

Article

Nonstationary Process Monitoring Based on Alternating Conditional Expectation and Cointegration Analysis

Jingzhi Rao, Cheng Ji , Jiatao Wen, Jingde Wang * and Wei Sun * 

College of Chemical Engineering, Beijing University of Chemical Technology, Beijing 100029, China

* Correspondence: jingdewang@buct.edu.cn (J.W.); sunwei@buct.edu.cn (W.S.)

Abstract: Traditional multivariate statistical methods, which are often used to monitor stationary processes, are not applicable to nonstationary processes. Cointegration analysis (CA) is considered an effective method to deal with nonstationary variables. If there is a cointegration relationship among the nonstationary series in the system, it indicates that a stable long-term dynamic equilibrium relationship exists among these variables. However, due to the complexity of modern industrial processes, there are nonlinear relations between variables, which are not considered by the traditional linear cointegration theory. Alternating conditional expectation (ACE) can perform nonlinear transformation on these variables to maximize the linear correlation of the transformed variables. It will be helpful to deal with the nonlinear relations by modeling with transformed variables. In this work, a new monitoring strategy based on ACE and CA is proposed. The data are first transformed by an ACE algorithm, CA is performed after that, and then monitoring statistics are calculated to determine whether the system is faulty. The strategy is applied to the monitoring of a simulation case and a catalytic reforming unit in a petrochemical company. The results show that the strategy can realize the monitoring of nonstationary process, with a higher fault detection rate and a lower false alarm rate compared with the monitoring strategy based on traditional cointegration theory.

Keywords: nonlinear transformation; actual industrial process; long-term equilibrium trend



Citation: Rao, J.; Ji, C.; Wen, J.; Wang, J.; Sun, W. Nonstationary Process Monitoring Based on Alternating Conditional Expectation and Cointegration Analysis. *Processes* **2022**, *10*, 2003. <https://doi.org/10.3390/pr10102003>

Academic Editor: Jie Zhang

Received: 31 August 2022

Accepted: 26 September 2022

Published: 4 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the increasing scale and complexity of the modern production industry, the probability of system fault also increases, which may cause economic losses or even major safety incidents. Therefore, it is particularly necessary to monitor the production process. Due to the wide application of distributed control system (DCS), a large number of industrial process data have been recorded [1,2], and data-driven process monitoring methods have developed rapidly with the application of computer technology and artificial intelligence technology in process monitoring. The characteristics of historical data can be extracted by data-driven monitoring methods without detailed modeling of the internal mechanism of the process.

As a traditional data-driven method, multivariate statistical process monitoring (MSPM) has been widely used in the monitoring of stationary processes. The high-dimensional samples are projected into the low-dimensional subspace through MSPM, and the monitoring statistics in the low-dimensional subspace are calculated to monitor the process operation status [3].

Among the methods based on multivariate statistics, principal components analysis (PCA) is one of the classical algorithms. The covariance matrix of the process data set is calculated to obtain the eigenvector of the matrix, so as to determine the direction of the reduced dimension projection [4]. Partial least squares (PLS) is similar to PCA. The basic idea of PLS is to establish a small number of input and output comprehensive variables, so as to reflect the change information contained in the original variables more intensively, and then a linear regression model is established [5]. The square prediction error (SPE)

and Hotelling's statistics (T^2) are calculated to achieve process monitoring in these two methods [6]. However, it is usually assumed in traditional MSPM that the relationship among process variables is linear and variables are stationary. Dynamic and nonstationary characteristics in the process are not taken into account, which refers to the autocorrelation and time-varying characteristics of variables. The abnormal deviations at an early stage of process faults could be buried in these nonstationary trends, and cannot be effectively detected in time [7]. Therefore, several new monitoring methods are proposed for the complex characteristics of the process.

In order to deal with the dynamic characteristics of the process, Ku proposed dynamic principal components analysis (DPCA) on the basis of PCA [8]. The time-lagged variables of the original data were extended to reflect the dynamic relationship among variables. However, Rato [9] pointed out that the principal components extracted by DPCA still retain strong autocorrelation, which result in reduced monitoring performance of T^2 and SPE. A common way to deal with the nonstationary process is to make several differences on the nonstationary variables to obtain the stationary variables and then establish the monitoring model. The autoregressive integrated moving average model (ARIMA) was first proposed by Box and Jenkins [10], whose basic idea is to analyze the characteristics of autocorrelation and partial autocorrelation functions of stationary series after difference, and the parameters of the model are calculated to test the effectiveness of the model to predict the future time series. However, the dynamic information of the process is lost after the difference, which makes the monitoring model less effective.

Cointegration analysis (CA) was first proposed by Engle to deal with nonstationary economic variables [11]. Hendry proposed the error correction model (ECM) in 1978, through which the nonstationary series are converted into stationary series without difference modeling. Granger proposed the relationship between cointegration and ECM in 1981 [12]. In 1987, Engle and Granger integrated the vector autoregressive model (VAR), ECM and cointegration theory to form the Granger representation theorem. Through this theory, the advantages of short-term and long-term models in time series analysis are combined, which provides a better solution for the modeling of nonstationary time series. Due to the internal physical and chemical mechanism of modern industrial process, there is a long-term dynamic equilibrium relationship among variables, which can be handled by cointegration theory. Therefore, CA has been widely used in the industrial field in recent years. Chen applied CA to the industrial field for the first time and introduced the reduced-order model diagnosis method to isolate the system fault. The simulation example in the fluid catalytic cracking unit (FCCU) system showed that cointegration has a good prospect in the application of condition monitoring and fault diagnosis for engineering systems [13].

Xu revealed the shortcomings of the monitoring strategy based on the traditional unit root test method which is insensitive to some system faults and the limitations of the reduced order cointegration model method through some examples, and put forward the method of using the unit root of structural mutation and the Gregory–Hansen cointegration test to carry out system condition monitoring and fault diagnosis. The results showed that the deficiencies of the ADF test were remedied and the variables containing fault information can be directly determined [14]. Yu proposed an adaptive monitoring scheme based on recursive CA to address the issues that when the cointegration relationship changes, the operation status of future nonstationary process could not be reflected accurately by the previous CA. Three monitoring statistics were developed to reflect the operation status of the industrial process, and experimental results of two real industrial processes showed that the adaptive monitoring strategy based on recursive CA could effectively adapt to normal process changes without frequent model updating [15]. A new monitoring index that contains multiple order moments was proposed by Wen [16] to capture different statistical features of the stationary data set. The results showed that the use of multiple order moments as a monitoring index based on cointegration analysis can provide early alarms for abnormal conditions and can effectively identify normal changes and abnormalities.

In simple terms, if a linear combination of a group of nonstationary time series is stationary, it means that linear cointegration exists in the time series. However, the traditional linear cointegration model is not sensitive enough to some faults when the relationship of most industrial process variables is nonlinear, which should be considered when establishing the monitoring model.

Alternating conditional expectation (ACE) was proposed by Breiman [17] and improved by Xue [18] through replacing the conditional expectation calculation of finite data sets with a data smoothing technique called Supermoore. The basic principle of ACE is to transform the dependent variable and the independent variable to maximize the linear correlation between the transformed dependent variable and the independent variable.

Zhang [19] applied it to answer the question that if the sequences Y_t and X_t are nonstationary and non-cointegrated, then under what conditions the cointegration relationship exists between the nonlinear transformation $f(Y_t)$ and $g(X_t)$. He pointed out that it is not necessary to consider whether there is a cointegration relationship between the transformed sequence $f(Y_t)$ and $g(X_t)$ any more if there is a cointegration relationship identified between Y_t and X_t . In such a case, the theory of linear cointegration can perform well. On the contrary, if there is no cointegration relationship between Y_t and X_t , the existing linear cointegration theory will no longer be applicable. In this case, the establishment of the structural model of the transformed sequence $f(Y_t)$ and $g(X_t)$ will broaden the application scope of the cointegration theory.

Based on above discussion, a new monitoring strategy is proposed in this paper. An ACE algorithm is used to transform historical data to maximize the linear correlation between variables, then CA is used to analyze the cointegration relationship among variables and establish the cointegration model, and finally the statistics are calculated to monitor the process. The strategy is applied to the monitoring of a simulation case and a catalytic reforming unit in a petrochemical company. The results show that this method can realize the monitoring of nonstationary process and find the equilibrium relationship which cannot be found by the traditional cointegration method, and it can improve the sensitivity of the monitoring model.

The rest of this paper is organized as follows: the theories and methods used in this paper are introduced in Section 2. The detailed steps of the method proposed in this paper are introduced in Section 3. Three cases are used to verify the effectiveness of the proposed method in Section 4, including two simulation cases and a real industrial case. Finally, the paper is concluded in Section 5.

2. Theory and Method

2.1. Difference and Unit Root Test

If a nonstationary series becomes stationary by differencing it d times, it is called integrated of order d , which can be also represented as: $X \sim I(d)$. When $d = 1$, the difference process is called unit root process. In this paper, we only discuss the case of $d = 1$ as most nonstationary signals can be considered as the 1st order cointegration, if linear cointegration exists. A unit root test can be used to determine whether a time series is consistent with a unit root process.

A popular tool to test whether a time series is stationary is Augmented Dickey–Fuller (ADF) [20]. An autoregressive model can be written as the following form:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + u_t, \quad (1)$$

where p is the lag order, u_t is the random part of a sequence. The characteristic Equation of the above formula is:

$$\lambda^p - a_1 \lambda^{p-1} - \dots - a_p = 0, \quad (2)$$

If the obtained p nonzero eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ are all within the unit circle, the sequence y_t is stationary. Otherwise, there is at least one unit root, assuming $\lambda_1 = 1$, Equation (2) can be written as:

$$1 - a_1 - a_2 - \dots - a_p = 0, \quad (3)$$

This shows that if the series is nonstationary, the sum of the regression coefficients is 1. Equation (1) can be transformed into:

$$y_t = \rho y_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta y_{t-i} + u_t, \quad (4)$$

in which: $\rho = \sum_{j=1}^p \alpha_j$, $\theta_i = -\sum_{j=i+1}^p \alpha_j$, ($i = 1, 2, \dots, p-1$). When $\rho < 0$, the series is stationary.

Parameters in Equation (4) are estimated by ordinary least squares and ρ is tested to determine whether the series is stationary.

2.2. Cointegration Theory

The basic idea of cointegration theory is that although a set of variables are nonstationary over time, they will change together and maintain a common long-term random trend. The random trend can be eliminated by linear combination, and a stationary series can be obtained. It can be considered that there is a long-term equilibrium relationship between them.

On the concept of simple integration, the cointegration theory can be further explained. $x_{1t}, x_{2t}, x_{3t}, \dots, x_{it}$ are nonstationary series, where $x_{it} \sim I(1)$, $i = 1, 2, 3, \dots, i$. These series meet the cointegration relationship if the following equation can be established by a set of coefficients, $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_i$:

$$\xi = \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \dots + \alpha_i x_{it}, \quad (5)$$

where ξ is a stationary series of $I(0)$, and α_i is called the cointegration coefficient. The equation above can be explained as follows: the linear combination of a group of $I(1)$ variables can derive an $I(0)$ variable, then these variables are cointegrated.

2.3. Cointegration Test

2.3.1. Engle and Granger Test

In order to test whether there is a cointegration relationship between two sequences Y_t and X_t , Engle and Granger proposed a two-step test, also known as the EG test, in 1987.

Step 1: estimate the following equation with ordinary least squares regression:

$$Y_t = c + \beta X_t + \mu_t, \quad (6)$$

in which c is the constant term, β is the regression coefficient, and μ_t is the residual sequence, which can be calculated as follows:

$$\hat{\mu}_t = Y_t - \hat{Y}_t = Y_t - \hat{\alpha} - \hat{\beta} X_t, \quad (7)$$

Step 2: carry out the ADF test mentioned above on $\hat{\mu}_t$, if the residual sequence is stationary, Y_t and X_t are cointegrated.

The EG test is mainly applicable to the cointegration test between two variables, and not applicable when there are multiple cointegration relationships between variables.

2.3.2. Johansen Test

The Johansen test [21,22] based on the VAR model is a method which is used to determine the cointegrated relationship among multiple nonstationary variables. The VAR model can be described as follows:

$$Y_t = \sum_{j=1}^p A_j Y_{t-j} + D_t + U_t, \quad (8)$$

where $Y_t \in R^m$ is a m-dimensional series, $A_j \in R^{m \times m}$ is the coefficient matrix, and U_t is an m-dimensional white noise that conforms to normal distribution.

The vector error-correction (VEC) model transformed from Equation (8) can be expressed as:

$$\Delta Y_t = -\Pi Y_{t-1} + \sum_{j=1}^{p-1} A_j^* \Delta Y_{t-j} + D_t + U_t, \tag{9}$$

$$\Pi = A(1) = I - \sum_{j=1}^p A_j, \tag{10}$$

$$A_j^* = -\sum_{i=j+1}^p A_i, j = 1, 2, \dots, p - 1, \tag{11}$$

Π can be estimated by the maximum likelihood estimate (MLE). Depending on its rank, if $\text{rank}(\Pi) = 0$: there is no cointegrated relationship between the series. The number of cointegrated relationship can be calculated by testing the significance of the characteristic roots of Π . The characteristic roots of Π are sorted in descending order, $\lambda_1 > \lambda_2 > \dots > \lambda_m$. If there are r cointegrated vectors, the other characteristic roots should be zero. The number of cointegration relationships can be obtained by a trace test of characteristic roots, of which statistics can be calculated by:

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^m \ln(1 - \hat{\lambda}_i), \tag{12}$$

The critical value of the statistic can be obtained by the Monte Carlo method, then the number of independent cointegration vectors can be determined. Π is decomposed into two full rank matrices:

$$\Pi = \Gamma B', \tag{13}$$

where $\Gamma, B \in R^{m \times r}, 0 < r < m$, then:

$$\zeta_t = B'^{x_t}, \tag{14}$$

in which ζ_t is a stationary series, and B' is the cointegration coefficient matrix.

2.4. Alternating Conditional Expectation

A method is provided by the ACE algorithm to derive a pair of nonlinear transformation functions $f(\cdot)$ and $g(\cdot)$ to maximize the correlation between $f(Y_t)$ and $g(X_t)$. If $f(Y_t)$ is regressed to $g(X_t)$ to calculate goodness of fit R^2 , the ACE algorithm is equivalent to selecting nonlinear transformation function to maximize goodness of fit R^2 , which makes Equation (15) a significant regression relation expression.

$$f(Y_t) = g(X_t) + \mu_t, \tag{15}$$

where μ_t is the random errors. $f(\cdot)$ and $g(\cdot)$ are used to minimize the sum of squares of residuals in the above regression relationship, that is:

$$(f, g) = \arg \min_{f, g} \sum_t^T [f(Y_t) - g(X_t)]^2, \tag{16}$$

The process of finding the optimal transformation function by the ACE algorithm is an iterative process. The mean square error of the simple linear regression process is defined:

$$e^2(f, g) = E[f(Y_t) - g(X_t)]^2, \tag{17}$$

the $e^2(f, g)$ for $f(Y_t)$ are minimized and if $E[f^2] = 1$, the result is:

$$f_1(Y) = \frac{E[g(X_t)|Y_t]}{\|E[g(X_t)|Y_t]\|} \tag{18}$$

where $\|\cdot\| = [E(\cdot)^2]^{1/2}$, then the next step of ACE algorithm is minimizing the $e^2(f_1, g)$ for $g(X_t)$. The result is:

$$g_1(X_t) = E[f_1(Y_t)|X_t], \quad (19)$$

The process to calculate nonlinear transformation functions $(f_1(\cdot), g_1(\cdot))$ is the first iteration in the whole ACE calculation. The iterative process will continue until the value of $e^2(f, g)$ is no longer reduced. The iterative basis is defined as follows:

$$f_m(Y_t) = \frac{E[g_{m-1}(X_t)|f_{m-1}(Y_t)]}{\|E[g_{m-1}(X_t)|f_{m-1}(Y_t)]\|}, \quad (20)$$

$$g_m(X_t) = E[f_m(Y_t)|g_{m-1}(X_t)], \quad (21)$$

where the initial values are $f_0(Y_t) = Y_t/\|Y_t\|, g_0(X_t) = X_t$.

2.5. Monitoring Statistics

A stationary multivariate series can be combined by cointegration test:

$$\zeta_t = \hat{\beta}x_t + \hat{\mu}, \quad (22)$$

where ζ_t is a stationary multivariate series, so it can be monitored by traditional multivariate statistical methods. When the process becomes abnormal, the long-term equilibrium relationship among variables will be broken, and the statistical characteristics of ζ_t will also change.

The monitoring of process operation status can be realized through monitoring the statistical characteristics of ζ_t . The original nonstationary variables are projected into a new space through the cointegration vector, and the common random trend among variables is eliminated. The stochastic trends in residual subspace can be extracted as:

$$\tau_t = \beta^{\perp T} x_t, \quad (23)$$

τ_t is differenced to eliminate its nonstationary trend.

$$\Delta\tau_t = \beta^{\perp T} (x_t - x_{t-1}), \quad (24)$$

T_τ^2 statistic is constructed as:

$$T_\tau^2 = \Delta\tau^T \Lambda_z^{-1} \Delta\tau, \quad (25)$$

T_τ^2 is used to monitor the change of nonstationary part [23], which is similar to T^2 statistic in PCA. The control limit C can be obtained by kernel density estimation.

3. The Proposed Monitoring Strategy

Nonstationary working conditions often exist in actual industrial production. The nonstationary trend of variables cannot be distinguished from the trend caused by an abnormal process through the traditional multivariate statistical method. The change of a long-term equilibrium relationship among nonstationary variables can be monitored instead of the nonstationary variables themselves by CA. However, the traditional linear cointegration model is not sensitive to some faults if there is a nonlinear relationship among certain process variables. The purpose of the ACE algorithm is to find a pair of nonlinear transformation functions that maximize the linear correlation of the transformed sequences. Therefore, establishing the structural model of the transformed sequence will broaden the application range of the cointegration theory.

The monitoring strategy based on ACE and CA is proposed in this article. The algorithm block diagram is as shown in Figure 1.

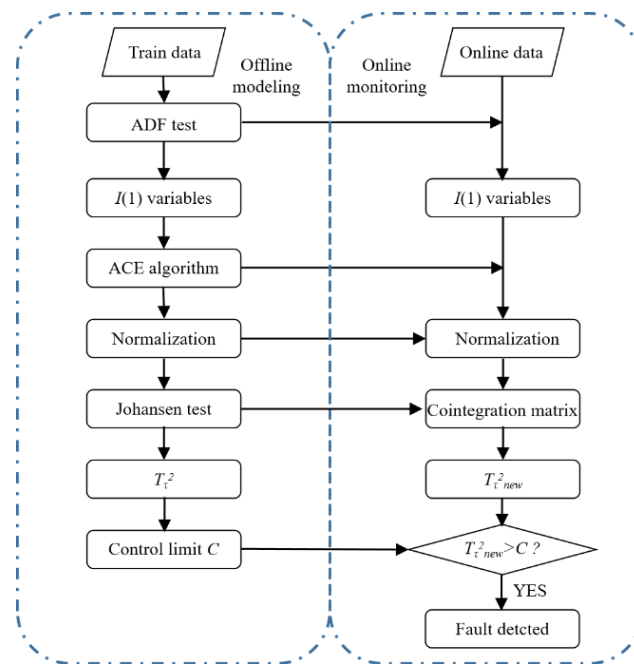


Figure 1. Monitoring strategy based on ACE and CA.

3.1. Offline Modeling

Step 1: Perform ADF test on the training data to distinguish the nonstationary variables, $I(1)$ variables will be selected as the modeling variables.

Step 2: Select one of the $I(1)$ variables as target variable Y_t , which is a general variable prone to abnormal changes, other variables as X_t . A group of transform data $X_{t-trans}$ and $Y_{t-trans}$ are calculated through ACE. Then Y_t is replaced with the average of $Y_{t-trans}$. The original variables are polynomial fitted to the transformed variables to obtain the nonlinear transformation equation.

Step 3: Normalize the transformed variables with the following formula:

$$x^* = \frac{x - \bar{x}}{s}, \quad (26)$$

where \bar{x} is the average of $X_{t-trans}$, s is the standard deviation. The number of cointegration vectors r is obtained through Johansen test, and the cointegration coefficient matrix $B = (\beta_1, \beta_1, \dots, \beta_r)$ is obtained from maximum likelihood estimation.

Step 4: Construct the monitoring statistics T_τ^2 and control limit C so that online data can be monitored.

3.2. Online Monitoring

Step 1: Select the $I(1)$ variables determined by the training data as the model input variables.

Step 2: Transform variables by the nonlinear transformation equations in step 2 above.

Step 3: Normalize the transformed variables by the process variables calculated in step 3 above. Project the transformed variables onto the cointegration coefficient matrix obtained in step 3 above:

$$\zeta_{new} = BX_{t-new}, \quad (27)$$

Step 4: Construct monitoring statistics $T_{\tau-new}^2$. When the monitoring statistics exceed the control limit C , the system will trigger an alarm.

4. Case and Result

4.1. Two-Dimensional Simulation Case

4.1.1. Data Construction

In order to verify whether the two sets of sequences $f(Y_t)$ and $g(X_t)$ transformed by the ACE algorithm are cointegrated, the following nonstationary variables X and Y are constructed.

$$a_t = a_{t-1} + e_{1t}, \quad (28)$$

$$x_t = -0.01a_t^2 + e_{2t}, \quad (29)$$

$$y_t = a_t + e_{3t} + 3, \quad (30)$$

where $a_0 = 0$, $e_{1t} \sim N(0, 1)$, $e_{2t}, e_{3t} \sim N(0, 0.5)$. The total number of samples is 1000, and X and Y are shown in the Figure 2:

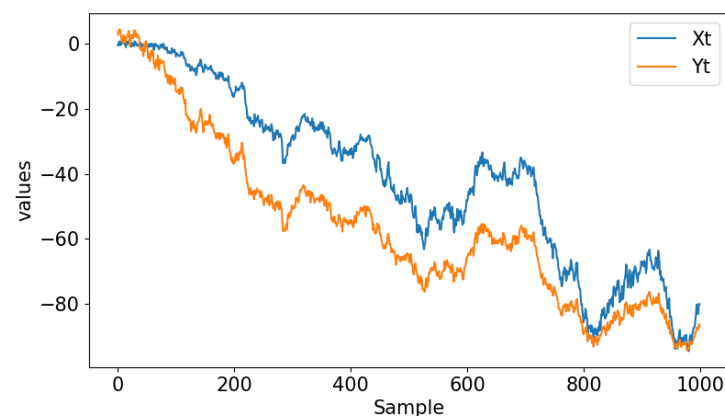


Figure 2. Two-dimensional simulation data.

4.1.2. ACE and EG Test

Least square regression is performed between X_t and Y_t , the following equation is obtained:

$$Y_t = -17.7144 + 0.9250X_t + \mu_t, \quad (31)$$

μ_t is shown in Figure 3:

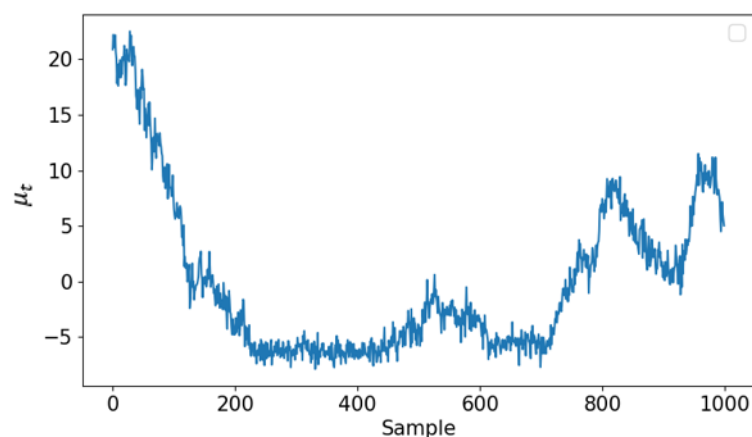


Figure 3. Residual sequence of raw data.

The ADF test shows that the series is nonstationary. The result of the EG test indicates that there is no cointegration relationship between X_t and Y_t . The constructional data after ACE is shown in Figure 4:

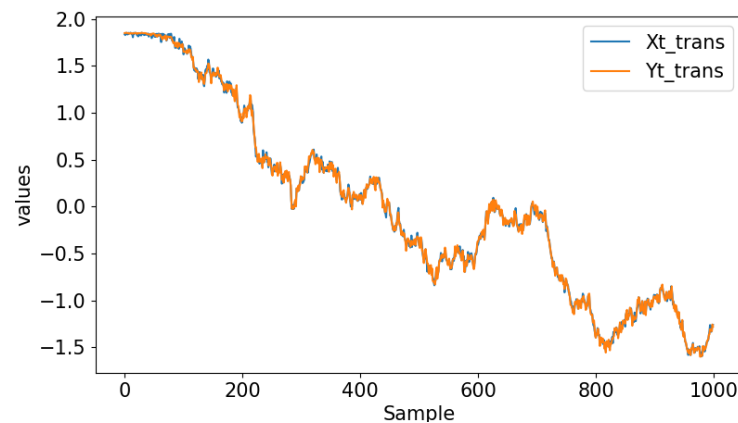


Figure 4. Two-dimensional transformed data.

Least square regression is performed between $X_{t-trans}$ and $Y_{t-trans}$, the following equation is obtained:

$$Y_{t-trans} = 1.4844 \times 10^{-11} + 1.00024X_{t-trans} + \mu_t, \quad (32)$$

μ_t is shown in Figure 5.

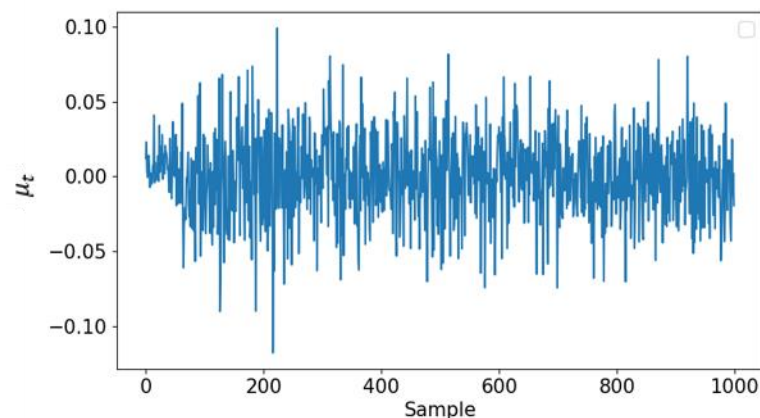


Figure 5. Residual sequence of transformed data.

The ADF test shows that the series is nonstationary. The result of the EG test indicates that the transformed data are cointegrated. It is proved that the ACE algorithm can transform two non-cointegration variables into cointegration variables.

4.2. Multidimensional Simulation Case

4.2.1. Data Construction

To verify the effectiveness of the proposed monitoring method, a set of simulation variables is constructed as follows:

$$x_t = x_{t-1} + e_{1t}, \quad (33)$$

$$y_{0t} = -0.02x_t^2 + e_{2t}, \quad (34)$$

$$y_{1t} = x_t + e_{3t}, \quad (35)$$

$$y_{2t} = -0.01x_t^2 + e_{4t} - 10, \quad (36)$$

$$y_{3t} = 0.01x_t^2 + e_{5t} + 4, \quad (37)$$

$$y_{4t} = 0.03x_t^2 - e_{6t} - 20, \quad (38)$$

where $e_{1t} \sim N(0, 1)$, $e_{2t}, e_{3t}, e_{4t}, e_{5t}, e_{6t} \sim N(0, 0.5)$, $x_0 = 0$. The total number of samples is 1600 and the training data sample are shown in Figure 6:

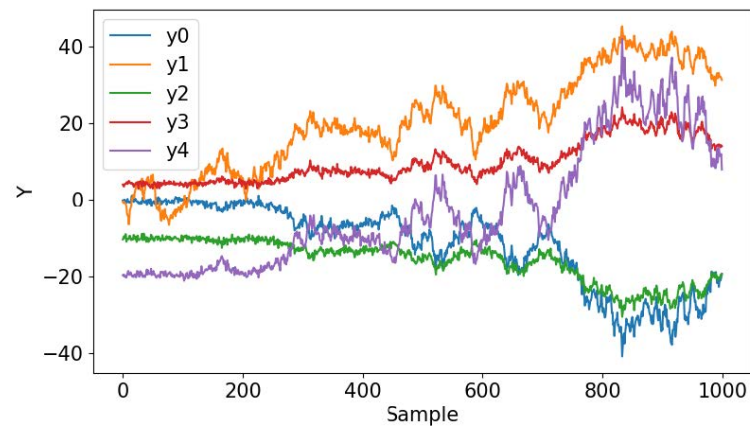


Figure 6. Training data sample.

Figure 6 shows that trends of these variables are roughly the same. At the 300th sample of the test data, a fault is introduced to y_1 to break the long-term equilibrium relationship between variables. After 300 samples, y_1 is constructed as follows:

$$y_{1t} = y_{1t-1} + e_t, \quad (39)$$

where $e_{1t} \sim N(0, 0.5)$, y_1 changes as shown in the Figure 7, where the solid line indicates normal data and the dotted line indicates fault. After the fault is introduced, the long-term equilibrium trend no longer exists between y_1 and other variables.

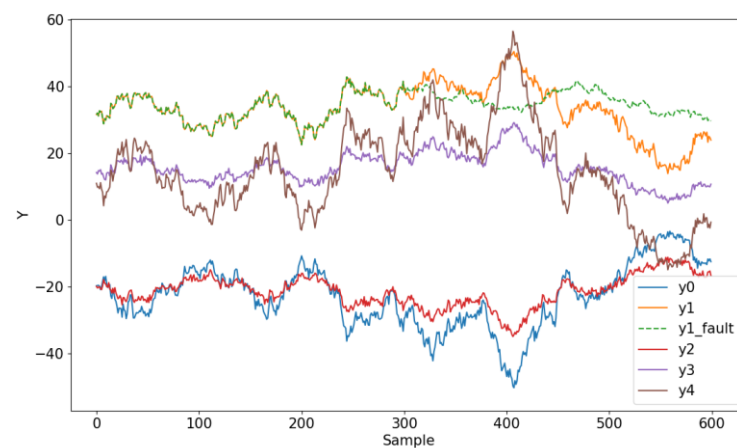


Figure 7. Test and fault data sample.

4.2.2. ACE and Johansen Test

Transformed data by ACE are shown in Figure 8:

The following quartic polynomial fitting on the transformed data is performed to obtain the nonlinear transformation function, and the results are shown in Table 1.

$$f(x) = Ax^4 + Bx^3 + Cx^2 + Dx + E, \quad (40)$$

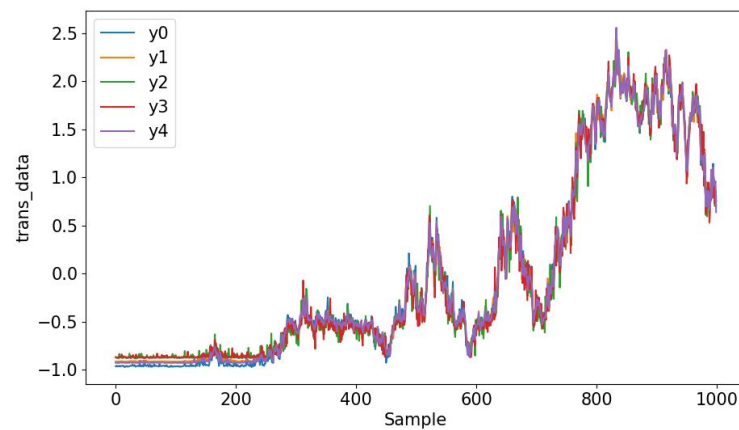


Figure 8. Transformed train data by ACE.

Table 1. Polynomial fitting coefficient table.

	y_0	y_1	y_2	y_3	y_4
A	3.803×10^{-7}	-2.346×10^{-6}	7.105×10^{-6}	-5.025×10^{-7}	-2.341×10^{-8}
B	9.533×10^{-5}	1.788×10^{-4}	1.308×10^{-3}	-7.445×10^{-4}	-2.182×10^{-5}
C	4.193×10^{-3}	-1.713×10^{-3}	5.973×10^{-2}	3.087×10^{-2}	5.489×10^{-4}
D	-4.106×10^{-2}	2.945×10^{-3}	8.172×10^{-1}	-1.893×10^{-1}	6.948×10^{-2}
E	-9.707×10^{-1}	-8.798×10^{-1}	2.565	-5.595×10^{-1}	6.563×10^{-2}

The test data are transformed through the equation above.

In Figure 9, it can be observed that the change trend of the training and test data is the same in addition to the fault data, so that the Johansen test can perform well. The Akaike information criterion is used to determine the lag order of the VAR model, which is calculated as 2. Table 2 shows the results:

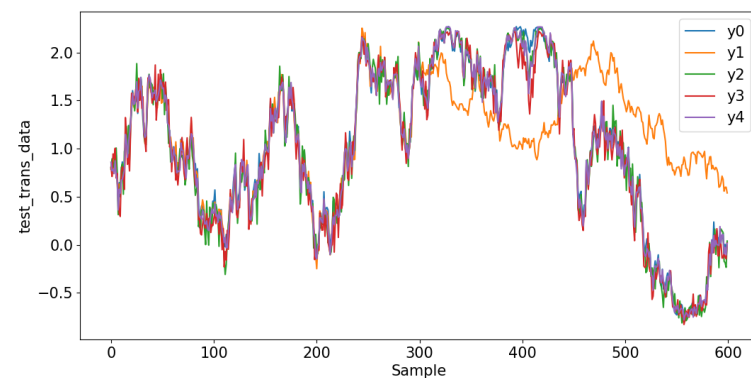


Figure 9. Transformed test data through fitting curves.

Table 2. Johansen test results with ACE.

Johansen H0 Hypothesis	Trace Statistics	Critical Value (5%)
$r \leq 0$	549.353	3.8415
$r \leq 1$	286.844	15.4943
$r \leq 2$	65.3606	29.7961
$r \leq 3$	1.85367	47.8545
$r \leq 4$	0	69.8189

Table 2 shows the results of the Johansen test with ACE, while the results of Table 3 are obtained without ACE transformed. At the 5% significance level, the assumption that

$r \leq 3$ is acceptable in Table 1. The number of cointegration relationship is 3, that is, there are three forms of linear combination that can eliminate the nonstationary trend among these variables.

Table 3. Johansen test results without ACE.

Johansen H0 Hypothesis	Trace Statistics	Critical Value (5%)
$r \leq 0$	566.01	3.8415
$r \leq 1$	276.362	15.4943
$r \leq 2$	11.0395	29.7961
$r \leq 3$	2.89588	47.8545
$r \leq 4$	0	69.8189

However, in Table 3, the assumption that $r \leq 2$ is acceptable. It shows that there are more cointegration relationships among variables after ACE algorithm transform.

4.2.3. Monitoring Results

The monitoring results based on common cointegration and the method proposed in this paper are shown in Figures 10 and 11:

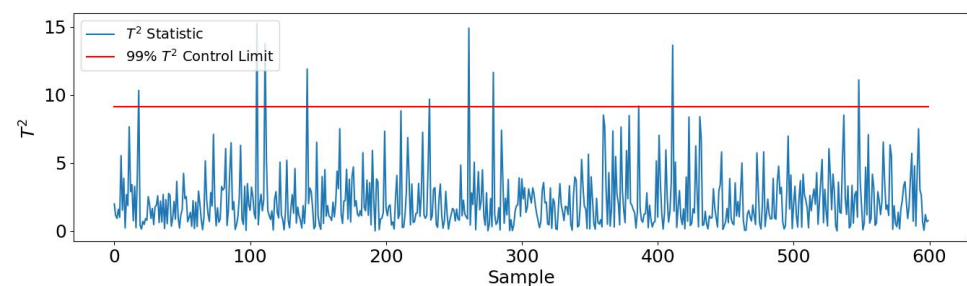


Figure 10. Monitoring results of common cointegration analysis.

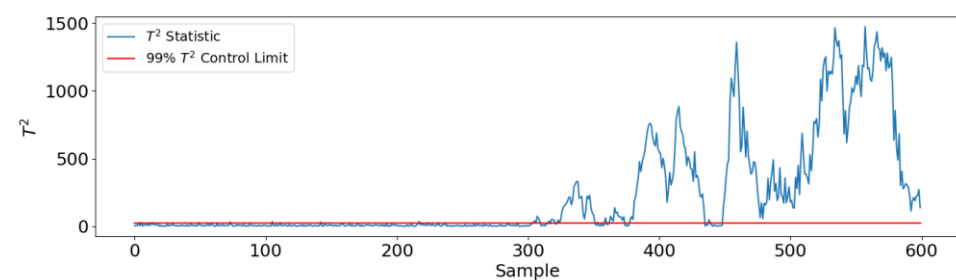


Figure 11. Monitoring results of cointegration analysis with ACE.

The blue solid line represents the monitoring statistics, and the red horizontal solid line represents the monitoring control limit. If the statistics exceed the control line, it is considered to be in an abnormal state. It can be observed from Figures 10 and 11 that the monitoring method based on common cointegration analysis will not trigger an alarm when a fault occurs, while the method proposed in this paper can trigger an alarm in time, which shows that the monitoring strategy proposed in this paper is more sensitive to this type of fault.

4.3. Industrial Case

4.3.1. Introduction

The pressure drop at the hot end of the heat exchanger in the catalytic reforming process unit of a petrochemical company often rises abnormally, which may lead to safety incidents if it is not handled quickly. Therefore, it is particularly necessary to monitor the process.

A section of historical data with an abnormal rise of pressure drop is selected, including 2000 sample points. The sampling frequency is 1 min, and each sample consists of 27 variables.

The pressure drop at the hot end of the heat exchanger is mainly affected by the feed rate of naphtha and circulating hydrogen. Figure 12 shows the changes of pressure drop at the hot end of the heat exchanger, naphtha feed rate and circulating hydrogen feed rate with time from top to bottom. The naphtha feed rate has been stable, while the change trend of the pressure drop at the hot end of the heat exchanger is basically the same as that of the circulating hydrogen feed. Near about the 1400th sample points (the red frame), that is, the 500th sample point in the test set, the pressure drop at the hot end of the heat exchanger shows an upward trend, while the feed rate of circulating hydrogen and naphtha changes steadily, indicating that an abnormal increase in the pressure drop at the hot end of the heat exchanger occurred near this point.

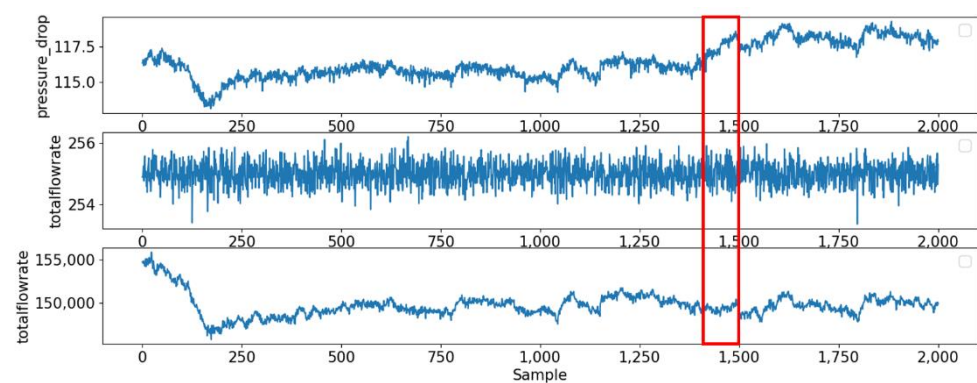


Figure 12. Variation trend of pressure drop at hot end.

The first 900 samples are used as the training data to establish the model, and the remaining samples are used as the test set to verify the effectiveness of the method.

4.3.2. Modeling

In total, 27 variables in the training data are tested by ADF, and I (1) variables are selected to establish the monitoring model.

If the test statistic of ADF is greater than the critical value under 1% significance, the variable can be considered as nonstationary. These nonstationary variables are stationary after difference of the first order. Finally, eight variables are determined as modeling variables. The test statistic and critical value are showed in Table 4.

Table 4. Unit root test results of training data.

Variable	Test Statistic	Critical Value (1%)	Variable	Test Statistic	Critical Value (1%)
Inlet flow of cold end	−30.5959	−3.43764	Inlet temperature of hot end	−4.4610	−3.43765
Hydrogen flow rate	−2.8303	−3.43766	Inlet temperature of cold end	−5.0377	−3.43765
Inlet pressure of cold end	−3.3003	−3.43766	Outlet temperature of cold end	−4.0436	−3.43766
Hydrogen pressure	−2.8968	−3.43766	Outlet temperature of the first furnace	−15.7487	−3.43766
Outlet pressure of hot end	−11.3220	−3.43766	Outlet temperature of the second furnace	−19.9573	−3.43765
Pressure drops of hot end	−3.1284	−3.43766	Outlet temperature of the first reactor	−3.3998	−3.43766
Pressure drops of cold end	−3.4391	−3.43766	Outlet temperature of the second reactor	−3.1017	−3.43766
Pressure drops of the first reactor	−3.6682	−3.43766	Outlet temperature of the third furnace	−19.2434	−3.43766

Table 4. Cont.

Variable	Test Statistic	Critical Value (1%)	Variable	Test Statistic	Critical Value (1%)
Pressure drops of the second reactor	-2.8726	-3.43766	Outlet temperature of the third reactor	-4.2464	-3.43766
Pressure drops of the third reactor	-3.5858	-3.43766	Outlet temperature of the fourth furnace	-18.8509	-3.43766
Pressure drops of the fourth reactor	-3.8137	-3.43766	Temperature drops of the third furnace	-9.2383	-3.43766
Inlet pressure of the fourth reactor	-3.8399	-3.43766	Temperature drops of the fourth furnace	-7.7637	-3.43766
Pressure drops of cold end filter	-12.1079	-3.43766	Temperature drops of the second furnace	-4.3286	-3.43766
Outlet temperature of hot end	-2.8410	-3.43766			

The relationship of some variables in training data transformed by the ACE algorithm is shown in Figure 13, where the abscissa represents the original values and the ordinate represents the transformed values.

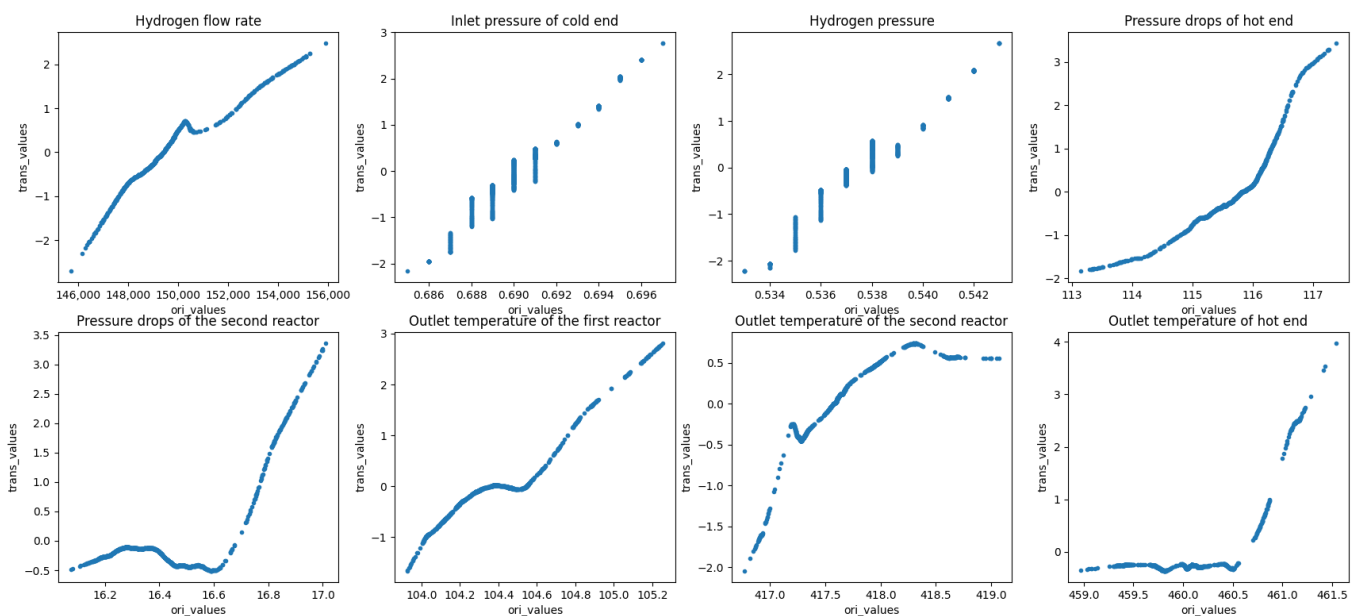


Figure 13. Ace transformation relation of the selected variables.

Then the Johansen test is performed on these variables. The results of the Johansen test and the Johansen test with the ACE algorithm are shown in Tables 5 and 6.

Table 5. Johansen test results for industrial case.

Johansen H0 Hypothesis	Trace Statistics	Critical Value (5%)
$r \leq 0$	419.515	3.8415
$r \leq 1$	215.018	15.4943
$r \leq 2$	121.076	29.7961
$r \leq 3$	50.9084	47.8545
$r \leq 4$	26.0231	69.8189
$r \leq 5$	9.88992	95.7542
$r \leq 6$	2.44853	125.618
$r \leq 7$	0	159.529

Table 6. Johansen test with ACE results for industrial case.

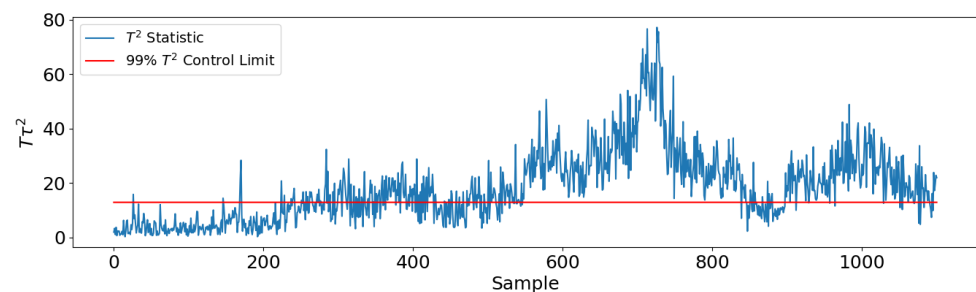
Johansen H0 Hypothesis	Trace Statistics	Critical Value (5%)
$r \leq 0$	419.486	3.8415
$r \leq 1$	250.82	15.4943
$r \leq 2$	161.232	29.7961
$r \leq 3$	96.6288	47.8545
$r \leq 4$	54.8984	69.8189
$r \leq 5$	20.8353	95.7542
$r \leq 6$	6.35215	125.618
$r \leq 7$	−0	159.529

At the 5% significance level, the assumption that $r \leq 4$ is acceptable while $r \leq 3$ is rejected. Thus, there are four cointegration relationships among these variables. Although both methods have tested four cointegration relationships, the trace statistics of the Johansen test processed by the ACE algorithm are larger than that of the ordinary Johansen test, which means that the ACE algorithm still has the possibility to test more cointegration relationships.

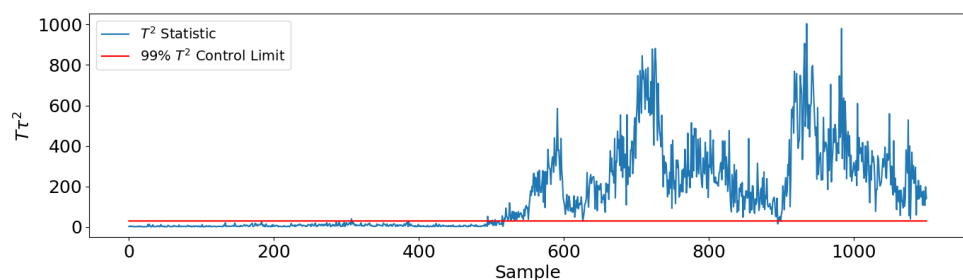
4.3.3. Monitoring Results

The control limit C is calculated with training set data, and T_T^2 statistics are calculated for test data. The monitoring results of the two methods are as follows:

It can be observed from Figure 14 that the T_T^2 statistics of the monitoring method based on conventional cointegration analysis have exceeded the control limit before the abnormal rise of the pressure drop at the hot end of the heat exchanger, and there are a large number of false alarms, which indicates that inherent features of process data are not well captured by the traditional cointegration method.

**Figure 14.** Monitoring results of industrial case based on cointegration analysis.

In Figure 15, the system triggers an alarm at approximately the 517th sample point, when the monitoring statistics exceed the control limit, which shows that the monitoring method based on cointegration analysis with ACE also shows a good monitoring performance in practical industrial cases, which also verify that nonlinear cointegration exists in the considered process.

**Figure 15.** Monitoring results of industrial case based on cointegration analysis with ACE.

5. Conclusions

In this work, a nonstationary process monitoring strategy based on CA and an ACE algorithm is proposed. Through traditional CA, only the linear cointegration relationship between variables can be extracted. As a nonparametric method, the ACE algorithm only depends on the extremely weak distribution assumption, and a variety of nonlinear transformation forms of data can be obtained, so that the nonlinear characteristics of different forms of variables can be described [24]. Thus, the nonlinear cointegration relationship can be extracted by CA combined with ACE.

Aiming at nonlinear and nonstationary industrial data, nonlinear transformation derived by ACE is first performed on non-cointegration series. These transformations converge gradually to an optimal transformation obtained through the nonparametric data smoothing technique, namely the optimal ACE transformation, which is similar to robust optimization [25,26]. If there is a certain long-term nonlinear relationship between these series, this long-term equilibrium relationship among these transformed series could be extracted by traditional CA, which means that the transformed series become cointegrated and the nonlinear and nonstationary data characteristics can be extracted.

The strategy proposed is also applied in multi-dimensional simulation data and industrial data. The results show that the strategy can trigger an alarm in time when the fault occurs, while the traditional monitoring strategy based on cointegration theory results in a large number of false alarms, which means that the strategy has a wider application range and higher sensitivity.

Author Contributions: Conceptualization, J.R. and W.S.; methodology, J.R. and J.W. (Jiatao Wen); software, J.R. and J.W. (Jiatao Wen); validation, J.R.; formal analysis, J.R.; investigation, J.R.; resources, J.R., J.W. (Jingde Wang) and W.S.; data curation, J.R. and J.W. (Jingde Wang); writing—original draft preparation, J.R.; writing—review and editing, J.R., C.J. and W.S.; visualization, J.R.; supervision, W.S. and J.W. (Jingde Wang); project administration, W.S. and J.W. (Jingde Wang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 22278018).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ge, Z.; Song, Z.; Gao, F. Review of Recent Research on Data-Based Process Monitoring. *Ind. Eng. Chem. Res.* **2013**, *52*, 3543–3562. [[CrossRef](#)]
2. Ji, C.; Sun, W. A Review on Data-Driven Process Monitoring Methods: Characterization and Mining of Industrial Data. *Processes* **2022**, *10*, 335. [[CrossRef](#)]
3. Julieta, C.; Daniel, S.; Barbara, G. Monitoring Wine Fermentation Deviations Using An ATR-MIR Spectrometer and MSPC Charts. *Chemom. Intell. Lab. Syst.* **2020**, *201*, 104011.
4. Tong, C.; Shi, X. Mutual Information Based PCA Algorithm with Application in Process Monitoring. *CIESC J.* **2015**, *10*, 6.
5. Li, Z.; Liang, L.; Han, C. Multi-Rate Process Fault Detection Based on Partial Least Squares. *Comput. Simul.* **2016**, *10*, 5.
6. Pollanen, K.; Hakkinen, A.; Reinikainen, S. Dynamic PCA-Based MSPC Charts for Nucleation Prediction in Batch Cooling Crystallization Processes. *Chemom. Intell. Lab. Syst.* **2016**, *84*, 126–133. [[CrossRef](#)]
7. Ji, C.; Ma, F.; Wang, J.; Sun, W. Early Identification of Abnormal Deviations in Nonstationary Processes by Removing Non-Stationarity. *Comput. Aided Chem. Eng.* **2022**, *49*, 1393–1398.
8. Ku, W.; Storer, R.H.; Georgakis, C. Disturbance Detection and Isolation by Dynamic Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 179–196. [[CrossRef](#)]
9. Rato, T.J.; Reis, M.S. Advantage of Using Decorrelated Residuals in Dynamic Principal Component Analysis for Monitoring Large-Scale Systems. *Ind. Eng. Chem. Res.* **2013**, *52*, 13685–13698. [[CrossRef](#)]
10. Box, G.; Jenkins, G.M. *Time Series Analysis, Forecasting, and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
11. Engle, R.F.; Granger, C. Cointegration and Error-Correction: Representation, Estimation and Testing. *Econometrica* **1987**, *55*, 251–276. [[CrossRef](#)]
12. Granger, C. Some Properties of Time Series Data and Their Use in Econometric Model Specification. *J. Econom.* **1981**, *16*, 121–130. [[CrossRef](#)]

13. Chen, Q.; Pan, Y. Application of Cointegration Testing Method to Condition Monitoring and Fault Diagnosis of Nonstationary FCCU System. *Acta Pet. Sin.* **2007**, *23*, 69–76.
14. Xu, Z. Nonstationary Process Monitoring and Fault Diagnosis Using Cointegration with Structural Change Testing Method. Master's Thesis, Nanjing University of Aeronautics and Astronautics, Nanjing, China, December 2007.
15. Yu, W.; Zhao, C.; Huang, B. Recursive Cointegration Analytics for Adaptive Monitoring of Nonstationary Industrial Processes with both Static and Dynamic Variations. *J. Process Control* **2020**, *92*, 319–332. [[CrossRef](#)]
16. Wen, J.; Li, Y. Nonstationary Process Monitoring Based on Cointegration Theory and Multiple Order Moments. *Processes* **2022**, *10*, 169. [[CrossRef](#)]
17. Breiman, L.; Friedman, J.H. Estimating Optimal Transformations for Multiple Regression and Correlation. *Publ. Am. Stat. Assoc.* **1985**, *80*, 580–598. [[CrossRef](#)]
18. Xue, G. Optimal Transformations for Multiple Regression: Application to Permeability Estimation from Well Logs. *Spe Form. Eval.* **1997**, *12*, 85–94. [[CrossRef](#)]
19. Zhang, X.; Zhang, S. A Reconsideration to the Nonlinear Transformation of the Integrated Time Series. *J. Syst. Eng.* **1998**, *13*, 8.
20. Lu, F. Complex Dynamic Engineering System Condition Monitoring Research Using Cointegration Theory. Ph.D. Thesis, Nanjing University of Aeronautics and Astronautics, Nanjing, China, November 2010.
21. Johansen, S. Statistical Analysis of Cointegration Vectors. *J. Econ. Dyn. Control* **1988**, *12*, 231–254. [[CrossRef](#)]
22. Johansen, S. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* **1991**, *59*, 1551–1580. [[CrossRef](#)]
23. Li, G.; Qin, S.J.; Yuan, T. Nonstationarity and Cointegration Tests for Fault Detection of Dynamic Processes. *IFAC Proc. Vol.* **2014**, *47*, 10616–10621. [[CrossRef](#)]
24. Zhu, H.; Li, S.; Zeng, H. Test for Bayesian Nonlinear Cointegration in Nonparametric ACE Transformed Model. *J. Manag. Sci. China* **2011**, *14*, 52–64.
25. Özmen, A.; Weber, G.W.; Batmaz, İ. RCMARS: Robustification of CMARS with Different Scenarios under Polyhedral Uncertainty Set. *Commun. Nonlinear Sci. Numer. Simul.* **2011**, *16*, 4780–4787. [[CrossRef](#)]
26. Özmen, A.; Weber, G.W.; Batmaz, İ. The New Robust CMARS (RCMARS) Method. *Vectors* **2010**, *1*, 362–368.