ORIGINAL ARTICLE   **Open Access**

Digital Twin-based Quality Management Method for the Assembly Process of Aerospace Products with the Grey-Markov Model and Apriori Algorithm

Cunbo Zhuang[1,2], Ziwen Liu[1], Jianhua Liu[1,2*], Hailong Ma[3], Sikuan Zhai[1] and Ying Wu[1]

Abstract

The assembly process of aerospace products such as satellites and rockets has the characteristics of single- or small-batch production, a long development period, high reliability, and frequent disturbances. How to predict and avoid quality abnormalities, quickly locate their causes, and improve product assembly quality and efficiency are urgent engineering issues. As the core technology to realize the integration of virtual and physical space, digital twin (DT) technology can make full use of the low cost, high efficiency, and predictable advantages of digital space to provide a feasible solution to such problems. Hence, a quality management method for the assembly process of aerospace products based on DT is proposed. Given that traditional quality control methods for the assembly process of aerospace products are mostly post-inspection, the Grey-Markov model and T-K control chart are used with a small sample of assembly quality data to predict the value of quality data and the status of an assembly system. The Apriori algorithm is applied to mine the strong association rules related to quality data anomalies and uncontrolled assembly systems so as to solve the issue that the causes of abnormal quality are complicated and difficult to trace. The implementation of the proposed approach is described, taking the collected centroid data of an aerospace product's cabin, one of the key quality data in the assembly process of aerospace products, as an example. A DT-based quality management system for the assembly process of aerospace products is developed, which can effectively improve the efficiency of quality management for the assembly process of aerospace products and reduce quality abnormalities.

Keywords: Quality management, Digital twin, Assembly process, Aerospace product, Grey Markov model, Apriori algorithm

1 Introduction

Aerospace products include satellites, missiles, and rockets, which are complex in their customer requirements, product composition, manufacturing processes, and project management. Assembly is the addition or connection of parts to form a complete product, which is important in the research and development (R&D) and production of aerospace products. As the assembly process is the most important link in the delivery of aerospace products, to improve assembly quality and achieve quality management of the process has important engineering significance [1].

Aerospace product assembly is a typical discrete assembly, with the characteristics of single- or small-batch production and a long assembly cycle, involving many professional fields, scattered assembly data, and much rework and repair [2]. The assembly process generates much quality data, including indirect data such as the number of defective products, first pass rate, and repair

*Correspondence: jeffliu@bit.edu.cn
[1] Laboratory of Digital Manufacturing, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China
Full list of author information is available at the end of the article

rate, and direct data such as product length, weight, and assembly errors. Assembly quality data include data that can reflect product quality, such as quality cost loss, production batch, inventory backlog, and invalid operation time [3]. These data are the basis for the evaluation of assembly quality, which can be used to measure product quality and provide guidance for the continuous improvement of assembly quality.

Quality management is a process of organizing and coordinating related activities to meet certain quality requirements, i.e., to manage and control the above quality data. The complete quality management process aims to grasp the status of quality data, predict its future status, and adjust the assembly process accordingly, so as to maintain quality within a reasonable range. Assembly quality data are collected, stored and managed through a manufacturing execution system (MES) [4]. However, assembly quality management adopts post-inspection, i.e., checking and controlling the quality of the final product. If the data meet the standard, the product is warehoused or it enters the next process, and otherwise it is returned for repair or scrap. Assembly process quality management for aerospace products has issues to be resolved.

(1) *Quality prediction when the sample volume of data is small.* A large amount of quality data is generated in the assembly of aerospace products. Their variety is wide, and the volume of a particular type of data may be insufficient for forecasting. Accuracy of data forecasting and sample data volume show some positive correlation. Hence, it can be difficult to obtain accurate predictions.

(2) *Quick location of abnormal causes when influence factors of quality data are complicated.* The assembly quality of aerospace products is affected by the factors of man, machine, material, method, measurement, and environment (5M1E), each having several influencing factors, making it difficult to directly determine specific causes of abnormal quality data.

(3) *Quality control is time-consuming and has poor real-time performance.* Assembly quality control is implemented based on the quality management process of post-optimization, including problem definition, investigation and measurement of relevant factors, analysis and determination of key factors, and control and improvement of influence factors. This can take a long time, precluding the avoidance of imminent quality problems.

The concept of digital twin (DT) was first proposed by the Air Force Research Laboratory in 2011, and was first applied to industry [5]. Its purpose is to optimize the simulation through virtual environment before production to avoid undesirable conditions. With the deepening of research, DT technology has been considered the key to realizing the interaction and fusion between the physical and information worlds [6, 7]. It provides a clear path for the implementation of a cyber-physical system (CPS) and introduces the idea of making full use of digital space to mirror, predict, guide and control the physical space [8]. Additionally, it provides a novel way to assist humans in understanding the physical world from a multi-time dimension including the past, present and future. Hence, the application of DT technology to the quality management in the assembly process of aerospace products can realize quality status monitoring, prediction, and anomaly traceability in virtual spaces based on the collected real-time data and can enable adjustments in physical spaces. It conforms to the basic ideas of status monitoring and prediction, anomaly traceability, and the timely regulation of quality data. Therefore, DT provides a feasible approach to realize quality management of assembly process for aerospace products.

The rest of this paper is organized as follows. Section 2 summarizes the state-of-the-art of quality data prediction, traceability technology, and DT applications in the manufacturing phase. In Section 3, the implementation framework of quality management for the assembly process of aerospace products based on DT is established. Three key technologies are discussed in Section 4, including numerical prediction of quality data based on Grey Markov model, assembly system statues prediction based on T-K statistical control chart, and association rule mining for quality abnormalities based on the Apriori algorithm. In Section 5, taking the centroid data of an aerospace product's cabin, one of the key quality data in the assembly process of aerospace products, as an example, the application process of the proposed method is elaborated in detail. Additionally, a DT-based assembly process quality management system for aerospace products is developed and verified. Finally, the main contributions of the paper and future work are summarized in Section 6.

2 Related Work

This paper focuses on the use of DT technology to achieve quality management for the assembly process of aerospace products. We start by reviewing the quality management technology including quality data prediction and traceability, and pointing out the future trend of taking full advantage of digital space, including low cost, high efficiency, multiple iterations, and predictability, to promote it. And then the state-of-the-art of DT applications in the production phase is overviewed.

Zhuang *et al. Chinese Journal of Mechanical Engineering*   (2022) 35:105

Page 3 of 21

2.1 Quality Data Prediction and Traceability Technology

The key to quality management is to forecast the future development trend of quality data, and even its specific value. Data prediction includes qualitative prediction, i.e., of change trends, and quantitative prediction, i.e., of specific values. We study quantitative prediction for assembly quality management of aerospace products. Quantitative data prediction includes time series analysis and causal analysis. Time series analysis methods include the simple sequential time mean, weighted sequential time mean, moving average, weighted moving average, exponential smoothing, seasonal trend prediction, and market life cycle prediction. Taylor et al. analyzed several time series algorithms based on intraday power demand data from 10 European countries and found the prediction results of exponential smoothing to be best [9]. Guo et al. presented a chaotic time series prediction algorithm to predict wind speed, modified the model with the parallel rule algorithm, and verified the method with real wind data [10]. Time series analysis focuses on predicting large amounts of time series data, but the volume of assembly quality data for aerospace products is generally small and does not necessarily have strict time series. Causal analysis includes linear regression, support vector machines, neural network prediction, Grey prediction, and Markov prediction. To improve the data prediction accuracy, Zhou et al. investigated the fine tuning approach for the parameters of least-squares support vector machines to predict one-step ahead short-term wind speed [11]. For the case of TFT–LCDs which is a small sample size prediction problem, Li et al. developed a new approach involves three-steps. They are K-means clustering, attribute extension using the fuzzy membership function in each cluster, and put the data with new generate attributes into a backpropagation neural network (BPNN) machine learning algorithm [12]. Additionally, grey prediction well also discovers the intrinsic correlation behind the hazy phenomenon and is suitable to predict small datasets. Li et al. used trend price tracking to extract hidden information on the behavior of manufacturing sample data and constructed an adaptive Grey prediction model, AGM(1,1), to forecast industrial data [13]. Chang et al. presented a modified grey forecasting model to forecast the short-term manufacturing demand [14]. Markov models can also be used to predict the trends of small datasets and analyze the future trends of discrete random processes, i.e., to predict the future state of a variable from its present state and change trend [15]. However, how to improve prediction accuracy based on small volumes of sample data is an issue.

The most commonly used approach for quality data tracing, the relational data model, builds the corresponding management model and database in the assembly process to trace abnormal quality data. Materials are marked to trace quality data. Zhuang et al. realized material traceability and management in the assembly process based on workflow technology [16]. Besides, pictures and videos data are collected to initiatively discover the existing quality problems and trace the causes of quality abnormities. On this basis, vision-based recognition methods from a feature perspective are applied to discover the defect [17]. Such methods achieve traceability of quality data through correlation but do not mine the complicated association between multi-source heterogeneous scattered data, so the efficiency of quality problem traceability is low.

However, due to the unstable process, long assembly and adjustment period, and strict quality control in the R&D stage, quality management still has problems such as much rework and repair, and difficulty tracing quality issues. With the rapid development of new-generation information technology such as the Internet of Things, big data, and artificial intelligence, the deep integration of digital and physical space has become a common bottleneck in various countries' industrial strategies, including Industrial 4.0, the Industrial Internet, and "Made in China 2025." Therefore, to fully utilize the advantages of digital space to improve the quality management of the assembly process for aerospace products will be the development trend [18–20].

2.2 Overview of DT Applications in the Production Phase

DT technology has been widely applied in all stages of the product lifecycle [21, 22], including design [23], manufacture [24], and service [25]. In the production stage, Tao et al. introduced the digital twin shop-floor (DTS), a paradigm for shop-floor instances of cyber-physical production systems (CPPSs) [26]. Leng et al. incorporated DT in the parallel control of automated manufacturing systems [27]. Park et al. built a DT-based CPPS architectural framework for personalized production [28]. Bao et al. investigated an ontology-based modeling and evolution method of DT for the assembly shop-floor to deal with for the issues such as discreteness of assembly process, diversity of assembly resource, and complexity of dataflow in the assembly task execution [29]. Son et al. presented a DT-based CPS for abnormal scenarios involving automotive body production lines, which can forecast whether a product can be manufactured where abnormal scenarios occur [30]. Zhang et al. presented an information modeling method for a CPPS based on DT and AutomationML, which integrated various physical resources into CPPS to support information interaction between resources [31]. Yildiz et al. discussed the demonstration and evaluation of a DT-based virtual factory [32].

Zhuang *et al. Chinese Journal of Mechanical Engineering*   (2022) 35:105

Page 4 of 21

Additionally, Sun et al. investigated DT-based assembly commission approach for high precision products to solve the issues of low assembly efficiency and poor-quality consistency caused by traditional manual method [33]. Zhang et al. discussed the hybrid prediction approach of physical model and data to achieve quality assurance for composite components [34]. DT technology is the key in the realization of the virtual-physical fusion of CPS, which makes full use of the digital and physical spaces, and can improve the assembly quality control of aerospace products. In the assembly process of aerospace products, not only the DT model can truly and dynamically mirror the quality status and process of the physical counterpart, but also can be applied to achieve quality anomaly traceability and quality status prediction. Hence, the DT can be used to assist shop-floor managers in physical space to carry out quality optimization to effectively avoid some quality anomalies and reduce quality problem processing time. The above research uses DT in the production phase but does not address its application to assembly process quality management and control.

3 Framework of DT-based Quality Management for the Assembly Process of Aerospace Products

There are two main quality abnormalities in the assembly process of aerospace products: abnormal product quality data and uncontrolled assembly systems. The first case indicates quality problems in the assembly process. The latter is a hidden quality problem. If the assembly system is uncontrolled, quality anomalies are more likely in the subsequent assembly process. Considering these two anomalies, we propose a framework of quality management for the assembly process of aerospace products based on DT, as shown in Figure 1.

On the physical assembly shop-floor, a shop-floor Internet of Things (IoT), including RFIDs, sensors, barcodes, industrial Ethernet and wireless network, is constructed to realize real-time perception of manufacturing resources, quality data collection and transmission.

On the shop-floor data layer, real-time quality data are dynamically collected based on the built shop-floor IoT and managed based on product assembly BOM, which include inspection data, measurement data, assembly process parameters, environmental data, equipment operation data, material usage data, process completion data, and technical problem data. To ensure the real-time quality data, the inspection data and measurement data, such as centroid, moment of inertia, and weight, are automatically collected and transmitted using the corresponding inspection and inspection equipment. The assembly process parameters, environmental data and equipment operation data are obtained using the corresponding sensors such as displacement, speed, temperature and humidity sensors. The material usage data are acquired through RFID (Radio Frequency Identification) and scanning the barcode corresponding to the material. The process completion data and technical problem data are collected with human computer interaction.
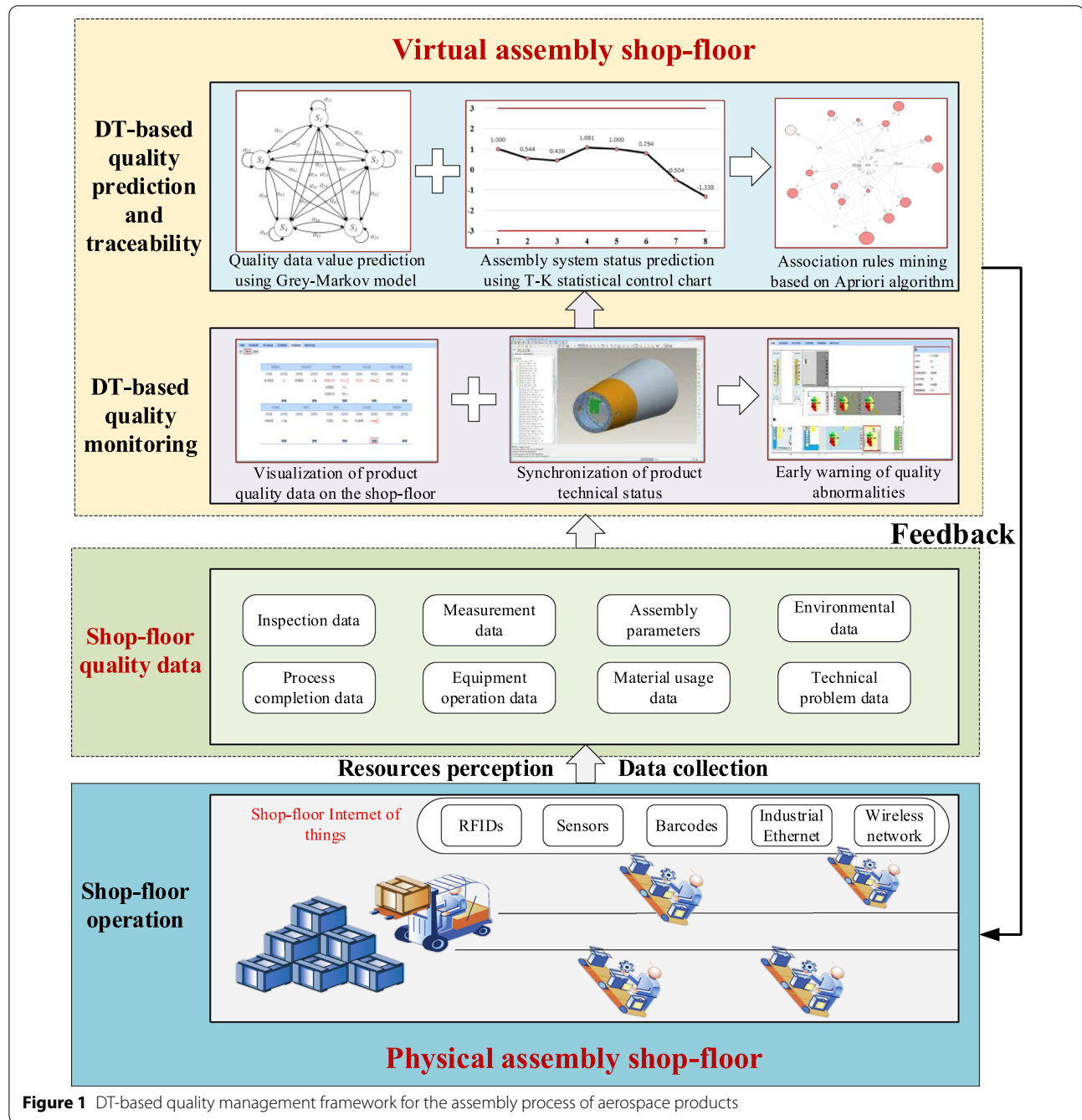
On the virtual assembly shop-floor, a DT model of quality management for the assembly process of aerospace products is constructed. The DT model is composed of two parts: one is the shop-floor visualization model, which can be used for quality monitoring, and the other is the quality prediction and traceability model, which can be used for calculation and decision-making. Therefore, the virtual shop-floor contains two levels.

At the DT-based monitoring level, based on the shop-floor visualization model, data and model visualization are applied to mirror and monitor the product quality status based on the built DT model of the shop-floor, and warnings are issued if abnormalities occur.

At the DT-based quality prediction and traceability level, a Grey-Markov model forecasts the future value of quality data according to historical and current data. An abnormal predicted value indicates quality abnormality in the current assembly process, triggering cause traceability; if the predicted value is normal, verification will continue. Then, according to the Grey-Markov forecasting results and historical quality data values, observation samples are selected from different batches, a T-K statistical control chart is established according to the sample data, changes of the mean and standard deviation of quality sample data are observed, and it is predicted whether the assembly system will be within the control range at the next moment. If an uncontrolled assembly system is forecast, quality anomalies are highly likely during subsequent assembly, triggering the cause traceability process of quality abnormality; if the assembly system is forecast to be in control, the process is repeated until no products are available. When the cause tracing process of quality abnormality is triggered, it is necessary to analyze factors that may affect the quality data on the assembly shop-floor according to 5M1E. The values of these factors, together with quality data values and assembly system status, form a project set. We then use the Apriori algorithm to mine the strong association rules related to quality data abnormalities and an uncontrolled assembly system. Through these strong association rules, the influencing factors related to quality anomalies are traced and stored in the quality management knowledge base.

The key technologies illustrated in Section 4 focus on relevant technologies included in the level of DT-based quality prediction and traceability, which contains numerical prediction of quality data using a Grey-Markov model, status prediction of the assembly system based on a T-K

Figure 1 DT-based quality management framework for the assembly process of aerospace products

control chart, and mining of association rules for quality abnormities based on the Apriori algorithm.

4 Key Technologies

4.1 Numerical Prediction of Quality Data Using Grey-Markov Model

4.1.1 Grey Model

When only part of the information in a system is known, the unknown information can be predicted using the Grey model (GM), i.e., a Grey system theory model. While appearing to be random, the quality data generated during the assembly process of aerospace products is actually ordered and time-dependent, so the Grey model can be used.

For the quality data of the assembly process of aerospace products, a GM(1,1) model is established, i.e., a first-order linear differential equation.

The GM(1,1) modeling process for assembly quality data of aerospace products is as follows.

Assume the original series of quality data is:

$$y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \ldots, y^{(0)}(n)), y^{(0)}(t) \geq 0, \quad (1)$$
$$t = 1, 2, \ldots, n.$$

Calculate the scale ratio of the original series of quality data

$$\lambda(t) = \frac{y^{(0)}(t-1)}{y^{(0)}(t)}, t = 2, 3, \ldots, n. \quad (2)$$

If all scale ratios of quality data fall within the acceptable coverage interval $Y = (e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}})$, the original series of quality data $y^{(0)}$ can be used to establish the GM(1,1) model. Otherwise, the data must be transformed, such as through a shift transformation

$$x^{(0)}(t) = y^{(0)}(t) + c, t = 1, 2, \ldots, n, \quad (3)$$

where c is a constant that can bring all scale ratios fall within acceptable coverage.

New series can be obtained by ratio validation and processing of original data:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)), x^{(0)}(k) \geq 0, \quad (4)$$
$$k = 1, 2, \ldots, n.$$

Accumulate the original quality data to weaken fluctuation and randomness that may exist in the random series to obtain the quality data accumulation series

$$\begin{cases} x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)), \\ x^{(1)}(t) = \sum_{i=1}^{t} x^{(1)}(i), t = 1, 2, \ldots, n. \end{cases} \quad (5)$$

Because the solutions of first-order differential equation show an exponential growth trend, similar to that of the sequence $x^{(1)}(t)$, the sequence $x^{(1)}$ is considered to satisfy the first-order differential equation.

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u, \quad (6)$$

where a is the development coefficient, the effective interval is $(-2, 2)$, and u is the Grey action, and both are undetermined factors. As long as the parameters a and u are obtained, $x^{(1)}(t)$ can be obtained, as can the predicted value of $x^{(0)}$.

According to the definition of derivative,

$$\frac{dx^{(1)}}{dt} = \lim_{\Delta t \to 0} \frac{x^{(1)}(t + \Delta t) - x^{(1)}(t)}{\Delta t}, \quad (7)$$

i.e.,

$$x^{(0)}(t) = -a \times \frac{1}{2}[x^{(1)}(t) + x^{(1)}(t-1)]. \quad (8)$$

The following matrices are obtained:

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \cdots \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -0.5[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -0.5[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \cdots & 1 \\ -0.5[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ u \end{bmatrix} \quad (9)$$

Let

$$\begin{cases} y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \cdots \\ x^{(0)}(n) \end{bmatrix}, \\ B = \begin{bmatrix} -0.5[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -0.5[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \cdots & 1 \\ -0.5[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{bmatrix}, \\ \hat{a} = \begin{bmatrix} a \\ u \end{bmatrix}. \end{cases} \quad (10)$$

Solving the Grey parameters by the least squares method, we obtain

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}Y_n. \quad (11)$$

We substitute $\hat{a}$ into Eq. (6) to obtain

$$\hat{x}^{(1)}(t+1) = (x^{(0)}(1) - \frac{u}{a})e^{-at} + \frac{u}{a}, \quad (12)$$

which is the time response function model of GM(1,1).

On this basis, the forecasting equation of the original quality data is obtained as

$$\begin{aligned} \hat{x}^{(0)}(t+1) &= \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) \\ &= (1 - e^a)(x^{(0)}(1) - \frac{u}{a})e^{-at}. \end{aligned} \quad (13)$$

After establishing the Grey model, it is necessary to check whether it can be used to forecast the target data. Three test methods are selected: residual, correlation, and posterior.

(1) Residual Test

Residual test is relatively intuitive and only needs to compare the predicted and original values and observe whether the relative error can meet the requirements.

The residual of the original data column $x^{(0)}(t)$ and predicted data column $\hat{x}^{(0)}(t)$ is

$$e(t) = x^{(0)}(t) - \widehat{x}^{(0)}(t). \tag{14}$$

The relative error Δ_t and average relative error $\overline{\Delta}$ are calculated as

$$\Delta_t = \left| \frac{e(t)}{x^{(0)}(t)} \right| \times 100\%, \tag{15}$$

$$\overline{\Delta} = \frac{1}{n} \sum_{t=1}^{n} \Delta_t. \tag{16}$$

The fitting accuracy is

$$p = 1 - \overline{\Delta}. \tag{17}$$

The Grey prediction model of quality data passes the residual test if p is greater than 80%.

(2) Correlation Degree Test

The correlation degree test is a geometric test to study the similarity of model curves of the original and predicted value. The more similar the geometry of the two curves, the more the values are correlated.

The correlation coefficient of the original quality data column $x^{(0)}(t)$ and predicted quality data column $\widehat{x}^{(0)}(t)$ is

$$\eta(t) = \frac{\min_{i=1}^{14} \left| \widehat{x}^{(0)}(i) - x^{(0)}(i) \right| + \rho \cdot \max_{i=1}^{14} \left| \widehat{x}^{(0)}(i) - x^{(0)}(i) \right|}{\left| \widehat{x}^{(0)}(t) - x^{(0)}(t) \right| + \rho \cdot \max_{i=1}^{14} \left| \widehat{x}^{(0)}(i) - x^{(0)}(i) \right|}, \tag{18}$$

where ρ is the resolution coefficient, usually taking a value in (0,1). A larger ρ indicates a smaller difference between correlation coefficients, and weaker discrimination ability.

The correlation degree between the original quality data column $x^{(0)}(t)$ and predicted quality data column $\widehat{x}^{(0)}(t)$ is

$$r = \frac{1}{n} \sum_{t=1}^{n} \eta(t). \tag{19}$$

The closer r is to 1, the better the forecast accuracy. If r is greater than 0.6, then Grey prediction passes the correlation degree test.

(3) Posterior Variance Test

A posterior variance test is based on the probability distribution of the residual predicted by quality data.

We calculate the variance of the original quality data s_1^2, and of the prediction residuals s_2^2

$$s_1^2 = \frac{1}{n} \sum_{t=1}^{1} [x^{(0)}(t) - \overline{x}]^2, \tag{20}$$

$$s_2^2 = \frac{1}{n} \sum_{t=1}^{1} [e^{(0)}(t) - \overline{e}]^2. \tag{21}$$

Then the ratio of the mean squared error (MSE) is

$$C = \frac{s_2}{s_1}. \tag{22}$$

The residual probability is

$$P = P\left\{ \left| e^{(0)}(t) \right| < 0.6745 s_1 \right\}. \tag{23}$$

As shown in Table 1, a smaller C and larger P indicate a more accurate Grey model. Grade I indicates the highest prediction accuracy, and Grade IV the lowest. Generally, if a prediction model is evaluated as Grade I, II, or III, it can be considered to pass the posterior variance test.

4.1.2 Revision of Predicted Residuals Using Markov Model

The assembly quality data of aerospace products are greatly affected by external information such as operators and the operating environment. This external information is considered random, so the correlation between the changes of quality data is not strong. Therefore, the Markov method is used to forecast and correct the value residuals in the Grey model.

The residuals of Grey predicted values are divided into different states, and a state transition matrix is established, which is composed of all one-step transition probabilities of random processes,
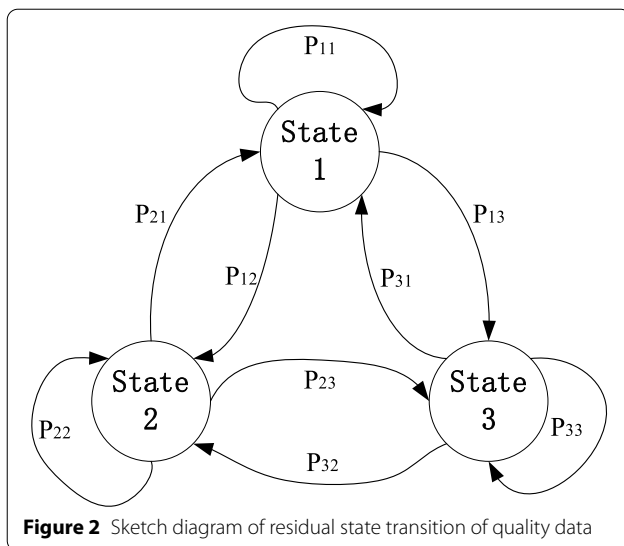
$$P = \begin{bmatrix} p_{11} \cdots & p_{1n} \\ \vdots \ddots & \vdots \\ p_{n1} \cdots & p_{nn} \end{bmatrix}, \tag{24}$$

where p_{ij} is the one-step transition probability from state i to state j.

It is worth noting that all elements in the state transition matrix are nonnegative, and they sum to 1. The

Table 1 Check table of Grey model accuracy

Grade	Average relative error q	Ratio of MSE C	Residual probability P
I	<0.01	<0.35	>0.95
II	<0.05	<0.50	<0.80
III	<0.10	<0.65	<0.70
IV	>0.20	>0.65	<0.60

Figure 2 Sketch diagram of residual state transition of quality data
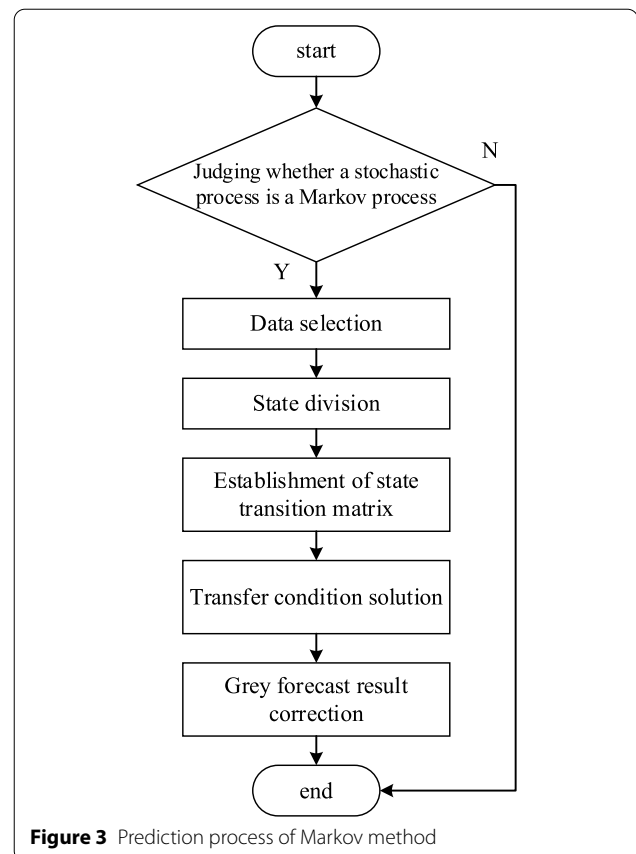

Figure 3 Prediction process of Markov method

state transition matrix is an important part of the Markov prediction model. After calculating it, the subsequent state of the residual value of quality data can be calculated according to the initial state of the residual value. The sketch diagram of the residual state transition, as shown in Figure 2, is convenient for observing the transition probability of each residual state of quality data.

Figure 3 shows the process of correcting residual values of Grey forecast results using a Markov prediction model.

After confirming that the variation of residuals of assembly quality data is a Markov process, it is necessary to collect residual data and classify the residual state. The state transfer matrix is built dynamically according to the specific changes of quality residual data and is used to solve the prediction state of the assembly quality data residuals of aerospace products. On this basis, the quality data obtained by the Grey forecasting model are corrected.

4.2 Assembly System Status Prediction Using T-K Statistical Control Chart

The T-K control chart does not require a large sample and is independent of the standard deviation of the parent. It is applicable to predict the assembly system status.

4.2.1 T-control Chart for Monitoring the Mean of Quality Data

The T-statistic is used to monitor the fluctuation of the mean value of quality data and is suitable for small samples. In the actual assembly process of aerospace products, the mean value of certain quality data is usually uncertain and will change constantly. Therefore, when constructing T statistics, it is assumed that the mean value of quality data is unknown.

X is set as the quality data items. 15 observation samples are selected for each batch of products and $\{X_{i,j,1}^{(r)}, \ldots, X_{i,j,n}^{(r)}\}$ as the group i sample, where i=1, 2, ...; j indicates the serial number of the product type corresponding to the batch sample, j=1, 2, ..., P; n is the sample size; and the superscript r indicates the serial number of a batch of the same type of product.

Quality data within a batch and between batches are independent, sample data of the same variety obey the same normal distribution, and sample data of different varieties obey different normal distributions, i.e.,

$$X_{i,j,k} \sim N(\mu_j, \sigma_j), i = 1, 2, \ldots, j = 1, 2, \ldots; j = 1, 2, \ldots, P; k = 1, 2, \ldots n. \tag{25}$$

where, μ_j and σ_j are the mean and standard deviation of the distribution of product j quality data under the controlled state. The mean and standard deviation of group i samples are

$$\overline{X}_{i,j} = \frac{1}{n}\sum_{k=1}^{n} X_{i,j,k}, \tag{26}$$

$$S_{i,j} = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(X_{i,j,k} - \overline{X}_{i,j}\right)^2}. \tag{27}$$

The mean of the first $r-1$ batches of samples of different product types are defined as

$$\overline{\overline{X}}_j^{(r-1)} = \frac{1}{r-1}\sum_{h=1}^{r-1}\bar{X}_j^{(h)}, j = 1, 2; r = 2, 3, 4. \tag{28}$$

Then the T-statistic is

$$\begin{cases} T_{i,j}^{(1)} = \bar{X}_j^{(h)}, \\ T_{i,j}^{(r)} = \frac{\bar{X}_{i,j}^{(r)} - \overline{\overline{X}}_j^{(r-1)}}{S_{i,j}^{(r)}}\sqrt{\frac{n(r-1)}{r}}, j = 1, 2, r > 1. \end{cases} \tag{29}$$

The control limits of the T-control chart are

$$\begin{cases} UCL = G_t^{-1}(1 - \frac{\alpha}{2}|n-1), \\ \qquad CL = 0, \\ \qquad LCL = -UCL, \end{cases} \tag{30}$$

where $G_t^{-1}(\cdot|n-1)$ is the inverse function of the cumulative T-distribution function with degree of freedom $n-1$, and α is the significance level. According to statistical process control theory, the upper and lower control limits correspond to the positions of $\pm 3\sigma$, so α is 0.0027.

During the assembly of aerospace products, if an assembly system is in a controlled state and the mean value of quality data does not deviate, for different kinds of products, as long as the sample size of each group is the same, T-statistics calculated from groups of sample data with the same sample size will be independent of each other, subject to the same T-distribution, and with the same control limits. T-statistics and control limits calculated from each batch of quality data can be used to plot the T-control chart and monitor assembly quality.

4.2.2 K-control Chart for Monitoring the Standard Deviation of Quality Data

The K-statistic is used to monitor the fluctuation of the standard deviation of quality data and is suitable for small samples. Similar to the process of calculating T-statistics, the mean of quality data is considered unknown.

The mean of the variance of sample quality data for the first $r-1$ batches of different product types is

$$\overline{S}_j^{2\,(r-1)} = \frac{1}{r-1}\sum_{h=1}^{r-1}S_j^{2\,(h)}, j = 1, 2; r = 2, 3, 4. \tag{31}$$

The intermediate variable is defined as

$$\lambda_{i,j}^{(r)} = \frac{S_{i,j}^{2(r)}}{\overline{S}_j^{2\,(r-1)}}, j = 1, 2; r = 2, 3, 4. \tag{32}$$

The K statistic is

$$K_{i,j}^{(1)} = 1, K_{i,j}^{(r)} = \Phi^{-1}[F_{n-1,(n-1)(r-1)}(\lambda_{i,j}^{(r)})], j = 1, 2, r > 1, \tag{33}$$

where $F_{v1,\,v2}$ are cumulative distribution functions of the F distribution with first and second degrees of freedom v1 and v2, respectively.

Since the intermediate variable in Eq. (33) follows the F-distribution with first degree of freedom $(n-1)$ and second degree of freedom $(n-1)$ $(r-1)$, the K-statistics are independent of each other and follow a normal distribution, and the control limits of the K-control chart are:

$$\begin{cases} UCL = 3, \\ \ CL = 0, \\ LCL = -3. \end{cases} \tag{34}$$

4.3 Association Rules Mining for Quality Exceptions Based on Apriori Algorithm

4.3.1 Analysis of Quality Influencing Factors

Quality abnormalities occurring in the assembly process of aerospace products may be caused by human, equipment, or environmental factors in the preassembly, assembly, or inspection stage, and the influencing factors of quality abnormalities differ by their type. We use the 5M1E analysis method to classify the influencing factors of quality anomalies into six categories.

(a) Man. The operator or inspector's knowledge of quality, health condition, technical proficiency, and other factors may cause abnormal quality.

(b) Machine. Precision and maintenance of equipment can affect quality. Because the volume of aerospace products is large, the equipment of each workstation is usually fixed, so machine factors relate to an assembly workstation.

(c) Material. The composition and physical and chemical properties of materials may cause abnormal quality.

(d) Method. The assembly process, fixture selection, and operating procedures can affect quality. Assembly processes vary by product.

(e) Measurement. The measurement method adopted for use with inspection equipment can cause a quality exception.

(f) Environment. Temperature, humidity, lighting, and cleaning conditions in the workplace are possible factors affecting quality abnormalities.

4.3.2 Principle of Apriori Algorithm

An association rules algorithm is used to mine the hidden correlation behind complex data, the classic being the Apriori algorithm, which is used as follows. Filter all items in a transaction dataset that are greater than or equal to the minimum support degree. Then, association rules are generated based on the most frequent item set and filtered according to the minimum confidence level to obtain strong association rules. According to shop-floor assembly data, Apriori was selected to determine the relationships between influencing factors and assembly quality data values and system status.

An association rule is evaluated by support, confidence, and lift. Support is the probability that items X and Y occur simultaneously in a project set, i.e., the ratio of the number of items including X and Y to the number of all items. It describes the universality of association rules and is calculated as

$$support(X\&Y) = T(X\&Y)/T. \qquad (35)$$

Confidence is the probability that item Y will occur when item X occurs, i.e., the ratio of the number of items containing X and Y to the number of items containing X. It describes the authenticity of association rules and is calculated as

$$confidence(X \rightarrow Y) = T(X\&Y)/T(X). \qquad (36)$$

Lift is a parameter that describes the probability change of item Y occurrence due to the emergence of item X, i.e., the ratio of confidence to support of item $X \rightarrow Y$,

$$lift(X \rightarrow Y) = confidence(X \rightarrow Y)/support(X\&Y). \qquad (37)$$

If the lift is 1, then there is no correlation between the two events; if it is less than 1, then events X and Y are incompatible.

5 Case Study and System Implementation

5.1 Numerical Prediction Results of Quality Data

5.1.1 Data Selection

Taking the assembly process of an aerospace product as an example, the centroid offset data of cabin A of model S1 are collected to verify the proposed key technologies and method. It is known that cabin A of model S1 has completed three batches of assembly since the last production environment change. In each batch 15 products are assembled, and the assembly of the 15th product in the fourth batch is currently underway. If the assembly time of each product is the same, the allowable error of the centroid offset of cabin A is 15 mm. The collected centroid offset data of cabin A of the fourth batch are shown in Table 2.

From the above data, we use 13 centroid offset data items, from A4-1 to A4-13 as raw data, A4-14 as validation data, and A4-15 as predicted data. The original series of centroid offset are obtained:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2) \ldots x^{(0)}(13))$$
$$= (5.918, 5.205, 4.393, 4.520, 5.105, 5.818, 6.647,$$
$$7.560, 8.467, 9.642, 10.811, 11.898, 13.548). \qquad (38)$$

The scale ratio of the original series $x^{(0)}$ for centroid offset is calculated as

$$\lambda(t) = (1.137, 1.185, 0.972, 0.885, 0.878, 0.875,$$
$$0.879, 0.893, 0.878, 0.892, 0.909, 0.878). \qquad (39)$$

According to the Grey prediction model, the tolerant coverage interval of the ratio is obtained as

$$Y = (e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}}) = (0.867, 1.154). \qquad (40)$$

By observing the scale ratio of the centroid offset, it is found that each item of the ratio series falls within the admissible coverage interval, which indicates that the GM(1,1) model is applicable for the original series of centroid offsets.

Table 2 Centroid offset data of fourth batch of cabin A

Product code	Centroid offset(mm)	Product code	Centroid offset(mm)
A4-1	5.918	A4-8	7.560
A4-2	5.205	A4-9	8.467
A4-3	4.393	A4-10	9.642
A4-4	4.520	A4-11	10.811
A4-5	5.105	A4-12	11.898
A4-6	5.818	A4-13	13.548
A4-7	6.647	A4-14	14.537

5.1.2 Establishment of Grey Forecasting Model

The original sequence of the centroid offset is accumulated to weaken the volatility and randomness that may exist in the random sequence, and the accumulated sequence of centroid offsets is obtained as

$$x^{(1)} = (5.918, 11.123, 15.515, 20.036, 25.141,$$
$$30.959, 37.605, 45.165, 53.633, 63.275, \tag{41}$$
$$74.086, 85.984, 99.532).$$

According to the Grey model, the cumulative sequence of centroid offsets is processed, and the matrix $\boldsymbol{B}$ and constant vector $\boldsymbol{y}_n$ are obtained:

$$\boldsymbol{B} = \begin{bmatrix} -0.5(X^{(1)}(1) + x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \cdots & 1 \\ -0.5(x^{(1)}(12) + x^{(1)}(13)) & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -0.5(5.918 + 11.123) & 1 \\ -0.5(11.123 + 15.515) & 1 \\ \cdots & 1 \\ -0.5(85.984 + 99.532) & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -8.520 & 1 \\ -13.319 & 1 \\ -17.776 & 1 \\ -22.588 & 1 \\ -28.050 & 1 \\ -34.282 & 1 \\ -41.385 & 1 \\ -49.399 & 1 \\ -58.454 & 1 \\ -68.681 & 1 \\ -80.035 & 1 \\ -92.758 & 1 \end{bmatrix} \tag{42}$$

$$\boldsymbol{y_n} = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \cdots \\ x^{(0)}(13) \end{bmatrix} = \begin{bmatrix} 5.205 \\ 4.393 \\ 4.520 \\ 5.105 \\ 5.818 \\ 6.647 \\ 7.560 \\ 8.467 \\ 9.642 \\ 10.811 \\ 11.898 \\ 13.548 \end{bmatrix}. \tag{43}$$

The Grey parameter matrix $\hat{\boldsymbol{a}}$ is calculated as

$$\hat{\boldsymbol{a}} = \begin{bmatrix} a \\ u \end{bmatrix} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}X_n = \begin{bmatrix} -0.112644 \\ 3.009679 \end{bmatrix}. \tag{44}$$

Substituting $\hat{\boldsymbol{a}}$ into the prediction model of the centroid offset, the prediction function of $\hat{X}^{(1)}(t)$ is obtained as

Table 3 Fitting results using GM(1,1) model

Product code	Original value (mm)	Predicted value (mm)	Residual (mm)	Relative error (%)
A4-1	5.918	5.918	0	0
A4-2	5.205	3.891	1.314	25.24
A4-3	4.393	4.355	0.037	0.85
A4-4	4.520	4.875	− 0.354	7.84
A4-5	5.105	5.456	− 0.351	6.87
A4-6	5.818	6.106	− 0.289	4.96
A4-7	6.647	6.835	− 0.188	2.83
A4-8	7.560	7.649	− 0.089	1.18
A4-9	8.467	8.562	− 0.094	1.11
A4-10	9.642	9.582	0.060	0.62
A4-11	10.811	10.725	0.086	0.79
A4-12	11.898	12.004	− 0.106	0.89
A4-13	13.548	13.435	0.113	0.84

$$\hat{x}^{(1)}(t+1) = \left(y^{(0)}(1) - \frac{u}{a}\right)e^{-at} + \frac{u}{a}$$
$$= 32.637e^{0.112644t} - 26.719. \tag{45}$$

Expressions $\hat{x}^{(1)}(t+1)$ and $\hat{x}^{(1)}(t)$ are discretized and restored to the original sequence of centroid offsets, whose prediction sequence is

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t)$$
$$= 32.637(e^{0.112644t} - e^{0.112644(t-1)}). \tag{46}$$

The predicted serial value of centroid offsets is obtained as

$$\hat{x}^{(0)} = (5.918, 3.891, 4.355, 4.875, 5.456, 6.106,$$
$$6.835, 7.649, 8.562, 9.582, 10.725, \tag{47}$$
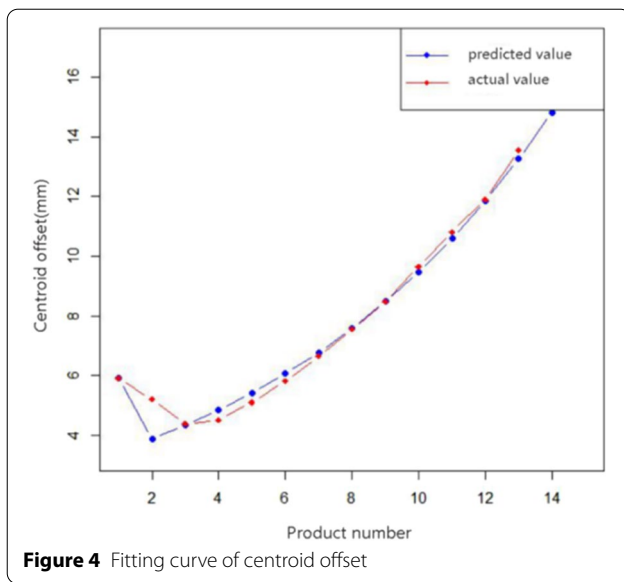$$12.004, 13.435, 15.037, 16.830).$$

5.1.3 Grey Prediction Model Test

(1) Residual Test

The predicted and original series of centroid offsets are put together, and the residual and relative error of each data item is calculated to obtain the fitting results in Table 3.

The fitting curve of centroid offsets drawn from the above fitting results is shown in Figure 4.

The calculated average relative error is 4.16%, and the fitting accuracy is 95.84%. Because the fitting accuracy is more than 80%, the forecast result passes the residual test.

Zhuang *et al. Chinese Journal of Mechanical Engineering*   (2022) 35:105

Page 12 of 21

Figure 4 Fitting curve of centroid offset

(2) Correlation Degree Test

According to the above fitting curve, $\min_{i=1}^{14}|\hat{y}^{(0)}(i) - y^{(0)}(i)| = 0, \max_{i=1}^{14}|\hat{y}^{(0)}(i) - y^{(0)}(i)| = 1.314$, and $\rho = 0.5$. According to Eq. (18), the corresponding correlation coefficients are

$$\eta(t) = (1, 0.333, 0.946, 0.650, 0.652, 0.695, 0.778,$$
$$0.880, 0.875, 0.916, 0.84, 0.861, 0.853). \quad (48)$$

The correlation degree between the two sets is 0.794, which is greater than 0.6; hence, the prediction result passes the correlation degree test.

(3) Posterior Deviation Test

The column of residuals extracted from the test results is

$$e^{(0)}(t) = (0, 1.314, 0.037, -0.354, -0.351,$$
$$- 0.289, -0.188, -0.089, -0.094, \quad (49)$$
$$0.060, 0.086, -0.106, 0.113).$$

We calculate that the variance of the original series $y^{(0)}(t)$ of the centroid offset is $s_1^2 = 8.284$, the variance of the residuals is $s_2^2 = 0.165$, and the ratio of the mean variance is $C = 0.141$. According to the definition, the probability of small error is $P = 1$.

Referring to Table 1, we can see that the mean variance and small probability error calculated above are in the range of Level I, but the average relative error is in the range of Level II.

5.1.4 Case of Grey-Markov Model

Based on the Grey prediction results, it can be found that the minimum residual error of the predicted and

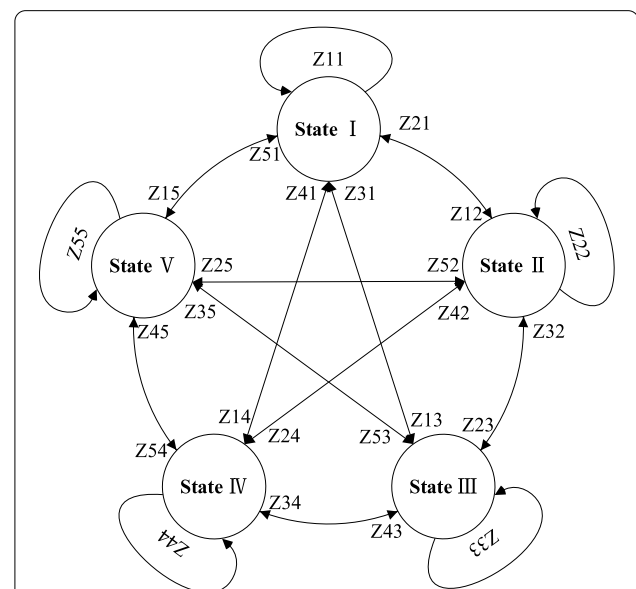**Table 4** State division of residual error value

Interval	State
[− 0.36, − 0.24]	State I: too large
[− 0.24, − 0.12]	State II: larger
[− 0.12, 0]	State III: slightly larger
[0, 0.12]	State IV: slightly smaller
[0.12, 1.32]	State V: smaller

actual values is −0.354, and the maximum is 1.314, i.e., the values of residual error are within the range [−0.36, 1.32].

According to the distribution of residual error values, the above large interval is divided into five subintervals: [−0.36, −0.24], [−0.24, −0.12], [−0.12, 0], [0, 0.12], [0.12, 1.32]. Each interval corresponds to a residual error state of the centroid offset, as shown in Table 4.

It is assumed that the predicted result of centroid offset will transfer between States I and V with a certain probability. The state transition diagram of the predicted results is shown in Figure 5. If the current forecasting result is State I, then the next forecasting result has a Z11 probability become State I, a Z12 probability become State II, a Z13 probability become State III, a Z14 probability become State IV, and a Z15 probability become State V.

From the column of residual error values of the centroid offset, we can see that State I occurs three times,

Figure 5 Diagram of state transition of residual error value of centroid offset

Table 5 Statistical results of state transfer of residual error

P_{ij}		Next moment residual status					Sum
		State I	State II	State III	State IV	State V	
Current residual status	State I	2	1	0	0	0	3
	State II	0	0	1	0	0	1
	State III	0	0	1	2	0	3
	State IV	2	0	1	1	1	5
	State V	0	0	0	1	0	1

Table 6 Comparison between Grey model and Grey-Markov model

Product code	Actual value (mm)	Grey model		Grey-Markov model	
		Predicted value (mm)	Relative error (%)	Predicted value (mm)	Relative error (%)
A4-14	14.537	15.037	3.44	14.737	1.38

twice transferred to State I and once to State II; State IV occurs five times, transferred once each to States I, III, IV, V, and the final State (the residual error of the predicted value at the moment of the final State is left of the interval where State I is located, which we consider as State I). The state transition of residual error is shown in Table 5.

It is assumed that the time interval of the measurement for each two cabins' centroid offset is one assembly cycle, i.e., the step size of the Markov model, which is the time required for state transition. According to the statistics above, the state transition matrix of the residual error can be obtained as

$$P = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & Z_{14} & Z_{15} \\ Z_{21} & Z_{22} & Z_{23} & Z_{24} & Z_{25} \\ Z_{31} & Z_{32} & Z_{33} & Z_{34} & Z_{35} \\ Z_{41} & Z_{42} & Z_{43} & Z_{44} & Z_{45} \\ Z_{51} & Z_{52} & Z_{53} & Z_{54} & Z_{55} \end{bmatrix} = \begin{bmatrix} \frac{P_{11}}{P_1} & \frac{P_{12}}{P_1} & \frac{P_{13}}{P_1} & \frac{P_{14}}{P_1} & \frac{P_{15}}{P_1} \\ \frac{P_{21}}{P_2} & \frac{P_{22}}{P_2} & \frac{P_{23}}{P_2} & \frac{P_{24}}{P_2} & \frac{P_{25}}{P_2} \\ \frac{P_{31}}{P_3} & \frac{P_{32}}{P_3} & \frac{P_{33}}{P_3} & \frac{P_{34}}{P_3} & \frac{P_{35}}{P_3} \\ \frac{P_{41}}{P_4} & \frac{P_{42}}{P_4} & \frac{P_{43}}{P_4} & \frac{P_{44}}{P_4} & \frac{P_{45}}{P_4} \\ \frac{P_{51}}{P_5} & \frac{P_{52}}{P_5} & \frac{P_{53}}{P_5} & \frac{P_{54}}{P_5} & \frac{P_{55}}{P_5} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{2}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{50}$$

According to the state of residual error, the residual error of prediction values for A4–13 belongs to state IV, so the initial state can be considered to be (0, 0, 0, 1, 0). According to the state transition matrix of Eq. (50), the state after one-step transition is (2/5, 0, 1/5, 1/5, 1/5), so the residual error of the prediction value of A4-14 is state I.

We correct the predicted centroid offset value of A4-14 from the Grey model:

$$A4 - 14 \text{ predicted value} = 15.037 + \frac{1}{2} * (0.24 + 0.36) = 14.737. \tag{51}$$

The results are compared with those of the Grey prediction model, as shown in Table 6.

It is found that, unlike the Grey prediction model, the Grey-Markov prediction model takes into account the fluctuation of data, and it reduces the relative error of centroid offset prediction from 3.44% to 1.38%, which improves the accuracy of prediction.

Similarly, we calculated the state after two steps of transfer of A4-13 (0, 0, 0, 0, 1, 0) as (0.347, 0.133, 0.107, 0.373, 0.040), so the remaining error of A4-15 prediction is State IV.

We correct the Grey forecast value of A4-15:

$$A4 - 15 \text{ predicted value} = 16.830 + \frac{1}{2} * (0 + 0.12) = 16.890. \tag{52}$$

Therefore, it forecasts that the value of the centroid offset for cabin A of the 15th product will be 16.890 mm, exceeding the allowable error of 15 mm, and abnormal quality will occur soon.

5.2 Assembly System Status Prediction Results
5.2.1 Raw Data Preprocessing
Based on 14 centroid offset data items of the fourth batch of cabin A of model S1 in Table 2 and the predicted data of product 15, we can obtain 15 data items for the fourth batch of cabin A. With the centroid offset data of the first three batches of cabin A of model S1, 60 pieces of centroid offset data are obtained, as shown in Table 7, which are used as part of the sample data to

Zhuang *et al. Chinese Journal of Mechanical Engineering*   (2022) 35:105

Page 14 of 21

forecast the status of the assembly system by means of statistical process control.

Drawing on the idea of group technology, we collect the data of cabin A of model S2 similarly to those of cabin A of model S1. The other half of the sample data, the centroid offset data of four batches of cabin A of model S2, are shown in Table 8.

All the sample data required by the statistical process are obtained so that T-K control charts can be used to monitor the mean and standard deviation of the

centroid data and evaluate the stability of the assembly system.

5.2.2 Case of T Control Chart

In the selected case, we know the allowable error range of the centroid offset of cabin A, but we do not know the mean value, which satisfies the condition for the use of the calculation formula of the T statistic in Section 4.2.1. According to the original data in Tables 7 and 8, we first calculate the mean and standard deviation of group i

Table 7 Centroid offset data for batches 1–4 of cabin A of model S1
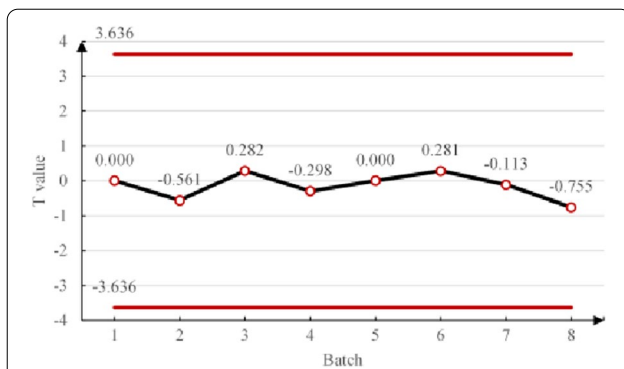
Batch	A1	A2	A3	A4	A5	A6	A7	A8
1	8.691	9.571	4.126	4.515	6.336	8.937	9.819	10.021
2	4.366	7.135	3.126	3.725	4.210	7.979	9.190	9.322
3	7.978	4.715	5.226	6.687	7.804	8.045	8.159	9.188
4	5.918	5.205	4.393	4.520	5.105	5.818	6.647	7.560
Batch	A9	A10	A11	A12	A13	A14	A15	Model
1	11.005	11.365	11.472	11.479	12.147	13.255	13.672	S1
2	10.083	10.124	10.864	11.081	11.677	12.781	12.973	S1
3	9.729	10.782	10.868	10.983	13.653	14.657	16.948	S1
4	8.467	9.642	10.811	11.898	13.548	14.537	16.890	S1

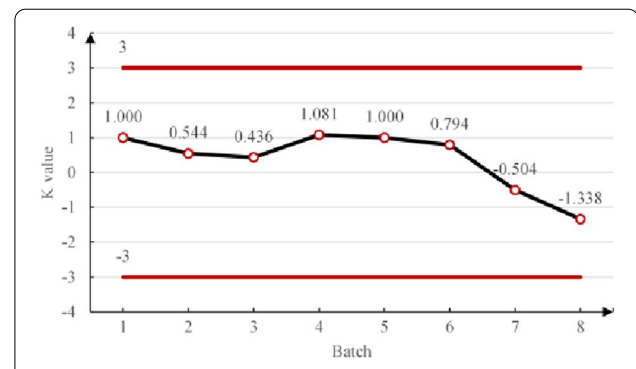**Table 8** Centroid offset data for batches 1–4 of cabin A of model S1

Batch	A1	A2	A3	A4	A5	A6	A7	A8
1	5.310	7.491	7.949	8.365	8.524	9.299	9.417	9.757
2	6.066	6.595	7.975	8.008	8.349	9.148	9.873	10.051
3	4.178	7.442	7.571	8.960	9.793	9.934	10.292	10.467
4	6.866	7.177	7.336	7.453	8.121	8.557	8.928	9.636
Batch	A9	A10	A11	A12	A13	A14	A15	Model
1	10.025	10.824	11.122	11.294	11.492	12.967	14.223	S2
2	10.785	10.910	12.261	12.340	13.484	13.948	15.688	S2
3	10.641	10.682	11.449	11.575	11.891	12.043	12.792	S2
4	9.791	9.966	11.192	11.394	11.395	11.445	11.604	S2

Table 9 Calculation results of T-statistic

Batch	A1	A2	…	A15	Model	Mean	Standard deviation	$\overline{\overline{x}}_j$	T
1	8.691	9.571	…	13.672	S1	9.761	2.880	–	0.000
2	4.366	7.135	…	12.973	S1	8.576	3.340	9.761	− 0.561
3	7.978	4.715	…	16.948	S1	9.695	3.410	9.168	0.282
4	5.918	5.205	…	16.890	S1	8.731	3.991	9.344	− 0.298
5	5.310	7.491	…	14.223	S2	9.871	2.244	–	0.000
6	6.066	6.595	…	15.688	S2	10.365	2.787	9.871	0.281
7	4.178	7.442	…	12.792	S2	9.981	2.216	10.118	− 0.113
8	6.866	7.177	…	11.604	S2	9.391	1.747	10.072	− 0.755

Zhuang *et al. Chinese Journal of Mechanical Engineering*   (2022) 35:105

Page 15 of 21

Figure 6 T-control chart of centroid offset of cabin A

Figure 7 K-control chart of centroid offset

samples, the mean $\overline{\overline{X_j}}$ of the first r-1 batches of samples, and then the T statistics, as shown in Table 9.

According to Eq. (30), we set n to 15, and the upper control limit of the T control chart is 3.636, the lower control limit is −3.636, and the center line is 0. We draw the T control chart of the centroid offset of cabin A, as shown in Figure 6, from which it can be found that the mean of the centroid offset for cabin A of models S1 and S2 is within the allowable range.

5.2.3 Case of K-control Chart

Similar to the calculation of the T-statistic, we only know the allowable range of the error of the centroid offset, and we do not know the average, which satisfies the condition of calculating the K-statistic as in Section 4.2.2. Based on the original data in Tables 7 and 8, we calculate the mean $\overline{S_j^2}$ of the variance of the first r-1 batches of sample data for different product types, and then calculate the intermediate variable $\lambda_{i,j}^{(r)}$ and the K statistic, as shown in Table 10.

According to Table 10 and Eq. (33), the K-control chart of the centroid offset is shown in Figure 7.

Observing the K-control chart, it can be found that the standard deviation of the centroid offset is within the allowable range. Combined with the results of the

T-control chart in Section 5.2.2, the assembly system can be considered to be in a controlled state according to the sample data. Since the centroid offset data of the 15th product of the fourth batch are a predicted value, the predicted results based on the T-K control chart indicate that the status of the assembly system at the next moment is controlled.
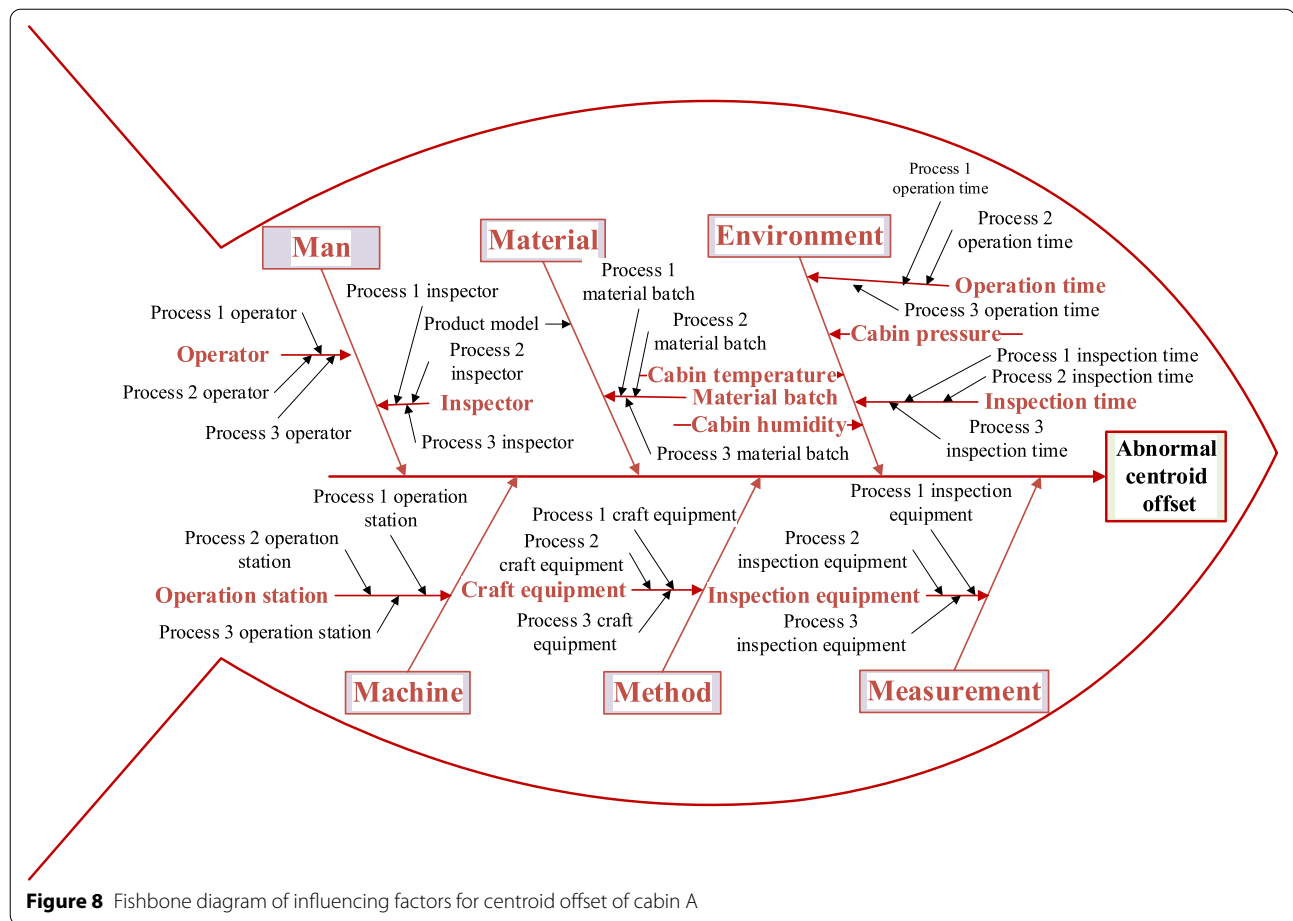
5.3 Mining Association Rules of Influence Factors for Abnormal Quality Data

Assuming that the assembly process of cabin A in this case has three processes, all kinds of materials required in each process are from the same batch, and all equipment of each station is fixed. According to the six categories of quality influencing factors mentioned above, the fishbone diagram of the influencing factors for the centroid offset of cabin A is shown in Figure 8, which includes 28 factors.
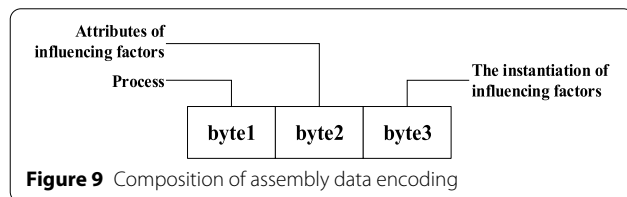
Although 28 factors are identified from the analysis, how these are combined to cause an abnormal centroid offset still must be determined. We take 150 pieces of quality data collected from the assembly process of a certain cabin A as an example, as shown in Additional file 1: Appendix 1.

Table 10 Calculation results of K statistic

Batch	A1	A2	…	A15	Model	Variance	$\overline{S_j^2}$	$\lambda_{i,j}^{(r)}$	K
1	8.691	9.571	…	13.672	S1	8.295	–	–	1.000
2	4.366	7.135	…	12.973	S1	11.157	8.295	1.345	0.544
3	7.978	4.715	…	16.948	S1	11.628	9.726	1.196	0.436
4	5.918	5.205	…	16.890	S1	15.927	10.360	1.537	1.081
5	5.310	7.491	…	14.223	S2	5.033	–	–	1.000
6	6.066	6.595	…	15.688	S2	7.766	5.033	1.543	0.794
7	4.178	7.442	…	12.792	S2	4.912	6.400	0.768	− 0.504
8	6.866	7.177	…	11.604	S2	3.052	5.904	0.517	− 1.338

Figure 8 Fishbone diagram of influencing factors for centroid offset of cabin A

Figure 9 Composition of assembly data encoding

To keep the centroid offset within the qualified range as much as possible, the ideal error range of centroid offset is determined as within 12 mm after the influencing factors are analyzed. Over 12 mm is considered to be abnormal, which requires attention. The data encoding is shown in Figure 9.
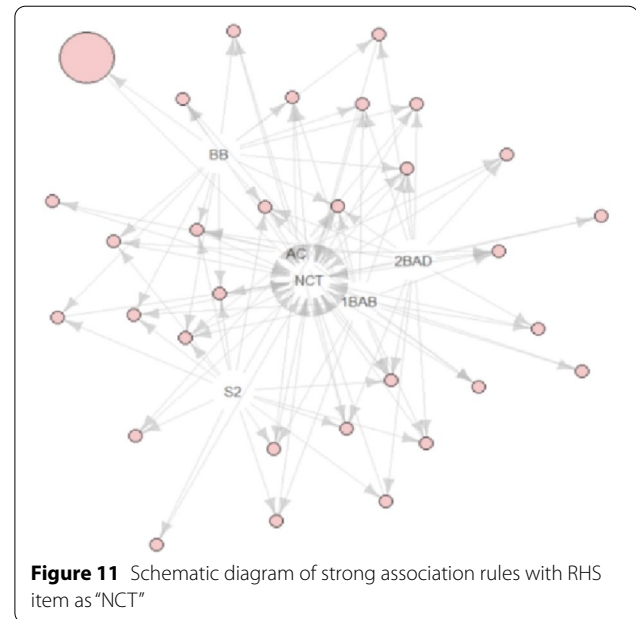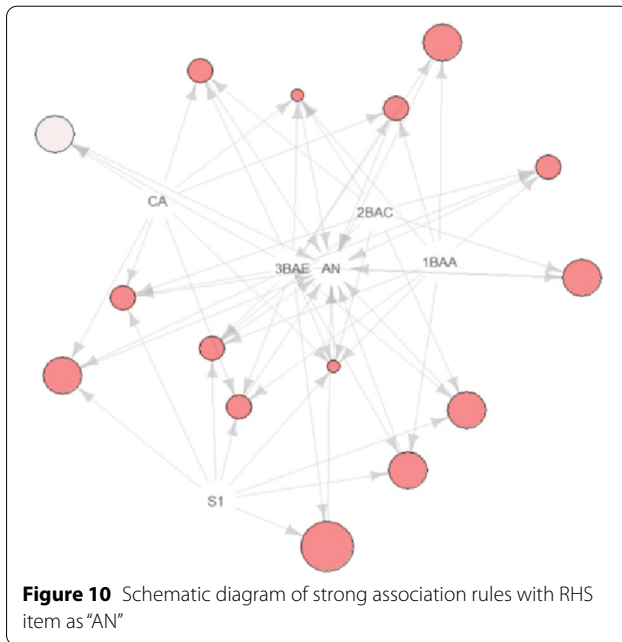
Each encoding consists of three bytes. Byte 1 represents the process to which it belongs, such as processes 1–3. Byte 2 represents the attributes of influencing factors, such as operators and inspectors. Byte 3 represents the instantiation of influencing factors, such as operator A and inspector B. It is worth noting that when bytes 2 and 3 of two encodings are the same, the corresponding objects are the same regardless of whether byte 1 is the same. For example, 1OPE and 2OPE both represent operator E, who performs process 1 in the first case and process 2 in the second. A total of 108 codes appear in Additional file 1: Appendix 1, and their meanings are given in Additional file 2: Appendix 2.

According to the principle of the Apriori algorithm, association rules are mined from 150 pieces of assembly quality data in Additional file 1: Appendix 1 by R system. We set the support degree to 0.16 and the minimum length of item to 2, to obtain 7250 frequent item sets, from which 15 strong association rules with RHS item "AN" are mined, as shown in Figure 10. The depth of color indicates the degree of lifting, and the size of circle indicates the size of support degree.

Strong association rules with RHS item "AN" are arranged in descending order of lifting in Table 11.

Observing the above rules, the lifts of the top 14 rules are greater than 3, which means they are effective. It is found that the probability of abnormal centroid offset is relatively high when the material batch of process 3 is E. Also, when the material batch of process 1 is A, the

Zhuang *et al. Chinese Journal of Mechanical Engineering*   (2022) 35:105

Page 17 of 21

Figure 10 Schematic diagram of strong association rules with RHS item "AN"

Figure 11 Schematic diagram of strong association rules with RHS item "NCT"

Table 11 Strong Association Rule with RHS item "AN"

	Left-hand-side (LHS)	Right-hand-side (RHS)	Support	Confidence	Lift
[1]	{1BAA,3BAE}	=> {AN}	0.180	1	3.191
[2]	{3BAE,S1}	=> {AN}	0.187	1	3.191
[3]	{2BAC,3BAE}	=> {AN}	0.180	1	3.191
[4]	{1BAA,3BAE,CA}	=> {AN}	0.173	1	3.191
[5]	{3BAE,CA,S1}	=> {AN}	0.180	1	3.191
[6]	{2BAC,3BAE,CA}	=> {AN}	0.173	1	3.191
[7]	{1BAA,3BAE,S1}	=> {AN}	0.180	1	3.191
[8]	{1BAA,2BAC,3BAE}	=> {AN}	0.173	1	3.191
[9]	{2BAC,3BAE,S1}	=> {AN}	0.180	1	3.191
[10]	{1BAA,3BAE,CA,S1}	=> {AN}	0.173	1	3.191
[11]	{1BAA,2BAC,3BAE,CA}	=> {AN}	0.167	1	3.191
[12]	{2BAC,3BAE,CA,S1}	=> {AN}	0.173	1	3.191
[13]	{1BAA,2BAC,3BAE,S1}	=> {AN}	0.173	1	3.191
[14]	{1BAA,2BAC,3BAE,CA,S1}	=> {AN}	0.167	1	3.191
[15]	{3BAE,CA}	=> {AN}	0.180	0.931	2.971

product model is S1, the material batch of process 2 is C, or the pressure in the cabin is A. Managers can try to avoid the combination of the above rules.

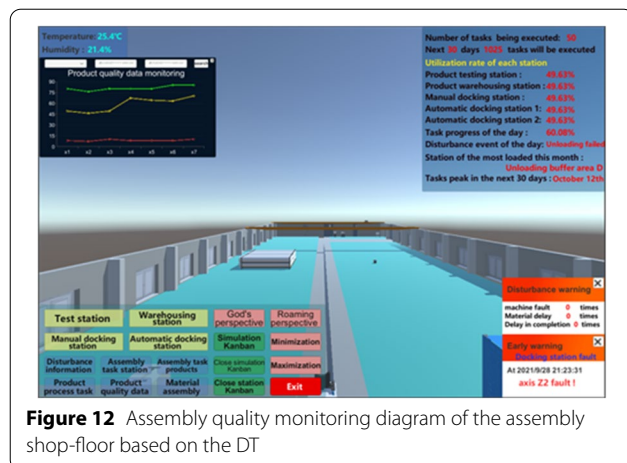### 5.4 Mining Association Rules of Influence Factors for Assembly System Status

Because no strong association rule with RHS item "NCT" is found, 30 uncontrolled data items of the assembly system are filtered in 150 collected pieces of data.

Mining association rules from this data, we set the support to 0.66, and the minimum length of item set to 2, from which 7250 frequent item sets are obtained. From these, 31 association rules with RHS item "NCT" are mined, as shown in Figure 11.

5.5 System Implementation

The proposed quality management method is achieved through data value prediction, assembly system status prediction, and association rule mining based on the constructed DT model, providing early warning for the assembly shop-floor and avoiding future quality abnormalities. According to the comparison of the results in Table 6, the Grey-Markov model considers the fluctuation of data compared with the general Grey prediction model, thus improving the accuracy of data prediction. A DT-based quality management system for the assembly process of aerospace products is developed, which achieves the monitoring of quality information and quick tracing of quality problems, thereby reducing the occurrence frequency and improving the processing efficiency of quality problems on the shop-floor. Currently, the system has been applied in an aerospace enterprise, and the quality data value of more than seven types of key indicators on a final assembly shop-floor and their trend for different products are monitored and predicted.

The interface for monitoring assembly shop-floor quality based on the DT is shown in Figure 12. The lower left quarter quadrant enables the monitoring of the operating status of the equipment, the bill of arrived materials, and the values of key quality indicators of the current

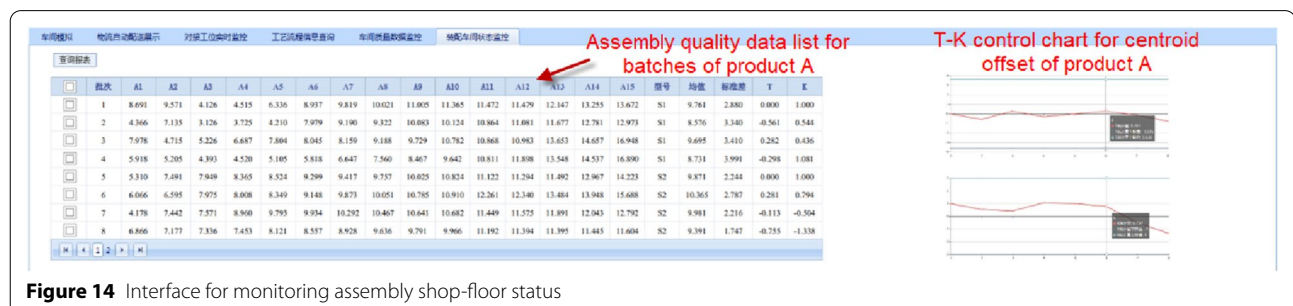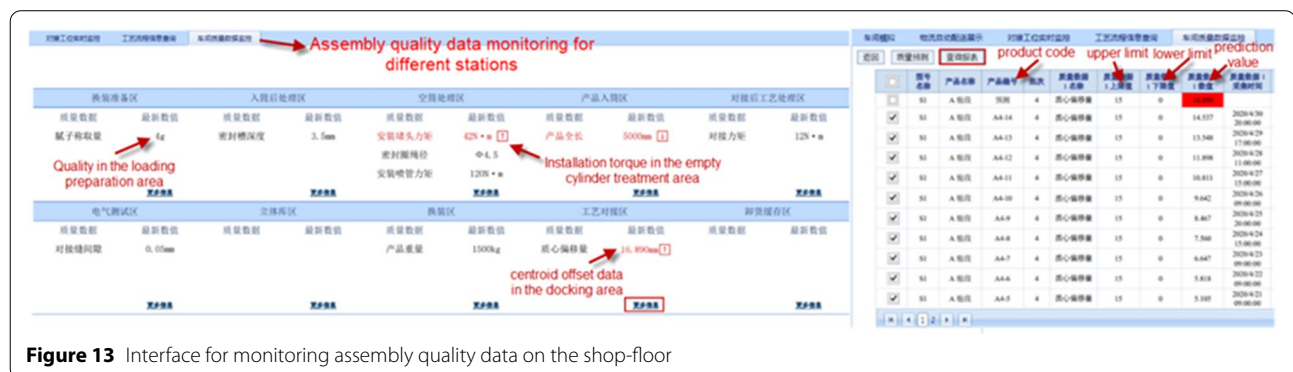**Figure 12** Assembly quality monitoring diagram of the assembly shop-floor based on the DT

station by clicking the corresponding button. The top-left corner monitors quality data value and change trends for different indicators. The lower-right corner is a pop-up window for early warning if an exception occurs or may occur. The upper-right corner shows the prediction results such as the utilization rate of each station and maximum load area.

The interface for monitoring all of the assembly quality data on the assembly shop-floor is as shown in Figure 13, including the prediction results at the next moment based on the built Grey-Markov model. On the left side, the assembly quality data monitoring for different stations on the shop-floor can be realized, including the loading preparation area, empty cylinder treatment area, and docking area. For a specific station, the quality data are monitored on the right side of the interface, and the warning prompt appears in a red font on the interface for monitoring shop-floor quality data on the left side, and on a red background on the right ride of the interface for monitoring station quality data. Quality abnormality records are automatically generated.

The interface for monitoring assembly shop-floor status is shown in Figure 14. The left side is an assembly quality data list for batches of aerospace product A. In the list, the values of key quality data of historical batches can be viewed, along with collected real-time data and predicted data of the current batch. Moreover, users also can monitor the mean, standard deviation, and T and K statistics of quality data in each batch. Meanwhile, users can view the T-K statistical diagram on the right side to observe whether the current status of the assembly system is controlled. If the T-K statistics of a batch exceed the upper or lower limits, then quality abnormality records will be automatically generated.

When dealing with quality problems, a craftsman or inspector can switch to the interface of managing strong association rules for quality exceptions, as shown in Figure 15. The left side is product BOM, through which quality data and rules are managed. The upper right is a search bar. Select product model, quality data, support,

Figure 13 Interface for monitoring assembly quality data on the shop-floor

Figure 14 Interface for monitoring assembly shop-floor status

Figure 15 Strong association rules for quality anomalies

confidence, and other conditions to search and view these rules which are listed in the lower right of the interface, and then the reference methods of processing quality problems can be provided. The list of strong association rules contains the combination of shop-floor resources that lead to abnormal quality data, which are obtained through the proposed approach and imported into the knowledge base in Excel format.

6 Conclusions and Future Work

The assembly process of aerospace products has the characteristics of single- or small-batch production and many exceptions. How to control assembly quality and improve production efficiency has been an issue in engineering applications. By introducing DT technology, the advantages of digital space, including low cost, high efficiency, and predictability, can be fully utilized to improve the management and control capability of the quality in the assembly process of aerospace products. The main contributions of this article include the following:

(1) To obtain more accurate predictions in the small sample volume of aerospace product quality data, the fluctuations of data are considered and a Grey-Markov model-based quality data prediction algorithm was presented. Moreover, group technology was used to increase the sample size of quality data and a T-K control chart was applied to obtain the mean and standard deviation of quality data, real-izing the prediction of assembly system status and avoiding quality problems caused by an uncontrolled assembly system.

(2) To improve the efficiency of quality problem tracing and handling in the assembly process of aerospace products, an Apriori algorithm-based traceability method of quality anomaly influencing factors was proposed. Strong association rules related to quality data anomalies and uncontrolled assembly systems were mined to trace the influencing factors of quality anomalies, which can assist related personnel in quickly locating abnormal causes and improving the efficiency of quality control.

(3) A DT-based quality management system for the assembly process of aerospace product is developed and has been applied in an aerospace enterprise, which promotes the application of DT in the assembly quality management.

This paper explores the application of DT to improve the quality management and problem tracing capability for the assembly process of aerospace products. Future research will focus on two areas: ① deeply studying the feedback link in quality control for the assembly process of aerospace products, so as to make the feedback process in real time and more automatic and intelligent; ② using text mining algorithms to obtain more quality problem processing knowledge, so as to assist in the rapid processing and decision-making of quality problems.

Zhuang *et al. Chinese Journal of Mechanical Engineering*   (2022) 35:105

Page 20 of 21

Supplementary Information

Acknowledgements

Author contributions
JL was in charge of the whole trial; CZ wrote the manuscript; ZL, HM, SZ, and YW assisted with sampling and laboratory analyses. All authors read and approved the final manuscript.

Authors' Information
Cunbo Zhuang, born in 1991, is currently an associate research fellow at *Laboratory of Digital Manufacturing, School of Mechanical Engineering, Beijing Institute of Technology, China*. He received his PhD degree from *Beijing Institute of Technology, China*, in 2018. His research interests include digital twin, manufacturing execution system, and shop scheduling.

Ziwen Liu, born in 1999, is currently a master's degree candidate at *Beijing Institute of Technology, China*.

Jianhua Liu, born in 1977, is currently a professor at *Laboratory of Digital Manufacturing, School of Mechanical Engineering, Beijing Institute of Technology, China*.

Hailong Ma, born in 1983, an engineer at *Shanghai Institute of Spacecraft Equipment, China*.

Sikuan Zhai, born in 1997, is currently a master's degree candidate at *Beijing Institute of Technology, China*.

Ying Wu, born in 1994, received her master's degree from *Beijing Institute of Technology, China* in 2020.

Funding

Competing Interests
The authors declare no competing financial interests.

Author Details
[1] Laboratory of Digital Manufacturing, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China. [2] Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing 314000, China. [3] Shanghai Institute of Spacecraft Equipment, Shanghai 200240, China.

References
[1] L P Liu, F Zhu, J Chen, et al. A quality control method for complex product selective assembly processes. *International Journal of Production Research*, 2013, 51(18): 5437-5449.

[2] C Zhuang, J Gong, J Liu. Digital twin-based assembly data management and process traceability for complex products. *Journal of Manufacturing Systems*, 2021, 58: 118-131.

[3] Y Hong. Data mining for classroom teaching quality based on fuzzy comprehensive evaluation. *Computer Science*, 2008, 35(2): 154-156, 170.

[4] S Zheng. Dynamic quality control in assembly systems. *LIE Transactions*, 2000, 32: 797–806.

[5] E J Tuegel, A R Ingraffea, T G Eason, et al. Reengineering aircraft structural life prediction using a digital twin. *International Journal of Aerospace Engineering*, 2011: 15498.

[6] F Tao, Q Qi, L Wang, et al. Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: correlation and comparison. *Engineering*, 2019, 5(4): 653-661.

[7] C Zhuang, T Miao, J Liu, et al. The connotation of digital twin, and the construction and application method of shop-floor digital twin. *Robotics and Computer-Integrated Manufacturing*, 2021, 68(4): 102075.

[8] C Zhuang, J Liu, H Xiong. Digital twin-based smart production management and control framework for the complex product assembly shop-floor. *International Journal of Advanced Manufacturing Technology*, 2018, 96: 1149-1163.

[9] J W Taylor, P E Mcsharry. Short-term load forecasting methods: an evaluation based on european data. *IEEE Transactions on Power Systems*, 2007, 22(4): 2213-2219.

[10] Z Guo, D Chi, J Wu, et al. A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm. *Energy Conversion and Management*, 2014, 84: 140-151.

[11] J Zhou, J Shi, G Li. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Conversion and Management*, 2011, 52(4): 1990-1998.

[12] D C Li, C C Chang, C W Liu, et al. A new approach for manufacturing forecast problems with insufficient data: the case of TFT-LCDs. *Journal of Intelligent Manufacturing*, 2013, 24(2): 225-233.

[13] D C Li, C W Yeh, C J Chang. An improved grey-based approach for early manufacturing data forecasting. *Computers & Industrial Engineering*, 2009, 57(4): 1161-1167.

[14] C J Chang, J Y Lin, P Jin. A grey modeling procedure based on the data smoothing index for short-term manufacturing demand forecast. *Computational and Mathematical Organization Theory*, 2017, 23(3): 409-422.

[15] X Kou, Q Zhang. The forecast for the wear trend of the diesel engine based on grey Markov chain model. *Lubrication Engineering*, 2007, 1: 288-291.

[16] C Zhuang, J Liu, C Tang, et al. Material dynamic tracking and management technology for discrete assembly process of complex product. *Computer Integrated Manufacturing Systems*, 2015, 21(1): 108-122. (in Chinese)

[17] Y Gao, X Li, X V Wang, et al. A review on recent advances in vision-based defect recognition towards industrial intelligence. *Journal of Manufacturing Systems*, 2022, 62: 753-766.

[18] F Tao, J Cheng, Q Qi, et al. Digital twin-driven product design, manufacturing and service with big data. *International Journal of Advanced Manufacturing Technology*, 2018, 94: 3563-3576.

[19] Y Lu, C Liu, I Kevin, et al. Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues. *Robotics and Computer-Integrated Manufacturing*, 2020, 61: 101837.

[20] D Jones, C Snider, A Nassehi, et al. Characterising the digital twin: a systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 2020, 29: 36-52.

[21] M Liu, S Fang, H Dong, et al. Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*, 2021, 58: 346-361.

[22] F Tao, H Zhang, A Liu, et al. Digital twin in industry: state-of-the-art. *IEEE Transactions on Industrial Informatics*, 2019, 15(4): 2405-2415.

[23] F Caputo, A Greco, M Fera, et al. Digital twins to enhance the integration of ergonomics in the workplace design. *International Journal of Industrial Ergonomics*, 2019, 71: 20-31.

[24] J Leng, Q Liu, S Ye, et al. Digital twin-driven rapid reconfiguration of the automated manufacturing system via an open architecture model. *Robotics and Computer-Integrated Manufacturing*, 2020, 63: 101895.

[25] B R Seshadri, T Krishnamurthy. Structural health management of damaged aircraft structures using the digital twin concept. *25th AIAA/AHS Adaptive Structures Conference*, 2017: 1–13.

[26] F Tao, M Zhang. Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing. *IEEE Access*, 2017, 5: 20418-20427.

[27] J Leng, H Zhang, D Yan, et al. Digital twin-driven manufacturing cyber-physical system for parallel controlling of smart workshop. *Journal of Ambient Intelligence and Humanized Computing*, 2019, 10(3): 1155-1166.

[28] K T Park, J Lee, H J Kim, et al. Digital twin-based cyber physical production system architectural framework for personalized production. *International Journal of Advanced Manufacturing Technology*, 2020, 106: 1787-1810.

[29] Q Bao, G Zhao, Y Yu, et al. The ontology-based modeling and evolution of digital twin for assembly workshop. *International Journal of Advanced Manufacturing Technology*, 2021, 117: 395–411.

[30] Y H Son, K T Park, D Lee, et al. Digital twin–based cyber-physical system for automotive body production lines. *International Journal of Advanced Manufacturing Technology*, 2021, 115: 291–310.

[31] H Zhang, Q Yan, Z Wen. Information modeling for cyber-physical production system based on digital twin and AutomationML. *International Journal of Advanced Manufacturing Technology*, 2020, 107: 1927-1945.

[32] E Yildiz, C Møller, A Bilberg. Demonstration and evaluation of a digital twin-based virtual factory. *International Journal of Advanced Manufacturing Technology*, 2021, 114: 185–203.

[33] X Sun, J Bao, J Li, et al. A digital twin-driven approach for the assembly-commissioning of high precision products. *Robotics and Computer-Integrated Manufacturing*, 2020, 61: 101839.

[34] M Zhang, F Tao, B Huang, et al. A physical model and data-driven hybrid prediction method towards quality assurance for composite components. *CIRP Annals-Manufacturing Technology*, 2021, 70(1): 115-118.