processes

MDPI

Article

Characteristic Canonical Analysis-Based Attack Detection of Industrial Control Systems in the Geological Drilling Process

Mingdi Xu [1], Zhaoyang Jin [1], Shengjie Ye [1] and Haipeng Fan [2,*]

[1] Wuhan Institute of Digital Engineering, Wuhan 430074, China; mingdixu@163.com (M.X.); jinzhaoyang37@163.com (Z.J.); shengjieily@hust.edu.cn (S.Y.)
[2] School of Automation, China University of Geosciences, Wuhan 430074, China
* Correspondence: fanhaipeng@cug.edu.cn

Abstract: Modern industrial control systems (ICSs), which consist of sensor nodes, actuators, and buses, contribute significantly to the enhancement of production efficiency. Massive node arrangements, security vulnerabilities, and complex operating status characterize ICSs, which lead to a threat to the industrial processes' stability. In this work, a condition-monitoring method for ICSs based on canonical variate analysis with probabilistic principal component analysis is proposed. This method considers the essential information of the operating data. Firstly, the one-way analysis of variance method is utilized to select the major variables that affect the operating performance. Then, a concurrent monitoring model based on probabilistic principal component analysis is established on both the serially correlated canonical subspace and its residual subspace, which is divided by canonical variate analysis. After that, monitoring statistics and control limits are constructed. Finally, the effectiveness and superiority of the proposed method are validated through comparisons with actual drilling operations. The method has better sensitivity than traditional monitoring methods. The experimental result reveals that the proposed method can effectively monitor the operating performance in a drilling process with its highest accuracy of 92.31% and a minimum monitoring delay of 11 s. The proposed method achieves much better effectiveness through real-world process scenarios due to its distributed structural division and the characteristic canonical analysis conducted in this paper.

Keywords: industrial control systems; performance monitoring; canonical variate analysis; principal component analysis

Citation: Xu, M.; Jin, Z.; Ye, S.; Fan, H. Characteristic Canonical Analysis-Based Attack Detection of Industrial Control Systems in the Geological Drilling Process. *Processes* **2024**, *12*, 2053. https://doi.org/10.3390/pr12092053

Academic Editor: Iqbal M. Mujtaba

Received: 30 July 2024
Revised: 9 September 2024
Accepted: 15 September 2024
Published: 23 September 2024

1. Introduction

Industrial control systems (ICSs) are increasingly critical in modern infrastructure and significant projects such as the hydraulic facility, transport, energy, and chemical industries. In this sense, ICS security is also directly linked to the smooth operation of critical infrastructures [1]. When the ICSs are attacked, it will directly harm the physical world by causing environmental pollution, power outages, oil leaks, and explosions. With the acceleration of the digitalization process of ICSs, the integration of industrialization and informatization has been gradually strengthened. Due to the increasing openness of industrial control systems, there are a increasing number of threats to the systems. Hence, timely and accurate anomaly detection in ICSs is essential in reflecting the security status of the production process and determining the vulnerability of the industrial control systems. Maintaining secure operation of industrial control networks is increasingly critical in improving production efficiency and safety [2,3].

ICSs are a series of control systems, which include supervisory control and data acquisition systems, distributed control systems and programmable logic controllers (PLCs), and other control systems and control units. An ICS ensures the safe, reliable, and secure operation of industrial processes. In ICSs, malicious attacks are possible due to the inherent

loopholes in communication protocols. Recently, industrial control networks have faced constant threats, such as the Stuxnet virus attacking PLC codes to achieve such an attack, thereby destroying the centrifuge's regular operation. Thus, numerous researchers have devoted themselves to constructing state models for ICSs to enable anomaly detection for different attacks.

The ICSs' layers interact and communicate with one another via the network while carrying out their specific assigned tasks. The ICS is vulnerable in both the network along physical layers due to its close coupling between cyberspace and physical space. An attacker may launch a cyberattack, which could result in malicious software and data asset theft or tampering with the equipment, leading to the loss of crucial control information and failure of crucial control commands.

The threat of attack has caused worldwide concern about the cyber security of ICSs. Given the numerous attack threats faced by ICSs, Teixeira et al. proposed an ICS pass-through attack model based on three-dimensional information and physical space to characterize the various attack means in different spaces and to illustrate the characteristics of multiple types of attacks [4]. Accordingly, Adepu et al. proposed a framework for describing physical attacks, cyberattacks, and other types of attacks by dividing them into domain, attacker, and attack models [5]. In light of the wide variety of attack types, complex attack paths, and variable attack strategies facing ICSs, it is challenging to construct a mathematical model covering all scenarios.

Currently, data-driven methods of extracting information from process data and modeling monitoring have become a hotspot in anomaly detection research. The advancement of sensor technology has allowed almost all industrial objects to be equipped with various types of sensors, which has resulted in a great deal of data being collected in industrial processes. By merging the data from various sources and examining the correlation between the information, data-driven anomaly detection methods can detect whether a system is under attack. A relational model that captures the intruder's identity, velocity, level of threat, and target of intrusion was developed, which serves as a foundation for continuous cyberspace state monitoring [6]. Lu et al. proposed a security monitoring method for industrial control networks based on an improved C-SVC (C-Support Vector Classifier), which can effectively identify multiple types of abnormal states and form situational awareness results [7]. A hidden Markov model-based attack detection for Stuxnet has been proposed in the industrial control system subject to random packet dropouts [6]. Despite being based on mechanistic models of attack-induced abnormal states, the methods above have inherent limitations when applied to large-scale complex industrial processes.

Considering the large scale and complexity of the system in question, as opposed to complex processes mechanisms, researchers have monitored network security status by analyzing the process data in industrial control networks. Multivariate statistical process monitoring (MSPM) methods have been widely studied and applied over the past few decades [8]. Rather than modeling a particular attack model, MSPM depicts the operational state of the system. Attack detection on ICSs is achieved by comparing the deviations from the operational state. Among the most well-known representative branches of statistical process monitoring is principal component analysis (PCA), which is regarded as an effective means of dimension reduction. PCA identifies the major changes in data by decomposing multiple related variables into several orthogonal principal components [9,10]. The PCA-based MSPM approach enables monitoring by modeling the variable space of the system where two different monitoring statistics, Hotelling T^2 and Squared prediction error Q, are viewed as the monitoring statistics [11,12].

Although PCAs are widely used to detect anomalies, they do not perform as well when their assumptions are incorrect. The underlying Gaussian assumption in the calculation of control limits of monitoring statistics in PCA makes it a poor monitoring tool for non-Gaussian processes. A variety of PCA variants have been proposed for nonlinear processes, including probability PCA (PPCA) [13] and kernel PCA (KPCA) [14], in which the data are projected into a high-dimensional space. In essence, KPCA remains a linear

dimensionality reduction method, and its effectiveness is heavily influenced by the choice of kernel function, which is not appropriate for systems with nonlinear or stochastic perturbations. Within the maximum likelihood framework, PPCA measures the similarity between new data points according to their probability density functions [15,16]. Canonical variate analysis (CVA), which provides a more accurate description of the process by maximizing the correlation between mainly dependent and quality variables [17,18], is another valid method for incorporating both static and dynamic process characteristics. Zhang et al. developed a CVA-based modeling and monitoring method for simultaneous static and dynamic analysis in three-phase flow processes [19,20]. A fault information-aided canonical variate analysis and a structured monitoring strategy has been proposed to improve anomaly detection rate [17]. However, the process is usually assumed to operate under one condition, whereas industrial processes always operate in multiple modes.

For plant-wide processes, multimodal methods were introduced as a solution to these problems. Generally, block division is the key step in sub-block modeling. These methods can be classified into two main categories: data-driven and knowledge-based. Based on field experience and prior process knowledge, knowledge-based methods usually divide process variables into blocks. A hierarchical multiblock total projection to latent structures (T-PLS) based on an operating performance assessment scheme was proposed to identify the anomalies in operating statuses [21]. Using prior process knowledge, Zhu et al. proposed the distributed parallel PCA process monitoring framework to decompose the high-dimensional process variables [22]. When there is a lack of accurate prior knowledge, monitoring and anomaly detection performance may be less than optimal if the process variables are not correctly divided.

Data-driven methods have also been extensively used to divide variable blocks in distributed process monitoring using the process measurements from industrial historians. The data-driven approach clusters variables into sub-blocks by evaluating the correlations between variables. For instance, Hu et al. used mutual information (MI) analysis to extract the complex relationships between each possible process variable and the burn-through point in the sintering process [23]. Zhang et al. investigated an improved mixture probability principal component analysis with clustering for nonlinear process monitoring where the *k*-means is subsequently utilized as a clustering algorithm to divide the variables into optimal sub-blocks [24]. Minimal redundancy maximal relevance was used to divide the most related variables into the same block and form a dynamic multiblock monitoring framework [25]. With mutual information-spectral clustering, the measured variables were automatically divided into sub-blocks on which a Bayesian inference-based multiblock KPCA monitoring model was established [26]. Combining knowledge-based and data-driven approaches, Cao et al. developed a hierarchical hybrid, distributed PCA for the plant-wide monitoring of chemical processes with two-layer manner sub-block division.

Although the aforementioned monitoring strategies have been demonstrated as effective, the monitoring performance may not be optimal when faced with sophisticated cyberattacks. On the one hand, network layer attacks such as data injection present more randomness and uncertainty than faults in the system. The above characteristics lead to traditional monitoring methods failing to identify the dynamic characteristics caused by attacks when modeling with normal samples. Specifically, PCA-based monitoring methods cannot fully extract the state-by-state characteristics of the system in the principal metric space, leading to omissions and false alarms in the monitoring results. Similarly, when confronted with large-scale complex systems, the traditional centralized modeling approach cannot adequately reconstruct the system's state characteristics.

Motivated by the above research status, a concurrent distributed monitoring method was proposed to tackle the ICS attack detection tasks. redUsing a two-stage distributed modeling approach, we can extract all the state characteristics of the system. By using the MI method, the decision variables are selected and the distributed structure is realized. Then, the PPCA models compute both the serially correlated subspace and its residual

subspace based on canonical variate analysis, which makes a complete interpretation of process dynamics under ICSs possible.

In the proposed framework, all detection variables were selected into the first layer by one-way variance analysis, and the detection variables were further divided into sub-blocks using a combination of general knowledge-based strategies with mutual information. Then, CVA–PPCA monitoring models were established for each sub-block, in which CVA was used to explore the serial correlations, and PPCA-based monitoring models were constructed for the variables of subspace. Finally, Bayesian inference was used to obtain comprehensive statistical indicators of the ICSs, which can realize plant-wide anomaly detection. Thus, the dynamic characteristics of the ICSs were restored, allowing for a deeper understanding of its security status. The main contributions of the present work are as follows.

1. An adaptive process variable selection and blocking method for distributed monitoring was implemented with combined knowledge-based strategies with mutual information.
2. Both linear and non-linear behaviors were analyzed and monitored, which can provide a meaningful interpretation for fine-scale identifying ICS attacks.

The rest of this paper is organized as follows. The problem description and monitoring framework are given in Section 2. Section 3 outlines the proposed concurrent distributed CVA-PPCA-based monitoring method in detail. Section 4 details a validation of the effectiveness of the proposed method on actual drilling processes. Finally, conclusions are made in Section 5.

2. Problem Description and Modeling Framework

In this section, the problems of ICS security monitoring are summarized. Based on these, a framework of monitoring model was designed.

2.1. Problem Description

ICS is an umbrella term for various network-connected control systems in the industrial field. Over the past few decades, ICSs have greatly enhanced the degree of industrial process automation and brought certain security risks. Figure 1 shows a typical industrial control network architecture for the geological drilling processes. A controller employs a communication network to regulate the operation of the controlled process by measurements from geographically dispersed sensors.

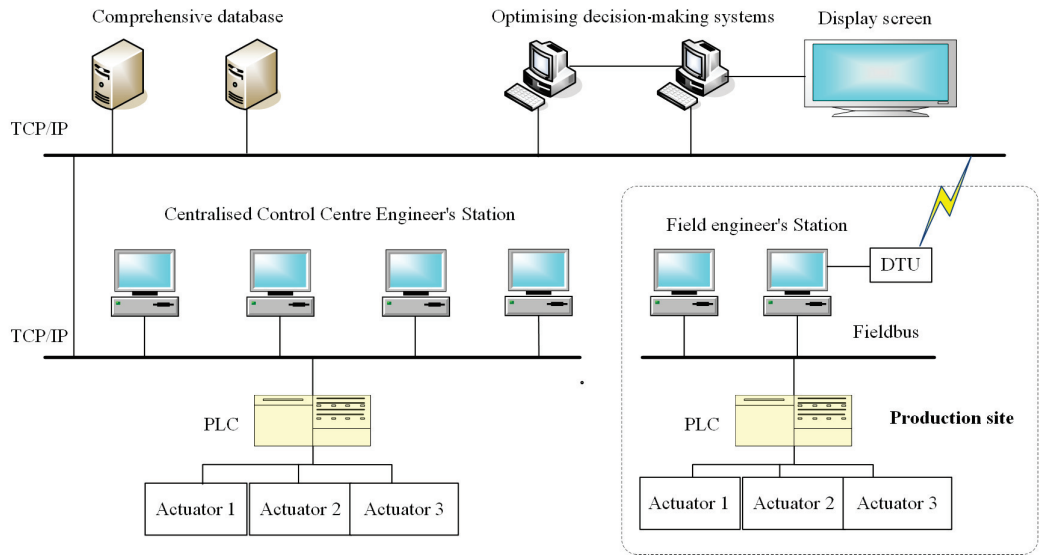

Figure 1. A typical industrial control system structure for the drilling process.

During the drilling process, the PLC is responsible for controlling the industrial control system in order to read the data from the field sensors. Additionally, the Profibus communication protocol was utilized in order to facilitate communication between the PLC and the industrial control machine. For the purpose of reading the data from the PLC over the OLE for Process Control (OPC) protocol, the WinCC configuration software was utilized. MVC (Model–View–Controller) architecture was utilized by the system, which enables intelligent optimization control, as well as complicated logic operations.

A system failure results from an attacker's deliberate destruction or manipulation of actuators, control units, etc., which is another manifestation of the ICS vulnerability in the physical layer. Network attacks and instrument malfunction both appear as anomalies in the data sampled by the sensors. The difference is that network attacks cause equipment failure, so the data usually show a causal relationship between them. Additionally, network attacks tend to maintain the statistical characteristics of the data sparingly, whereas equipment failures often result in outliers, missing values, and other easily observable changes. Due to the complexity of physical layer attacks, the attack detection algorithms in this paper only address attacks suffered at the network layer.

False data injection is a common network layer attack. In the event that sensors transmit sensing data to the PLC, the data may be tampered with, leading to the instability of the control system. In this attack, the original correct measurement value $z_i(t)$ of moment t will be tampered with, resulting in the measurement value $\tilde{z}_i(t)$ deviating from the normal value $z_i(t)$, which causes the feedback control system to perform incorrect responses. The attack process can be expressed as [7]

$$z_i(t) = \begin{cases} 0, t \notin T_{\text{atc}}, \\ z_i(t-1) + \tau\varphi^t, t \in T_{\text{atc}}, \end{cases} \tag{1}$$

where τ and φ are the impact index, which is usually a constant; and T_{atc} is the attack period. This paper assumed that the anomalous state of the system was caused by fake data that were imposed by the attacker.

In general, false data injection attacks include the manipulation of system measures while the attacker is aware of the setup of the system. These attacks are difficult to monitor directly since they are difficult to detect. The three primary types of attacks that fall under the category of fake data injection assaults are known as surge attacks, deviation attacks, and geometry attacks. To varied degrees and at varying rates, the normal operation of the system is disrupted in each of these instances, and, when it is severe, it is likely to result in serious accidents. Figure 2 presents histogram plots of the partial variables in the geological drilling process, such as the rate of permeation (ROP) as an example. Clearly, the distribution of data that is not ideal (shown by the red area) is mostly contained within the distribution of data that is optimal (represented by the blue area). Since this is the case, one of the most important concerns in ICS security monitoring is how to further parse data features. Monitoring the current status of network security can assist decision makers in determining whether or not an attacker intends to launch an attack. The operation of the system will be guaranteed to be stable and secure as a result of this.

In a data tampering attack, the attacker tampers with measured values of a system since he knows the system configuration and cannot be detected intuitively. Therefore, the following challenges need to be faced when investigating ICS-oriented attack detection methods.

1. Complexity: The number of current cyberattacks on ICSs is increasing, with attackers exploiting ICS vulnerabilities to deliver different types of attacks and threats.
2. Crypticity: There are insufficient means of identifying attack behavior, and the attack detection false alarm rate is high due to attackers deliberately confusing the attack with the normal operation of the control system.

Therefore, an essential component of achieving ICS attack detection involves developing a monitoring model that accurately captures the dynamic aspects of the attack behavior.

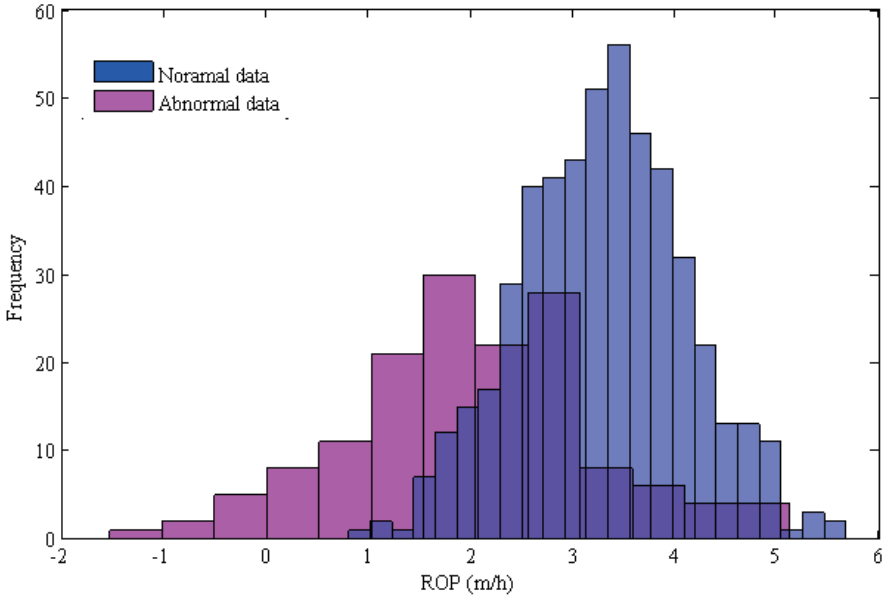

Figure 2. Histograms of the drilling data under optimal and non-optimal modes.

2.2. Modeling Framework

The objective of this study was to detect the abnormalities of ICSs by constructing a process monitoring model based on the sufficient normal data of related detection variables. A novel CVA-PPCA-based monitoring method was presented to overcome the shortcomings and improve the performance of network anomaly identification. The framework of the proposed network condition monitoring scheme is shown in Figure 3.

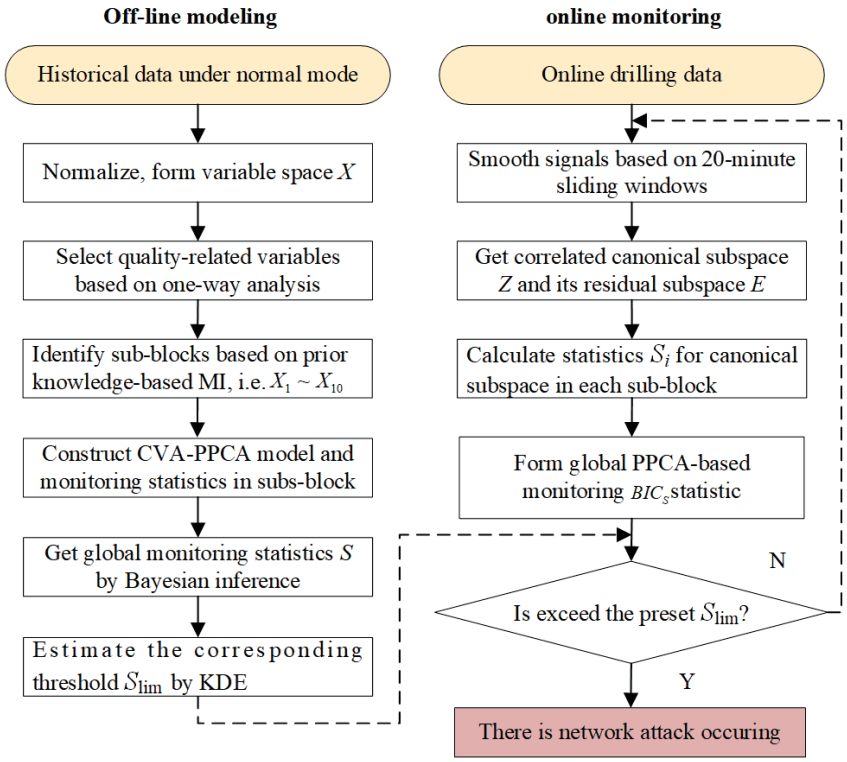

Figure 3. The framework of the proposed CVA–PPCA-based monitoring method.

The monitoring model consists of two parts: offline modeling and online monitoring. According to one-way analysis, the ICS performance quality-related detection variables were chosen; these were then further divided into reasonable sub-blocks by the MI analysis, which were added with prior knowledge. Within each sub-block, the CVA method was used to classify the variables according to their correlation into their correlated canonical subspace and residual subspace. Then, the PPCA-based monitoring model was established in canonical subspace. Finally, Bayesian inference was used to obtain comprehensive statistical indicators of the whole process, which can realize anomaly detection.

For online monitoring, real-time monitoring statistics can be compared with historical data to determine the overall performance of the integrated monitoring system and to define the detection thresholds according to attack type. Anomalies can then be detected by comparing the monitoring statistics to see if the limits have been exceeded.

3. Implementation of the Monitoring Model

In this section, the ICS security monitoring model is established. Firstly, sub-block division was carried out using one-way analysis and mutual information analysis. Using the CVA method, the original variable space was divided, and the PPCA monitoring model with preset control limits was constructed. To achieve online monitoring, the online data are used to calculate the monitoring statistics and compare them to the detection threshold.

3.1. Sub-Block Division Based on One-Way Analysis and Mutual Information Analysis

There are usually multiple industrial controls and multiple systems within ICSs. The whole process contains a number of detection variables. The multi-block modeling approach is an effective way to deal with the anomaly detection problem of large-scale processes. To fully extract the correlations between variables, sub-block division is necessary before offline modeling.

A two-stage delineation method was used in this study to create a multi-sub-block structure, with one-way analysis of variance being selected in the first stage to determine the operational state-related decision variables, which was followed by mutual information analysis and process knowledge for sub-block delineation.

In the first phase, one-way analysis of variance (ANOVA) can be used to determine the effect of the different operating modes on the distribution of variable data. By measuring the difference in the variance fluctuations caused by different operating conditions and random errors, ANOVA determines if changes in the operating conditions are a major factor in system operation.

There are five normal geological drilling conditions: drill up and down, rotary drilling, back reaming, hole sweeping, and sliding drilling. Assuming that the number of samples for each operating condition is selected as n_1, n_2, n_3, n_4, and n_5, then the drilling data for each condition is recorded as $x_{1,j}, x_{2j}, \ldots, x_{n_j,j} (j = 1, 2, \ldots 5)$. The degree of variation V_T between the drilling data can be calculated as follows:

$$\begin{aligned} V_T &= \sum_{j=1}^{5} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}\right)^2, \\ \bar{x} &= \frac{1}{N} \sum_{j=1}^{h} \sum_{i=1}^{n_j} x_{ij} = \frac{1}{n} \sum_{j=1}^{5} n_j \bar{x}_j, \\ \bar{x}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}, \end{aligned} \tag{2}$$

where $\bar{x}$ is the mean value of data collected for the variable, and $\bar{x}_j$ is the mean value of the variable in a data set for a mode. Furthermore, V_T can be decomposed into the sum of its

error sum of squares and effect sum of squares, which is denoted as $V_T = V_E + V_F$, and V_E and V_F are relative-independent, the details of which can be defined as

$$\begin{aligned} V_E &= \sum_{j=1}^{h} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2, \\ V_F &= \sum_{j=1}^{h} \sum_{i=1}^{n_j} \left(\bar{x}_j - \bar{x}\right)^2. \end{aligned} \tag{3}$$

According to the above definition, it is clear that V_T measures the distributional differences within different drilling conditions and also globally. Thus, it is possible to select the operating status-related variables related to the ICSs effectively. The degree of influence of a variable is measured by constructing a test statistic F_T and its test probability ρ:

$$F_T = \frac{(V_F)/(h-1)}{(S_E)/(N-h)} \sim F(h-1, N-h), \tag{4}$$

where F_T obeys F-distribution, $h = 5$, and the test probability is $\rho(F(h-1, N-h) \geq F_T)$. The smaller the test probability, the greater the effect of the parameter on the operating conditions.

Table 1 presents the test probability of each parameter based on 1800 samples of data collected from the industrial control network in the drilling process. Clearly, the probability of testing parameters d_{11} and d_{12} is significantly higher than those of the other variables, which is also consistent with the process knowledge. A total of 10 variables can be selected for $\rho \leq 0.001$, i.e., $X_1, X_2, \ldots, X_{10}$.

Table 1. Results of the one-way analysis of variance.

Parameter	Description	ρ
d_1	Rate of penetration (km/h)	6.27×10^{-8}
d_2	Weight on bit (kN)	0.43×10^{-10}
d_3	Rotation speed (r/min)	1.84×10^{-8}
d_4	Mud flow in (out) (L/s)	3.27×10^{-6}
d_5	Tank volume (m^3)	4.17×10^{-15}
d_6	Standpipe pressure (Mpa)	9.38×10^{-12}
d_7	Hookload (kN)	2.86×10^{-14}
d_8	Hook height (m)	1.22×10^{-18}
d_9	Rotary torque (kN·m)	8.86×10^{-18}
d_{10}	Depth (m)	1.09×10^{-9}
d_{11}	Bit dept h (m)	5.36×10^{-2}
d_{12}	Bit diameter (mm)	6.25×10^{-3}

In the second stage, the detection variable blocking is based on MI combined with prior knowledge. MI involves determining whether a detection parameter's data distribution and a performance indicator's distribution are interdependent. When several variables interact, MI is the entropy that was initially contained as it decays. It suggests that information entropy is not constant but rather varies with the number of events that occur. MI is commonly interpreted as a metric that quantifies the degree of dependence and strength between two variables. Specifically, given two random variables x_1 and x_2, the mutual information between them is defined as

$$I(X, Y) = \sum_{X} \sum_{Y} p(X, Y) \log \frac{p(X, Y)}{p(X)p(Y)}, \tag{5}$$

where $p(x)$ and $p(X)$ are the marginal probability density functions of X and Y, and $p(X, Y)$ is the joint probability of X and Y.

As this equation represents the uncertainty in x_2 after removing x_1, it confirms the intuitive meaning of MI as the amount of information one variable provides about another. By analyzing the physical mechanism of the drilling production, it can be seen that d_4, d_5, and d_6 are part of the mud system, and d_2 and d_3 are also one of the d_1-influencing parameters. Then, according to the blocking criterion [13], these variables were divided into three sub-blocks: $[X_1, X_2, X_3, X_9]$, $[X_4, X_5, X_6]$, and $[X_7, X_8, X_{10}]$.

Hence, the detection variables were blocked according to their interrelationships using the MI combined with prior knowledge, and the CVA-PPCA anomaly detection model is then applied on a distributed sub-block structure.

3.2. Canonical Subspace Identification Based on CVA

The drilling detection variables $d_1 \sim d_{10}$ are categorized into distinct sub-blocks based on current correlations. Then, state monitoring models would be constructed within each sub-block by parsing the data characteristics to accomplish anomaly detection for various attack methods.

Canonical variate analysis (CVA) is a dimension reduction algorithm that maximizes the alignment between two sets of variables. By maximizing the correlation between the "past" values and the "future" values of the system, the CVA-based approach generates state-space models from time-related data. Thus, CVA can be used to establish the relationships between process variables and quality variables, and the trained CVA model can be used for process monitoring related to quality.

In CVA, linear dimension reduction is used to reduce the size of variables so that it can be used to determine the most significant correlation between qualitative and primary dependent variables, as well as dynamic processes [20]. This study addresses the autocorrelation challenge of modeling the operational state of industrial control networks.

The past and future drilling data matrix is constructed using drilling data $x_k = [X_1, X_2, \ldots, X_n]^T (k = 1, 2, \ldots N; n = 12)$. Assume that, at moment k, the past vector $x_{p,k}$, comprising the past data, and the future vector $x_{f,k}$, containing the present and future observations, are defined as

$$\begin{aligned} x_{p,k} &= \left[\begin{array}{c} x^{\mathrm{T}}_{(k-1)}, x^{\mathrm{T}}_{(k-2)}, \ldots, x^{\mathrm{T}}_{(t-l)} \end{array} \right]^{\mathrm{T}}, \\ x_{f,k} &= \left[\begin{array}{c} x^{\mathrm{T}}_{(k)}, x^{\mathrm{T}}_{(k+1)}, \ldots, x^{\mathrm{T}}_{(k+l)} \end{array} \right]^{\mathrm{T}}, \end{aligned} \tag{6}$$

where the two vectors, i.e., $x_{p,k}$ and $x_{f,k}$, should first be normalized to a zero mean and with unit variance. To define the past and future matrices, vectors were arranged in the following Hankel matrix:

$$\begin{aligned} X_p &= \left[\begin{array}{c} x_{p(l+1)}, x_{p(l+2)}, \ldots, x_{p(l+N_1)} \end{array} \right], \\ X_f &= \left[\begin{array}{c} x_{f(l+1)}, x_{f(l+2)}, \ldots, x_{f(l+N_1)} \end{array} \right], \end{aligned} \tag{7}$$

where $N_1 = N - 2l + 1$ for a dataset with N samples.

The aim of CVA is to reveal the remarkable features of the ICS operating conditions by identifying the projection matrix L and J in order to identify a linear combination of the future and past observations that have the optimal linear performance. The problem of solving the projection matrix is defined as follows:

$$\begin{aligned} &\max_{J,L} J^T \Sigma_{pf} L, \\ &s.t. J^T \Sigma_{pp} J = I, \\ &L^T \sum_{ff} L = I. \end{aligned} \tag{8}$$

The projection matrix J and L can be calculated by singular-value decomposition (SVD) on the Hankel matrix H as follows:

$$H = \Sigma_{ff}^{-1/2}\Sigma_{fp}\Sigma_{pp}^{-1/2} = U\Lambda V^{\mathrm{T}}, \tag{9}$$

where the sample covariances $\Sigma_{ff}^{-1/2}$ and $\Sigma_{pp}^{-1/2}$ and the cross-covariance of Σ_{fp} of the past vector $x_{p,k}$ and the future vector $x_{f,k}$ are defined as follows:

$$\begin{bmatrix} \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{fp} & \Sigma_{ff} \end{bmatrix} = \frac{1}{N_1 - 1}\begin{bmatrix} X_p X_p^T & X_p X_f^T \\ X_f X_p^T & X_f X_f^T \end{bmatrix}, \tag{10}$$

where U and V consist of singular vectors that are orthogonal and only pairwise-correlated, and Λ is a diagonal matrix containing the canonical correlation coefficients. Thus, the projection matrices J and L can be calculated by taking the first r columns of U and V, respectively.

For the k moments of the ICS operation, the transformation matrices J and L are as follows:

$$\begin{aligned} J_r &= V_r^T \Sigma_{pp}^{-1/2}, \\ L_r &= U_r^T \Sigma_{pp}^{-1/2}. \end{aligned} \tag{11}$$

The canonical state subspace Z and its residual subspace E of the drilling data matrix x can be defined as

$$\begin{aligned} Z &= J_r X_p \in R^{r \times N_1}, \\ E &= F_r X_p \in R^{n_1 l \times N_1}, \end{aligned} \tag{12}$$

where the residual projection matrix $F_r = (I - V_r V_r^T)\Sigma_{pp}^{-1/2}$.

Therefore, the space of the primary and dependent variables Z, which are canonically correlated with the ICSs' operational performance, is extracted within each sub-block. Then, a PPCA-based monitoring model is built on it to detect cyberattacks.

3.3. Overall Monitoring Model

According to CVA, the ICS variable space for drilling processes consists of a correlated canonical and residual subspace. It is necessary to establish a model for monitoring subspace in order to implement the proposed scheme.

PPCA-based monitoring model: The PPCA method is a representation of PCA in probability space, where probability density functions measure the degree of the novelty of new data points. While PCA is a linear down-scaling method, PPCA can take into account the nonlinear and dynamic characteristics of the system fully. When dealing with non-linear characteristics, PCA is vastly improved by the incorporation of probability. Data x is believed to be generated by the latent variable z when viewed from the perspective of probability. In order to produce the standard PPCA, the following pattern is utilized [15]:

$$x = f(z, w) + \xi, \tag{13}$$

where $x \in R^d$ is the process observation variable, $z \in R^p$ is the vector of latent variables, $w \in R^{n \times q}$ is the associated model parameter vector like loading matrix, ξ is an independent noise vector, and $f(\cdot)$ describes the unknown function, which can be interpreted by a linear model in general.

$$X = WZ + \mu + \xi, \tag{14}$$

where $X \in R^{d \times n}$, $Z \in R^{q \times n}$, and μ is the monitoring delay. The model parameters are then determined using a maximum-likelihood technique given a set of observational data.

According to the canonical subspace $Z \in \mathbb{R}^{n \times m}$ acquired in the previous section. The PPCA algorithm seeks the projection matrix $W \in \mathbb{R}^m$ to further reveal both the static and

dynamic process variations in which the linear transformation $Z_c = ZW$ has the maximal variance. Like PCA, the problem of matrix projection can be expressed mathematically as

$$\arg\max_{W} \frac{1}{n-1} {Z_c}^{\mathrm{T}} Z_c = \arg\max_{W} \frac{1}{n-1} W^{\mathrm{T}} Z^{\mathrm{T}} Z W. \tag{15}$$

The transformed goal of the PPCA is to map the original m-dimensional data into a d-dimensional space, whose principal element model T can be expressed as

$$Z = \sum_{i=1}^{p} z_{c,i} w_i^{\mathrm{T}} + \sum_{i=p+1}^{m} z_{c,i} w_i^{\mathrm{T}} = Z_{\mathrm{c}} W^{\mathrm{T}} + E_c, \tag{16}$$

where W is the load matrix; Z_c is the scoring matrix; P is the number of principal components retained, which is commonly determined by a rule known as the cumulative percentage variance (CPV) [27]; and $E_c = Z - Z_c$ is the residual matrix, which represents process noise interference.

In general, the principal element is associated with a multivariate standard–normal distribution, while the noise residual is associated with a multivariate normal distribution, where $Z_c \sim \mathcal{N}(0, I)$, $E_c \sim \mathcal{N}(0, \sigma^2 I)$ and σ^2 is the noise variance. Then, the distribution of sample Z with respect to principal element Z_c is $Z|Z_c \sim G(Z_c W^{\mathrm{T}}, \sigma^2 I)$. According to Bayes' theorem, the distribution of the sample data X is $X \sim G(0, C)$, and $C = WW^{\mathrm{T}} + \sigma^2 I$.

Thus, the problem solved by the PPCA algorithm can be seen as forming observations Z from the distribution $G(0, C)$ by the hidden variable Z_c. The problem to be addressed translates into the estimation of the distribution parameters W and σ from the measurement samples [24]. This paper solves the probability distribution using the maximum-likelihood estimation problem. Expectation maximization (EM) is a powerful method for estimating the parameters of hidden variable models, which uses an expectation maximization algorithm that iterates repeatedly to find the parameters.

Online attack detection: To monitor the state of the ICSs online, the monitoring threshold must first be determined. Traditionally, PCA-based monitoring methods calculate two types of statistics, T^2 and Q, as well as the corresponding control charts. Specifically, the T^2 statistic is designed to monitor the data variations in the principal component space (PCS), while the Q statistic is used to monitor the data changes in the residual space. Observations of large deviations in the monitoring statistics may indicate an abnormal state of the industrial control network.

On the basis of the PPCA algorithm, the principal component space Z_c, contains systematic variation information and will be used to construct the T^2 statistic, while the residual E_c will form the Q statistic. The monitoring statistics are defined as

$$\begin{aligned} T^2 &= z_c^{\mathrm{T}} \Lambda^{-1} z_c, \\ Q &= \|z - z_c\|^2 = (z - w z_c)^{\mathrm{T}} (z - w z_c). \end{aligned} \tag{17}$$

In the case of a multivariate normal distribution for the process variables, the detection threshold for T^2 can be obtained using the F-distribution with α as the significance factor:

$$T^2 \sim \frac{r(n^2-1)}{n(n-p)} F_{r,n-p,\alpha}, \tag{18}$$

where p is the number of PCSs. As with the residual subspace, a weighted Chi-squared distribution can approximate the confidence limit of Q, such as

$$Q \sim d\chi^2_{g,a}, \tag{19}$$

where $d = v_c/2m_q$ and $g = 2{m_q}^2/v_c$, in which m_q is the mean value of Q, and v_c is the corresponding variance.

As the PPCA exclusively employs the Martensian paradigm for the detection of principal elements and noise [28], the comprehensive monitoring statistics, which consist of T^2 and Q, can be directly generated from the whitened values of the statistics. The following formats were used to calculate the comprehensive monitoring statistic S:

$$S = \left\| \left(WW^{\mathrm{T}} + \sigma^2 I \right)^{-0.5} z \right\|^2 = z_c^{\mathrm{T}} \left(\sigma_i^2 I + WW^{\mathrm{T}} \right)^{-1} z_c. \tag{20}$$

As a result of the proposed monitoring model, which effectively detects the data injection attacks on the ICSs, $S_{\lim}$ is the threshold determined by kernel density estimation (KDE) [29], which is the measurement of the degree of deviation from the normal operating conditions. Additionally, S is the monitoring statistic based on the PPCA, and $S_{\lim}$ is the threshold determined by the kernel density estimation (KDE). The threshold $S_{\lim}$ is given by

$$P(S \le S_{\lim}) = \int^{S_{\lim}} \hat{\phi}(s|W, \sigma) ds = 1 - \alpha, \tag{21}$$

where $\hat{\phi}(s|W, \sigma)$ is the probability density function of S estimated by KDE. If the corresponding detection logic satisfies, for example, $S \le S_{\lim}$, the operating performance is optimal; otherwise, it is non-optimal.

According to the previous discussion, there are several sub-blocks formed here. There is a need to integrate local statistics to construct comprehensive surveillance indicators for the whole process. This study used Bayesian inference to integrate the monitoring results of multiple sub-blocks into the overall monitoring results due to its excellent performance in sub-block decision fusion. Conceptually, the probability of each sub-model being under attack can be expressed as

$$P_S(F|x_i) = \frac{P_S(x_i|F) P_S(F)}{P_Q(x_i)}, \tag{22}$$

where the prior probability of x_i is calculated as

$$P_S(x_i) = P_S(x_i|N) P_S(N) + P_S(x_i|F) P_S(F), \tag{23}$$

and the conditional probabilities $P_S(x_i|N)$ and $P_S(x_i \mid F)$ are defined as

$$P_S(x_i|N) = e^{-S_i / S_{i,\lim}}, \tag{24}$$

$$P_S(x_i \mid F) = e^{-S_{i,\lim} / S_i}, \tag{25}$$

where S_i represents the statistic in the i-th sub-block and $S_{i,\lim}$ represents the control limits in the i-th mode blocks; N and F denote the optimal and non-optimal operating performance, respectively; $P_S(N)$ and $P_S(F)$ represent the prior probabilities under the confidence level α and 1-α; and $P_S(N) + P_S(F) = 1$. The intuitive interpretation is that the operating status expressed by sampling data is either normal or non-optimal in the drilling process.

After that, in the modeling phase, it is possible to obtain comprehensive monitoring indicators by integrating the PPCA sub-models for various operating modes based on Bayesian inference.

$$BIC_S = \sum_{i=1}^{m} \left\{ \frac{P_S(x_i \mid F) P_S(F \mid x_i)}{\sum_{i=1}^{m} P_S(x_i \mid F)} \right\}. \tag{26}$$

During the actual monitoring process, it can be determined that the ICSs have received an attack when the monitoring indicator exceeds the preset threshold.

4. Experimental Results and Analysis

This section verifies the validity of the methodology through practical examples, which are derived from the geological drilling process, and is divided into processes.

4.1. Geological Drilling Process

Geological exploration and resource extraction are contingent upon the successful completion of a geological drilling project. The drilling process is primarily conducted by drill rigs that are equipped with alternative current frequency conversion electric motors. Figure 4 illustrates the schematic of a typical geological drilling process. A few of the components that were used in the drilling process included the crown blocks, moving blocks, derrick, driller's residence, rotary table, drilling control system, mud pump, mud pit, sedimentation pit, drill string, bottom hole assembly, and drill bit. Figure 5 shows a geothermal well construction site with an on-site industrial control system.

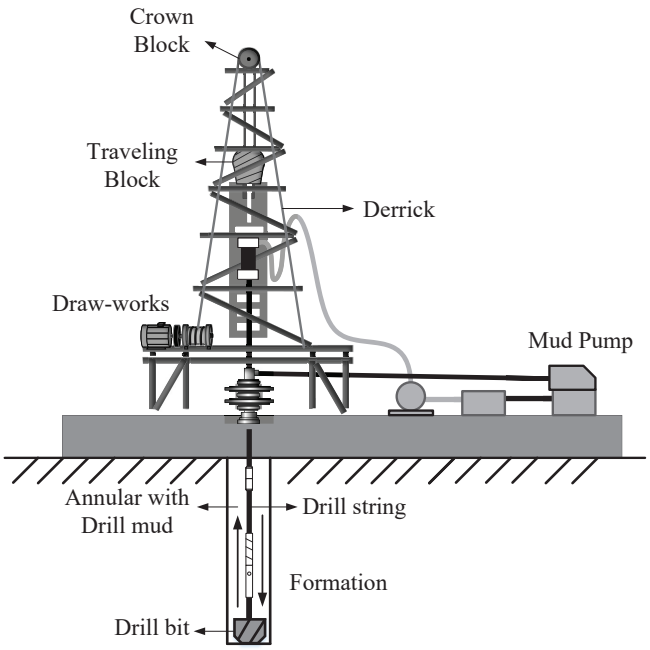

Figure 4. Schematic of a geological drilling process.

Figure 5. The drilling system of a real geological exploration well.

4.2. Overall Results of the ICS Attack Detection

In this paper, real-life case studies with drilling data from a geothermal well demonstrated the effectiveness and superiority of the proposed operating performance monitoring method. The selected running data contains the 12 process variables mentioned in Table 1 from 1052 m to 1058 m, with an interval of 1 s, totaling 2826 data samples. Figure 6 demonstrates time-series data of the actual running process of the ICSs during drilling. Despite the fact that the data injection attack on the network began at 160 s, no significant change was observed in the data curves of the detected variables. Therefore, more in-depth analyses of the data generated in the ICSs are needed to obtain a more accurate portrayal of the ICS operating state.

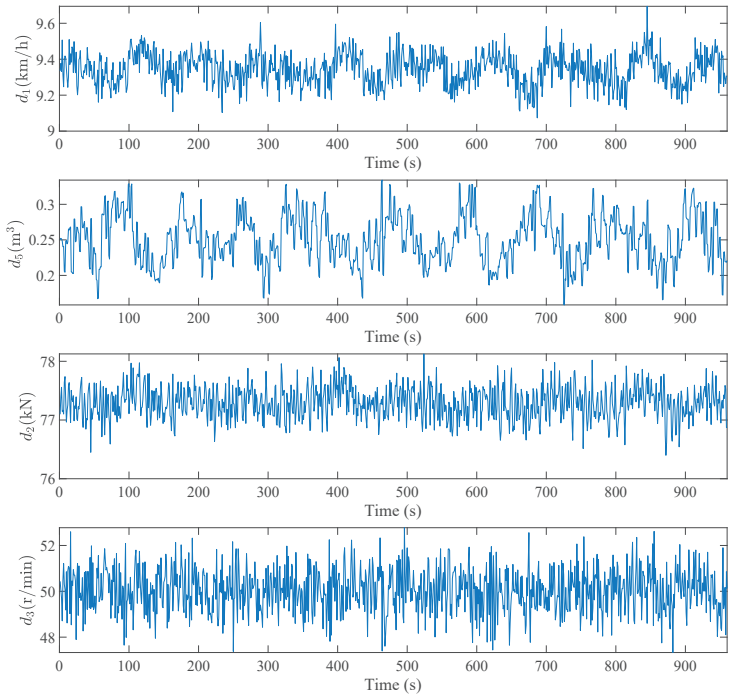

Figure 6. Time series plots of the drilling process under production.

Before constructing the ICS monitoring model, the data set under normal operations was obtained. A standard data matrix was created by selecting 10 decision variables based on a one-way analysis (ANOVA), i.e., $X_1, X_2, \ldots, X_{10}$. According to the blocking MI-based criterion, these variables were divided into three sub-blocks, $[X_1, X_2, X_3, X_9]$, $[X_4, X_5, X_6]$, and $[X_7, X_8, X_{10}]$. For each sub-block, the CVA-PPCA offline monitoring model is established on their canonical subspace, and the calculation of the composite discriminatory indicators and discriminatory thresholds are performed.

During the online monitoring phase, online data are collected according to a window of 20 min, and the monitoring statistic S_{new} is calculated to identify the attack conditions in comparison with the detection threshold. The length of the monitoring window has some effect on the quality of the monitoring. A long window may not detect the fluctuations caused by dual-use attacks, such as, for instance, when there is too short of a window, which may cause frequent alarms and may interfere with the driller's normal operation. Using the industrial control system at the drilling site and manual experience, this study specified a 20-min monitoring window, leading to better results.

Specifically, the principal components of the variables with CPV $= 98.2\%$ were selected to construct the monitoring model. In all of the monitoring charts, the KDE algorithm was adopted to preset the control limits at a confidence level of $\alpha = 0.05$ and monitoring statics $S_{\text{lim}} = 1.2961$.

In this paper, the anomalous state of the ICSs was the result of two categories of data tampering: surge attacks and biased attacks [30,31]. During a surge attack, a single piece of data is manipulated in order to provide the greatest amount of damage in the shortest amount of time, and it exhibits a step change. Contrary to this, a biased attacker adds non-zero constants to numerous parts of data in a sequence and shows a slow process of change. The monitoring model in this paper was intended to detect the assaults that the system has received by analyzing the monitoring statistics that had been generated by the attacks relevant to the change. Figure 7 illustrates the ICS attack detection results obtained through the proposed method.

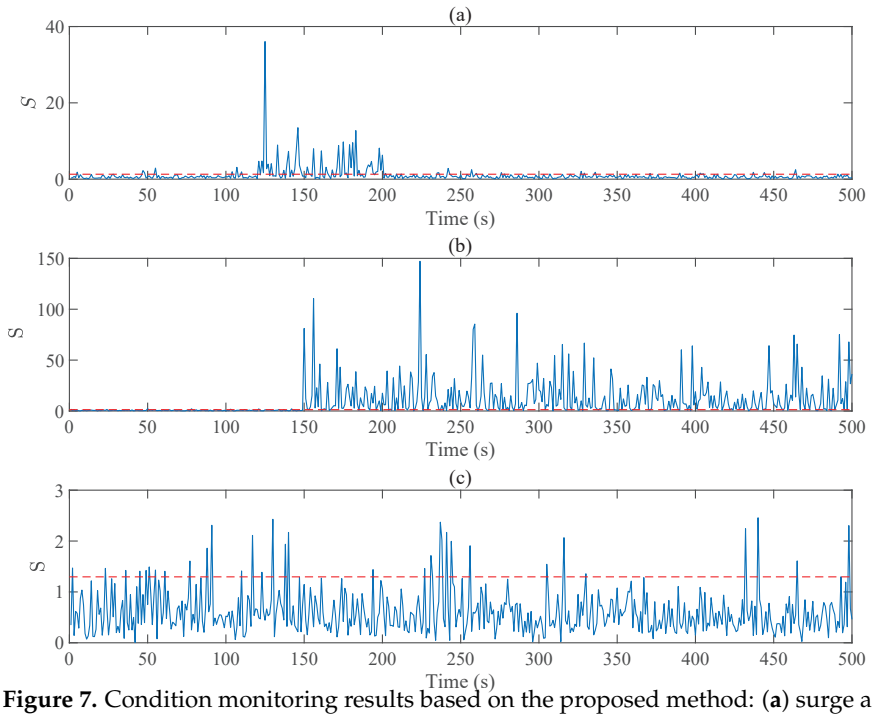

Figure 7. Condition monitoring results based on the proposed method: (**a**) surge attacks; (**b**) biased attacks; and (**c**) the normal conditions.

As shown in Figure 7, the red dashed line indicates the preset control limits, whereas the blue line represents the monitoring statistics calculated from the online data. The surge attack and deviation attack were performed at the 110th seconds of each experiment, as shown in Figure 7a and Figure 7b, respectively. In addition, Figure 7c shows the monitoring results under normal operating conditions. Based on the attacking records, the model successfully identified the impact of the step-wise and slowly varying deviations from the normal operating state. The experimental results revealed that the proposed method can effectively identify anomalies due to attacks with 92.31% accuracy and 12 s monitoring delay.

For greater clarity, the PCA-based process monitoring method was chosen to perform the comparative experiments as a monitoring strategy [32]. To realize the comparison, the integrated monitoring statistics of S_t, achieved by combining T^2 and Q, ere adopted in the attack detection task [33]. The control limit was set as $S_t = 7.9127$. It can be seen from Figure 8 that the PCA failed to detect the attacks because there was no significant change in the monitoring statistics. In both cases, the PCA method was less susceptible to the operational instability caused by assaults. As a result of the initial data structure being altered, the anomalies caused by data injection-type attacks did not rapidly accumulate and did not significantly affect the detection data. Consequently, the original PCA method was unable to extract the features that were related to operating conditions, resulting in unsatisfactory monitoring results. The monitoring process also suffered from more misses, false alarms, and longer anomaly detection delays than the method proposed in this study.

To effectively showcase the effectiveness of the proposed method in the monitoring processes, there were some sophisticated process monitoring methods that were selected for comparison such as the original PPCA [15] and mRMR-PCA [32]. The monitoring delay (μ) refers to the period between the incidence of attack performance and the detection of its reasons. Evaluating the performance monitoring involves assessing the non-detection rate (η) and false alarm rate (γ) according to specific criteria. The following matrix proves the definitions of the above indicators

$$\eta = \frac{n_{\mathrm{FP}}}{n_{\mathrm{TN}} + n_{\mathrm{FP}}} \times 100\%, \tag{27}$$

$$\gamma = \frac{n_{\mathrm{FN}}}{n_{\mathrm{TP}} + n_{\mathrm{FN}}} \times 100\%. \tag{28}$$

The variable n_{FN} represents the count of samples that are incorrectly classified into non-optimal modes when they should have been classified into optimal modes. The variable n_{TP} represents the count of samples that were correctly classified into optimal modes. The variable n_{FP} represents the count of samples that are incorrectly classified into optimal modes when they should have been classified into non-optimal modes. Lastly, the variable n_{TN} represents the count of samples that are correctly classified into non-optimal modes. Lower values for η and γ suggest a superior monitoring performance.

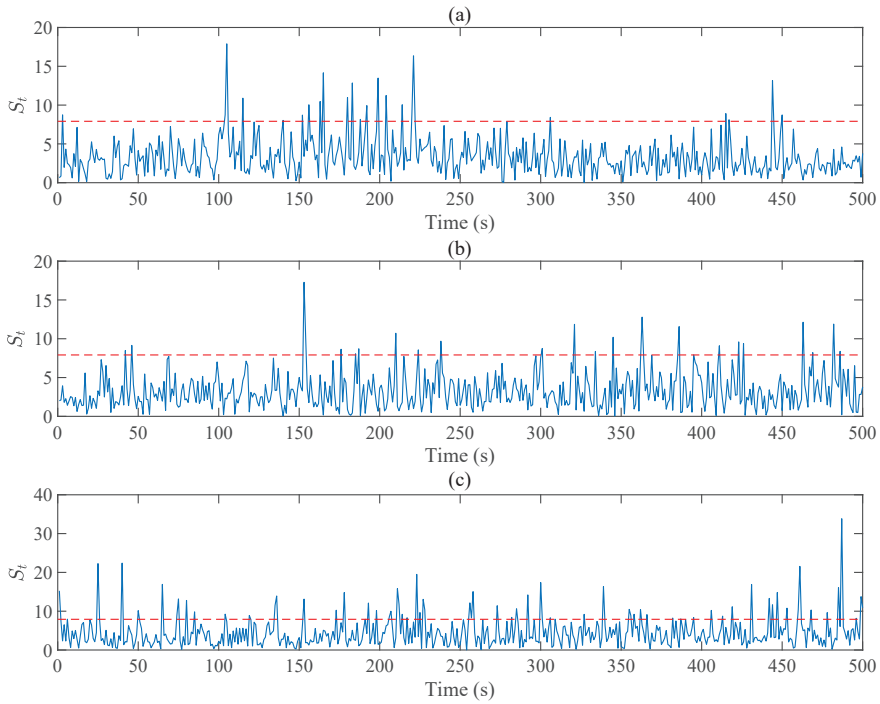

Figure 8. Condition monitoring results based on the PCA method: (**a**) surge attacks; (**b**) biased attacks; (**c**) the normal conditions.

The detection results of the different methods for monitoring data injection attacks are shown in Tables 2 and 3. It is essential to clarify that the typical PCA approach failed to detect both attacks because of its γ for the two statistics, which went up to 74.31% and 83.34%. The PPCA method is inadequate due to its failure to include the non-linear attributes of the data, rendering it unsuccessful in detecting abnormalities. The η of the Q-statistic calculated by mRMR-PCA was 6.05%, but it was 90.64% for the T^2-statistic in Case 1, which did not meet the needs of field applications. The mRNR-PCA based-monitoring method utilizes a distributed architecture, and, while it did not successfully identify the local attacks, its efficacy was attributed to the singular PCA model. The results show that our method has a comparatively better monitoring performance than the other methods.

In terms of statistical metrics, the maximum enhancement of η and γ reached 69.17 % and 9.67%, respectively, and the shortest detection delays of 11 s and 20 s were achieved in both cases.

Table 2. Attack detection for the different methods.

Type	Indexes	PCA [32]		PPCA [15]	mRMR-PCA [32]		Proposed Method
		T^2	Q	T_s^2	T^2	Q	S
Case 1	η (%)	74.31	21.43	15.2	90.64	6.05	5.14
	γ (%)	3.02	8.21	9.67	14.74	1.93	5.07
	μ (s)	64	42	35	-	40	11
Case 2	η (%)	72.16	83.34	5.14	75.13	5.84	7.76
	γ (%)	2.76	11.13	8.22	3.66	6.16	5.77
	μ (s)	-	48	20	-	104	21

Table 3. Attack detection results obtained with PCA, the original PPCA, mRMR-PCA, and the proposed method.

Type	Indexes	PCA [32]		PPCA [15]	mRMR-PCA [32]		Proposed Method
		T^2	Q	T_s^2	T^2	Q	S
Case 1	False Alarms	252	71	51	306	20	18
	Missed Alarms	10	27	31	48	7	17
	Accuracy(s)	5.33	10.08	77.65	4.41	80.33	90.35
Case 2	False Alarms	245	282	17	255	17	26
	Missed Alarms	7	37	27	10	21	20
	Accuracy(s)	5.21	4.87	83.55	5.67	85.86	94.3

In intuitive terms, the distributed structure ensures that the monitoring model can effectively extract the local and global features with finer-grained precision. In contrast, the typical correlation space combined with the data feature approach captures the latent data features of the ICSs and more accurately portrays the operational state of the process as a whole.

In summary, the proposed approach takes into account the relationship between variable spaces and residual spaces for online monitoring, whereas PCA just evaluates the interaction between variables. The findings suggest that an enhancement in performance monitoring can be achieved by partitioning the initial dataset using PPCA and CVA-based variable reconstruction.

5. Conclusions

This paper proposed a concurrent distributed ICS monitoring method for network attack detection using prior knowledge-based mutual information (MI) and canonical variate analysis with probabilistic principal component analysis (CVA-PPCA). While other centralized process monitoring methods treat all variables as a uniform modeling space, MI-based variable division is capable of probing the underlying local and global characteristics of ICSs comprehensively. Additionally, the CVA-PPCA method established in each sub-block can then more closely reflect and detect the external attack from different aspects.

Due to the complexity and variability of the stratum during geological drilling, as well as the randomness of the network attacks, it was necessary to improve the method's adaptability further by, for example, setting control limits and selecting monitoring windows. Aspects of anomaly tracing and small sample modeling are also important to consider for ICS security when dealing with unknown attack backgrounds. As the study progresses, it will be applied to a variety of industrial processes and recommendations will be provided in the decision-making phase. Further research will also focus on developing a monitoring scheme that takes into account the dynamic nature of variables.

Author Contributions: Conceptualization, M.X. and H.F.; methodology, H.F.; software, Z.J. and S.Y.; validation, M.X., Z.J., S.Y. and H.F.; formal analysis, Z.J.; investigation, S.Y.; resources, M.X.; data curation, H.F.; writing—original draft preparation, M.X., Z.J., S.Y. and H.F.; writing—review and editing, Z.J. and H.F.; visualization, S.Y.; supervision, M.X.; project administration, M.X.; funding acquisition, H.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National defense basic scientific research program (Grant JKCY2022211C007); National defense basic scientific research program (Grant JKCY2021206B104); the "CUG Scholar" Scientific Research Funds at the China University of Geosciences (Wuhan) (Project No. 2023095); the National Natural Science Foundation of China (Grant 6227021554); and the Fundamental Research Funds for the Central Universities, China University of Geosciences.

Data Availability Statement: Data are contained within the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper, and authors of this manuscript also have no conflicts of interest with any organization.

Abbreviations

The following abbreviations are used in this manuscript:

ICSs	Industrial Control Systems
MSPM	Multivariate Statistical Process Monitoring
CVA	Canonical Variate Analysis
PPCA	Probability Principal Component Analysis
ROP	Rate of Permeation
MI	Mutual Information
KDE	Kernel Density Estimate
ANOVA	One-Way Analysis Of Variance
CPV	Cumulative Percentage Variance

References

1. Lin, I.C.; Tseng, P.C.; Chang, Y.S.; Weng, T.C. IOTA Data Preservation Implementation for Industrial Automation and Control Systems. *Processes* **2023**, *11*, 2160. [CrossRef]
2. Gao, M.; Feng, D. Stochastic stability analysis of networked control systems with random cryptographic protection under random zero-measurement attacks. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 1098–1111. [CrossRef]
3. Jia, C.Q.; Feng, D.Q. Industrial control system devices security assessment with multi-objective decision. *Acta Autom. Sin.* **2016**, *42*, 706–714.
4. Teixeira, A.; Pérez, D.; Sandberg, H.; Johansson, K.H. Attack models and scenarios for networked control systems. In Proceedings of the 1st international conference on High Confidence Networked Systems, Beijing, China, 17–18 April 2012; pp. 55–64.
5. Adepu, S.; Mathur, A. Generalized attacker and attack models for cyber physical systems. In Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, USA, 10–14 June 2016; Volume 1, pp. 283–292.
6. Lu, G.; Feng, D.; Huang, B. Hidden Markov model-based attack detection for networked control systems subject to random packet dropouts. *IEEE Trans. Ind. Electron.* **2020**, *68*, 642–653. [CrossRef]
7. Lu, G.; Feng, D. Industrial control network security situation awareness based on improved C-SVC. *Control Decis.* **2017**, *32*, 1223–1228.
8. Chaouk, H.; Obeid, E.; Halwani, J.; Arayro, J.; Mezher, R.; Amine, S.; Gazo Hanna, E.; Mouhtady, O.; Younes, K. Application of Principal Component Analysis for the Elucidation of Operational Features for Pervaporation Desalination Performance of PVA-Based TFC Membrane. *Processes* **2024**, *12*, 1502. [CrossRef]
9. Huang, K.; Zhang, L.; Wu, D.; Yang, C.; Gui, W. Nonstationary industrial process monitoring based on stationary projective dictionary learning. *IEEE Trans. Control Syst. Technol.* **2022**, *31*, 1122–1132. [CrossRef]
10. Yang, C.; Zhang, J.; Wu, D.; Huang, K.; Gui, W. Variable partition based parallel dictionary learning for linearity and nonlinearity coexisting dynamic process monitoring. *Control Eng. Pract.* **2024**, *142*, 105750. [CrossRef]
11. Ji, H.; Hou, Q.; Wu, D. Modified performance-enhanced PCA for incipient fault detection of dynamic industrial processes. *J. Process Control* **2023**, *131*, 103107. [CrossRef]
12. Li, G.; Qin, S.J.; Ji, Y.; Zhou, D. Reconstruction based fault prognosis for continuous processes. *Control Eng. Pract.* **2010**, *18*, 1211–1219. [CrossRef]

13. Fan, H.; Lai, X.; Du, S.; Yu, W.; Lu, C.; Wu, M. Distributed monitoring with integrated probability PCA and mRMR for drilling processes. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–13. [CrossRef]
14. Kaib, M.T.H.; Kouadri, A.; Harkat, M.F.; Bensmail, A.; Mansouri, M. Improving kernel PCA-based algorithm for fault detection in nonlinear industrial process through fractal dimension. *Process Saf. Environ. Prot.* **2023**, *179*, 525–536. [CrossRef]
15. Zhang, J.; Chen, M.; Hong, X. Nonlinear process monitoring using a mixture of probabilistic PCA with clusterings. *Neurocomputing* **2021**, *458*, 319–326. [CrossRef]
16. Lu, C.; Zeng, J.; Dong, Y.; Xu, X. Streaming variational probabilistic principal component analysis for monitoring of nonstationary process. *J. Process Control* **2024**, *133*, 103134. [CrossRef]
17. Lou, S.; Wu, P.; Yang, C.; Xu, Y. Structured fault information-aided canonical variate analysis model for dynamic process monitoring. *J. Process Control* **2023**, *124*, 54–69. [CrossRef]
18. Li, L.; Dong, F.; Zhang, S. Manifold regularized deep canonical variate analysis with interpretable attribute guidance for three-phase flow process monitoring. *Expert Syst. Appl.* **2024**, *251*, 124015. [CrossRef]
19. Zhang, S.; Bao, X.; Wang, S. Common canonical variate analysis (CCVA) based modeling and monitoring for multimode processes. *Chem. Eng. Sci.* **2023**, *271*, 118581. [CrossRef]
20. Zhang, S.; Zhao, C.; Huang, B. Simultaneous static and dynamic analysis for fine-scale identification of process operation statuses. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5320–5329. [CrossRef]
21. Liu, Y.; Wang, F.; Gao, F.; Cui, H. Hierarchical multiblock T-PLS based operating performance assessment for plant-wide processes. *Ind. Eng. Chem. Res.* **2018**, *57*, 14617–14627. [CrossRef]
22. Zhu, J.; Ge, Z.; Song, Z. Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1877–1885. [CrossRef]
23. Hu, J.; Wu, M.; Cao, W.; Pedrycz, W. Soft-Sensing of Burn-Through Point Based on Weighted Kernel Just-in-Time Learning and Fuzzy Broad-Learning System in Sintering Process. *IEEE Trans. Ind. Inform.* **2024**, 20, 7316–7324. [CrossRef]
24. Zhang, J.; Chen, H.; Chen, S.; Hong, X. An improved mixture of probabilistic PCA for nonlinear data-driven process monitoring. *IEEE Trans. Cybern.* **2017**, *49*, 198–210. [CrossRef] [PubMed]
25. Zhong, K.; Ma, D.; Han, M. Distributed dynamic process monitoring based on dynamic slow feature analysis with minimal redundancy maximal relevance. *Control Eng. Pract.* **2020**, *104*, 104627. [CrossRef]
26. Melo, A.; Câmara, M.M.; Pinto, J.C. Data-Driven Process Monitoring and Fault Diagnosis: A Comprehensive Survey. *Processes* **2024**, *12*, 251. [CrossRef]
27. Xiao, B.; Li, Y.; Sun, B.; Yang, C.; Huang, K.; Zhu, H. Decentralized PCA modeling based on relevance and redundancy variable selection and its application to large-scale dynamic process monitoring. *Process Saf. Environ. Prot.* **2021**, *151*, 85–100. [CrossRef]
28. Kim, D.; Lee, I.B. Process monitoring based on probabilistic PCA. *Chemom. Intell. Lab. Syst.* **2003**, *67*, 109–123. [CrossRef]
29. Chen, X.; Zheng, J.; Zhao, C.; Wu, M. Full decoupling high-order dynamic mode decomposition for advanced static and dynamic synergetic fault detection and isolation. *IEEE Trans. Autom. Sci. Eng.* **2022**, *21*, 226–240. [CrossRef]
30. Moudoud, H.; Mlika, Z.; Khoukhi, L.; Cherkaoui, S. Detection and prediction of fdi attacks in iot systems via hidden markov model. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 2978–2990. [CrossRef]
31. Wang, S.; Wang, Y.; Yang, B.; Mo, F.; Zhang, Z. Variational Bayesian Learning with reliable likelihood approximation for accurate Process Quality Evaluation. *IEEE Trans. Ind. Inform.* **2023**, *20*, 815–823. [CrossRef]
32. Xu, C.; Zhao, S.; Liu, F. Distributed plant-wide process monitoring based on PCA with minimal redundancy maximal relevance. *Chemom. Intell. Lab. Syst.* **2017**, *169*, 53–63. [CrossRef]
33. Yue, H.H.; Qin, S.J. Reconstruction-based fault identification using a combined index. *Ind. Eng. Chem. Res.* **2001**, *40*, 4403–4414. [CrossRef]