 are relatively correlated with G34-32 water injection wells, and G23-32 is weakly correlated with G34-32 water injection wells.

The ternary composite oil recovery technology is an important means to further improve oil recovery in the later stage of high water cut oilfield. It can be divided into strong alkaline ternary composite flooding and weak alkaline ternary composite flooding according to the type of injected alkali. The use of different types of ternary composite flooding will also have different effects on the monthly data of oil production wells. Compared with water flooding and polymer flooding, the cost of strong alkaline ternary composite flooding is higher, and scaling is also more severe; The scaling phenomenon of weak alkaline ternary composite flooding is slightly better than that of strong alkaline ternary composite flooding, mainly manifested in delayed scaling time and fewer scaling wells. However, the configuration process of weak alkaline ternary composite flooding is relatively complicated, and the quality of the configuration cannot be guaranteed. In the on-site application of ternary composite flooding, the staff matched the advantageous wells with the advantageous oil displacement technology, fully amplifying the advantages of the oil displacement technology and greatly improving the mining efficiency.

Construction of external influencing factor characteristic matrix. The author mainly considers three factors: injection well related indicators, well construction problems, and ternary composite flooding types. The monthly water injection volume is selected as the main influencing factor for the injection well related indicators, and a 3-digit independent heat vector is used for encoding, corresponding to three levels of correlation between oil and water wells. The first digit is 1, indicating a strong correlation between oil and water wells, the second digit is 1, indicating a strong correlation between oil and water wells, while the third digit is 1, indicating a weak correlation between oil and water wells. The well construction situation is encoded using a 3-digit independent heat vector, which represents three situations: well construction in the current month, no well construction in the current month, and well construction in the past three months; The type of ternary composite flooding is also encoded using a 3-digit unique heat vector, representing the use of strong alkaline ternary composite flooding, weak alkaline ternary composite flooding, and no ternary composite flooding, respectively.

(3) The loss function of FAGTN. The ultimate goal of training the FAGTN model is to continuously optimize data accuracy detection methods to adapt to the spatiotemporal heterogeneity of development dynamic data, even if the error between the actual values of various attributes of monthly oil well data and the detection values processed by the model is minimized. The loss function during model training is shown in equation 2.1.

$$Loss = ||X - \hat{X}||_2 + \lambda L_2 \tag{2.1}$$

Among them, $\hat{X}$ represents the actual values of various detection attributes in the monthly data of oil production wells, X represents the detection value of the model, and L_2 represents the regularization term of the model, which is used to avoid overfitting of the model, λ for hyperparameters.

3. Experimental Results and Analysis . This experiment compares multiple detection methods and comprehensively evaluates them to complete the accuracy detection of dynamic data development. The brief description of the experimental design is as follows [8]. Elaborate on experimental preparation work, including introducing the experimental environment, describing experimental data, and listing experimental evaluation indicators. Analyze the performance indicators of the FAGTN model proposed by the author and the other three models under various conditions to demonstrate the advantages of FAGtN in terms of detection speed, model accuracy, and stability in certain scenarios. Comparative analysis of the comprehensive analysis method based on combination weighting and the changes in various indicators of the four models, in order to demonstrate the rationality and credibility of using the comprehensive analysis method to detect the accuracy of development dynamic data.

3.1. Experimental preparation.

(1) Experimental environment. Simulate the subsystem of a data quality inspection system for a certain onshore oilfield. In a real environment, the control center is responsible for the unified intelligent scheduling of resources, the transfer platform is responsible for data detection tasks, the data center is responsible for providing data support, and the detection model is responsible for accuracy detection of data.

(2) Data Description. The dataset selected for this experiment is the development performance dataset of a certain oilfield described in Chapter 3. The dataset contains key attributes of monthly data on oil production wells, as well as basic information about the wells. Specifically, there are oil pressure, casing pressure, dynamic liquid level, monthly oil production, monthly water production, well location information, well connectivity information, and external influencing factors. The target detection oilfield consists of no less than 1700 wells. Provide the well distribution and partial well connectivity of the target oilfield.

The oilfield data is summarized once a month, and the experiment uses data from 2008 to 2018. According to specific experimental requirements, the training set, validation set, and test set are divided. This experiment represents multiple attribute parameters of monthly oil well data as different detection tasks.

(3) Evaluation indicators. When comparing the performance of FAGTN with ARIMA, MGLN, and STGCN models, this experiment uses three evaluation indicators: Root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) to evaluate the performance of the four models. The specific calculation formula for indicators is shown below.

$$P_{RMSE} = \sqrt{\frac{1}{\gamma}\sum(\hat{X}_{v}^{t+1} - X_{v_i}^{t+1})^2} \tag{3.1}$$

$$P_{MSE} = \frac{1}{\gamma}\sum_{i=1}^{\gamma}|\hat{X}_{v_i}^{t+1} - X_{v_i}^{t+1}| \tag{3.2}$$

$$P_{MAPE} = \frac{1}{\gamma}\sum_{i=1}^{\gamma}\frac{|\hat{X}_{v_i}^{t+1} - X_{v_i}^{t+1}|}{X_{v_i}^{t+1}} \tag{3.3}$$

where $X_{v_i}^{t+1}$ and $\hat{X}_{v_i}^{t+1}$ respectively represent the next time point (t+1), the true and reference values of the attributes of well V, where V represents the number of wells. Both RMSE and MAE can reflect the error between the true value and the reference value, and the smaller the value of both, the higher the accuracy of the model. MAPE can reflect the ratio between error and true value.

The accuracy detection of developing dynamic data belongs to the binary classification problem, and the detection results only have two basic situations: accurate and inaccurate. Therefore, in the comprehensive evaluation experiment of the algorithm, the commonly used confusion matrix and its extended evaluation indicators for binary classification problems are selected, such as accuracy, recall, and fl_ Score and other criteria are used as evaluation criteria for the rationality of comprehensive evaluation methods. The abnormal data detected by the model is called a positive sample, and the normal data detected is called a negative sample. TP refers to positive samples that are correctly classified by the model, that is, real data is abnormal data, and the accuracy detection result is abnormal; FN refers to positive samples that have been misclassified by the model, where the true data is abnormal but the accuracy detection result is normal; FP refers to negative samples that have been misclassified by the model, where the true data is normal but the accuracy detection result is abnormal; TN refers to negative samples that are correctly classified by the model, meaning that the real data is normal and the accuracy test result is normal. Precision refers to the proportion of true data in a positive sample to a positive sample in the accuracy detection result. The higher the precision, the better the detection effect of the model. The calculation method is shown in equation 3.4. Recall rate refers to the proportion of correctly classified samples in real data samples, calculated as shown in equation 3.5. Fl score takes into account both accuracy and recall, and is an important criterion for measuring the accuracy of model detection. The calculation method is shown in equation 3.6 [9].

$$P = \frac{TP}{TP + FP} \tag{3.4}$$

$$R = \frac{TP}{TP + FN} \tag{3.5}$$

$$f1\_score = \frac{2 * P * R}{P + R} \tag{3.6}$$

Table 3.1: Comparison of experimental effects

Accuracy detection method	Serial Number	Detecting attributes	Accuracy (%)
ARIMA	1	oil pressure	66.46
	2	casing pressure	67.4
	3	dynamic liquid level	62.62
	4	monthly oil production	62.37
	5	Monthly water production	63.56
MGLN	1	oil pressure	79.38
	2	casing pressure	80.54
	3	dynamic liquid level	76.63
	4	monthly oil production	77.8
	5	Monthly water production	78.22
STGCN	1	oil pressure	75.33
	2	casing pressure	77.16
	3	dynamic liquid level	79. 98
	4	monthly oil production	81.15
	5	Monthly water production	81.12
FAGTN	1	oil pressure	80. 53
	2	casing pressure	79.74
	3	dynamic liquid level	81.76
	4	monthly oil production	79.22
	5	Monthly water production	82.16
A Comprehensive Analysis Method Based on Combination Weighting	1	oil pressure	83.28
	2	casing pressure	82. 47
	3	dynamic liquid level	84.56
	4	monthly oil production	81.29
	5	Monthly water production	83.62

3.2. Algorithm Comprehensive Evaluation Experiment . In order to solve the problems of weak credibility, large deviation, and insufficient support caused by using only one algorithm for dynamic data accuracy detection in development, the author proposes a comprehensive analysis method based on combination weighting. This experiment uses ARIMA for comparative analysis MGLN The results of using STGCN, FAGTN, and comprehensive analysis methods to detect the accuracy of development dynamic data verify that the comprehensive analysis method is more reliable and reasonable in solving the problem of accuracy detection of development dynamic data. The obtained experimental results are shown in Table 3.1.

The results in Table 3.1 indicate that the dynamic data accuracy detection method based on ARIMA has the worst performance, with detection accuracy below 70% in different detection attributes, which is relatively not high enough; The development of dynamic data accuracy detection method based on MGLN achieved an accuracy rate of 80.53% when detecting sleeve pressure, but the accuracy rate did not reach 80% when detecting oil pressure, dynamic liquid level, monthly oil and water production, and the detection effect was relatively unstable; The accuracy of developing dynamic data accuracy detection methods based on STGCN fluctuates around 80%; The accuracy of the dynamic data accuracy detection method based on FAGTN is higher than the previous models, with an accuracy rate of 82.15% when detecting monthly water production; The comprehensive analysis method based on combination weighting has an accuracy rate of over 80% in detecting the accuracy of five attributes, which is generally better than using any other algorithm alone. This indicates that the results of the comprehensive analysis method based on combination weighting have good credibility and applicability. In order to present the accuracy detection results more intuitively, as shown in Figure 3.1, a comprehensive analysis method was used to accurately detect the dynamic liquid level of a certain well over a period of time, and some abnormal points were marked.

According to the analysis rules of combination weighting, while determining outliers, the reference value

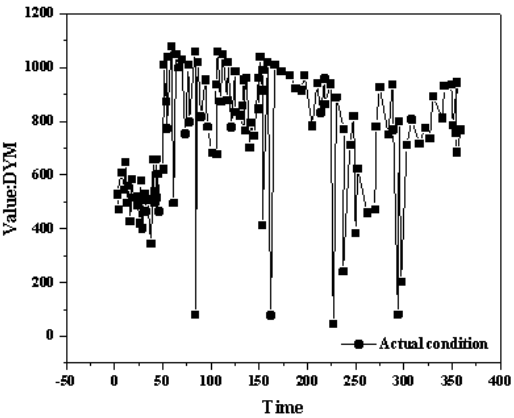

Fig. 3.1: Comprehensive analysis method detection results

Table 3.2: Comparison of reference value ranges

time	model	true value	reference value	confidence interval	Initial judgment	Final judgment	Reference range
44	MGLN	656.8	647.35	[517.88,776.82]	normal	abnormal	[434.72, 652.08]
	STGCN		505.92	[404.74,607.1]	abnormal		
	FAGTN		493.55	[394.84,592.26]	abnormal		
163	MGLN	77.7	85863	[686.9, 1030.35]	abnormal	abnormal	[680.05, 1020.07]
	STGCN		874.71	[699.77,1049.66]	abnormal		
	FAGTN		825.14	[660.11,990.17]	abnormal		

range of the model for determining outliers is also provided. Staff can refer to this range to complete data correction. Taking the data from the 44th and 163rd time points in the experimental area as an example, the reference value range is shown in Table 3.2 [10].

4. Conclusion. The author conducted in-depth research on the development of a comprehensive detection method for dynamic data accuracy, and designed a comprehensive analysis method for the results of dynamic data accuracy detection. Firstly, in response to the spatiotemporal heterogeneity in developing dynamic data, a multi detector combination algorithm selection concept based on spatiotemporal mixed patterns is adopted to complete the evaluation and selection of accuracy detection algorithms; Secondly, considering the various factors affecting the monthly data indicators of oil production wells, an improved accuracy detection method (FAGTN) is proposed by integrating GCN and TCN; Then, design and develop a comprehensive analysis method for the accuracy detection results of dynamic data, and complete the comprehensive evaluation of the accuracy detection results of dynamic data development; Finally, based on real data, experiments were conducted to compare the experimental results, proving the feasibility and effectiveness of this method in actual development of dynamic data accuracy detection.

REFERENCES

[1] Zhong, R., Hu, B., Feng, Y., Zheng, H., Hong, Z., & Lou, S., et al. (2023). Construction of human digital twin model based on multimodal data and its application in locomotion mode identification. Chinese Journal of Mechanical Engineering, 36(1),245.

[2] Su, Q., Bergquist, R., Ke, Y., Dai, J., He, Z., & Gao, F., et al. (2022). A comparison of modelling the spatio-temporal pattern

of disease: a case study of schistosomiasis japonica in anhui province, china. Transactions of the Royal Society of Tropical Medicine and Hygiene, 116(6), 555-563.

[3] Yi, T., Cheng, X., & Peng, P. (2022). Two-stage optimal allocation of charging stations based on spatiotemporal complementarity and demand response: a framework based on mcs and dbpso. Energy, 239(55),99.

[4] Liu, Z., Qian, R., Gao, S., Wang, Y., Jiang, J., & Zhang, Y. (2023). Modelling of damage spatiotemporal distribution in saturated cementitious materials and its chloride transport evolution. Journal of Advanced Concrete Technology, 21(4), 248-261.

[5] Wang, S., Tao, Z., Sun, P., Chen, S., Sun, H., & Li, N. (2022). Spatiotemporal variation of forest land and its driving factors in the agropastoral ecotone of northern china. Journal of Arid Land, 14(1), 1-13.

[6] Chen, M., Li, M., Ye, L., & Yu, H. (2022). Construction of canthaxanthin-producing yeast by combining spatiotemporal regulation and pleiotropic drug resistance engineering. ACS Synthetic Biology55(1), 11.

[7] Wang, L., & Gao, X. (2023). Construction of cross-border e-commerce financial risk analysis system based on support vector machine. Journal of Industrial Integration and Management, 08(01), 25-38.

[8] Kaganga, L. S. (2023). Data on land use land cover changes of urban and peri-urban areas: the case of mwanza city in tanzania. Current Urban Studies, 11(4), 15.

[9] Jia, Y., Sun, C., Wu, H., Luan, G., Zhu, S., & Zhao, F., et al. (2023). Spatiotemporal distribution characteristics of land resources in the middle reaches of the yellow river in china from 1980 to 2018: an asset perspective based on multi-source data. Natural Resources Research, 32(4), 1823-1838.

[10] Lin, Y., Peng, C., Shu, J., Zhai, W., & Cheng, J. (2022). Spatiotemporal characteristics and influencing factors of urban resilience efficiency in the yangtze river economic belt, china. Environmental Science and Pollution Research, 29(26), 39807-39826.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Jan 23, 2024
Accepted: Mar 12, 2024