Article

A Process Monitoring Method Based on Dynamic Autoregressive Latent Variable Model and Its Application in the Sintering Process of Ternary Cathode Materials

Ning Chen [1], Fuhai Hu [1], Jiayao Chen [1,*], Zhiwen Chen [1,2,*], Weihua Gui [1] and Xu Li [3]

[1] School of Automation, Central South University, Changsha 410083, China; ningchen@csu.edu.cn (N.C.); FuhaiHu@csu.edu.cn (F.H.); gwh@csu.edu.cn (W.G.)

[2] State Key Laboratory of High Performance Complex Manufacturing, Central South University, Changsha 410083, China

[3] Hunan Shanshan Energy Technology Co., Ltd., Changsha 410205, China; li.bo@shanshanenergy.com

* Correspondence: jychen@csu.edu.cn (J.C.); zhiwen.chen@csu.edu.cn (Z.C.)

Abstract: Due to the ubiquitous dynamics of industrial processes, the variable time lag raises great challenge to the high-precision industrial process monitoring. To this end, a process monitoring method based on the dynamic autoregressive latent variable model is proposed in this paper. First, from the perspective of process data, a dynamic autoregressive latent variable model (DALM) with process variables as input and quality variables as output is constructed to adapt to the variable time lag characteristic. In addition, a fusion Bayesian filtering, smoothing and expectation maximization algorithm is used to identify model parameters. Then, the process monitoring method based on DALM is constructed, in which the process data are filtered online to obtain the latent space distribution of the current state, and T^2 statistics are constructed. Finally, by comparing with an existing method, the feasibility and effectiveness of the proposed method is tested on the sintering process of ternary cathode materials. Detailed comparisons show the superiority of the proposed method.

1. Introduction

To ensure production safety and product quality, process monitoring technology has become an indispensable ingredient for industrial processes in recent years. It is commonly divided into model-based methods and data-driven methods. Compared with the former ones, the later ones can take advantage of the routine measurement and do not rely on process prior knowledge and precise mechanism models, which are unavailable or cost-intensive to obtained at times [1,2]. Therefore, they are widely used in modern industrial process.

During the past decades, many data-driven process monitoring methods have been published [3–6]. Kim et al. [7] proposed a probabilistic PCA to monitoring industrial processes, which firstly extracts redundant information from the variables and constructs feature distribution for monitoring, but it only extracts features of the input space. Zhao et al. [8] proposed the probabilistic PLSR process monitoring method to monitor quality-related faults, which can simultaneously consider the fault characteristics of the input and output spaces for monitoring. Furthermore, Chen et al. [9] proposed a probability-related PCA method for detecting incipient faults, which can greatly improve the detection ability of minor faults. Probabilistic framework modeling can overcome process noise [10]. However, the process monitoring methods currently proposed are all static methods, and the actual production processes are dynamical featured with variable time lag [11,12].

Process dynamics could refer to the mutual influence before and after the current sampling [13]. To deal with the dynamics of process, Ku et al. [14] built an augmented

matrix and extended the static PCA model to the dynamic PCA (DPCA) for process monitoring. However, the introduction of augmented matrix increases the parameter dimensions called the curse of dimensionality [15]. Motivated by DPCA, Li et al. [16] proposed a dynamic latent variable model for monitoring the Tennessee Eastman process. In this model, the autoregressive model is used to extract data dynamic information, and PCA is performed to reduce redundancy between variables. It divides variable order reduction and dynamic information extraction into two stages, which makes the system complex and not easy to tune. In addition, compared with process variables, quality variables also contain useful fault information [17]. For this reason, Ge et al. [18] proposed a supervised linear dynamic system model process monitoring method. This method uses a first-order autoregressive equation to simulate the first-order dynamic [19] but does not take the variable time lag into account.

Variable time lag refers to the delay between the effects of variables [20]. The existing monitoring methods considering time lag are usually divided into two categories [21]. One is to find the time lag between variables and translate the data to eliminate the time lag and then establish a static process monitoring model for the processed data. For example, Wang et al. [22] proposed a spatial reconstruction method to identify system time lag, then aligned the data and established a monitoring model, but the alignment operation will destroy the data structure and cause data loss. The other idea is to use time lag as an unknown parameter of the process monitoring model and identify the parameters through a data-driven method. For example, Huber et al. [23] proposed to take the time lag as a parameter of a high-order state space system model and then solve it uniformly with the model parameters, but this method relies on the setting of the time lag parameter and the parameter identification method.

From the above discussions, it can be observed that the variable time lag characteristic of a process makes the previous work unfavorable. However, this characteristic is common in industrial processes [24,25]. To deal with this problem, this paper proposes a process monitoring method based on a dynamic autoregressive latent variable model. Firstly, from the data point of view, a linear dynamic model is constructed between process variables and quality variables, and the dynamic information of process input and process output is compressed to latent variables, and then a dynamic autoregressive latent variable model (DALM) is constructed for latent variables to extract variable time lag information. In addition, a fusion Bayesian filtering, smoothing and expectation maximization algorithm is used to identify model parameters. Then, the DALM is applied to the industrial monitoring process. The process variables are filtered through improved Bayesian filtering technology to obtain the latent space distribution of the current state, and the T^2 statistics of the latent space are constructed and monitored [26] to realize the process monitoring task. The main contribution can be concluded as (1) a process monitoring method based on dynamic autoregressive latent variable model is proposed in this paper; (2) a dynamic autoregressive latent variable model (DALM) is developed to extract variable time lag information; (3) a fusion Bayesian filtering, smoothing and expectation maximization algorithm is improved to identify model parameters; (4) based on the DALM, the T^2 statistics of the latent space are constructed to realize the process monitoring task.

The main structure of the paper is arranged as follows. In the second section, a dynamic autoregressive latent variable model is proposed, and the parameter identification algorithm of the model is derived in detail. A process monitoring method based on DALM is proposed in the third chapter. The fourth section uses the monitoring method to monitor the sintering process of the ternary cathode material to verify the monitoring performance of the proposed method. Finally, the last section concludes.

2. Modeling Method Based on Dynamic Autoregressive Latent Variable Model

This section proposes a dynamic autoregressive latent variable modeling method for the accurate modeling of dynamical industrial processes. Bayesian filtering and smoothing

inference were used to obtain the spatial distribution of latent variables, and the parameters of the model were identified by combining with the EM algorithm.

2.1. Dynamic Autoregressive Latent Variable Model Structure

In order to consider the dynamic characteristics of the process, the traditional probability latent variable model [27] establishes the relationship between the current moment and the previous moment data, as shown in (1).

$$
\begin{aligned}
\mathbf{z}_t &= \mathbf{A}\mathbf{z}_{t-1} + \boldsymbol{\eta}_t^{\mathbf{z}}, \\
\mathbf{x}_t &= \mathbf{B}_{\mathbf{x}}\mathbf{z}_t + \boldsymbol{\eta}_t^{\mathbf{x}}, \\
\mathbf{y}_t &= \mathbf{B}_{\mathbf{y}}\mathbf{z}_t + \boldsymbol{\eta}_t^{\mathbf{y}},
\end{aligned}
\tag{1}
$$

where the structure consists of a linear Gaussian dynamic equation and two linear Gaussian observation equations, $\mathbf{z}_t$ is the latent variable of the process state at time t; $\mathbf{x}_t$ and $\mathbf{y}_t$ are the process variable and quality variable at time t, respectively; $\mathbf{A}$, $\mathbf{B}_{\mathbf{x}}$ and $\mathbf{B}_{\mathbf{y}}$ are their own load matrix. The Gaussian dynamic equation is used to describe the dynamic relationship of the process data. The observation equation compresses the information of the process data into low-dimensional latent variables. Therefore, an accurate mathematical model can be established for the dynamic process, but the structure does not consider the time lag characteristics of the process.

In order to further consider the characteristics of process time lag, on the basis of dynamic probabilistic latent variable model (DPLVM) [13], the trend similarity analysis algorithm [22] was first used to obtain the time lag coefficient L of the current process, and then an autoregressive equation was constructed for the latent variables to describe the variable time lag information. Among them, the autoregressive equation models the process dynamics and time lag characteristics, and the linear observation equation models the cross-correlation of data. The probability graph model of the model is shown in Figure 1, and the mathematical expression is shown in (2).

$$
\begin{aligned}
\mathbf{z}_t &= \mathbf{A}\mathbf{h}_{t-1} + \boldsymbol{\eta}_t^{\mathbf{z}}, \\
\mathbf{x}_t &= \mathbf{B}_{\mathbf{x}}\mathbf{z}_t + \boldsymbol{\eta}_t^{\mathbf{x}}, \\
\mathbf{y}_t &= \mathbf{B}_{\mathbf{y}}\mathbf{z}_t + \boldsymbol{\eta}_t^{\mathbf{y}},
\end{aligned}
\tag{2}
$$

where $\mathbf{z}_t \in R^d$ represents the latent variable of the process state at time t, $\mathbf{h}_{t-1} = \begin{bmatrix} \mathbf{z}_{t-1}^T & \mathbf{z}_{t-2}^T & \cdots & \mathbf{z}_{t-L}^T \end{bmatrix}^T \in R^{dL}$ is the augmented state variable containing the latent variables at time L in the past, $\mathbf{x}_t \in R^v$ is the observed value of the process variable at time t, $\mathbf{y}_t \in R^k$ is the observed value of the quality variable at time t and d, v, k, respectively, correspond to the dimensions of latent variables, process variables and quality variables. $\mathbf{A} \in R^{d \times dL}$ is the state transition matrix, L is the time lag value of the process, and $\mathbf{B}_{\mathbf{x}} \in R^{v \times d}$, $\mathbf{B}_{\mathbf{y}} \in R^{k \times d}$ are the state divergence matrices. $\boldsymbol{\eta}_t^{\mathbf{z}} \in R^d$, $\boldsymbol{\eta}_t^{\mathbf{x}} \in R^v$, and $\boldsymbol{\eta}_t^{\mathbf{y}} \in R^k$ are Gaussian noise terms of latent variables, process variables and quality variables, respectively. Assuming that the noises are independent of each other, the distributions obeyed are $\boldsymbol{\eta}_t^{\mathbf{z}} \sim N(0, \Sigma_{\mathbf{z}})$, $\boldsymbol{\eta}_t^{\mathbf{x}} \sim N(0, \Sigma_{\mathbf{x}})$ and $\boldsymbol{\eta}_t^{\mathbf{y}} \sim N(0, \Sigma_{\mathbf{y}})$, respectively. Latent variables represent the current state of the process, and this model is an extension of the traditional DPLVM.
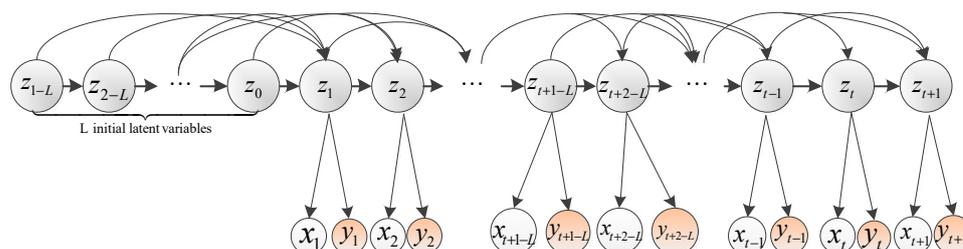

Figure 1. Probability graph model of dynamic autoregressive latent variable model.

2.2. Parameter Identification Based on EM Algorithm

Since only $\mathbf{x}_t$ and $\mathbf{y}_t$ can be observed in the process, and latent variables are abstracted to describe the state of the process and are unobservable, the EM algorithm was used to identify the parameters of the model [28]. Each iteration of the EM algorithm consisted of two steps: E step, seeking expectation (exception); M step, seeking maximization (maximization). This section uses Bayesian filtering and smoothing to infer the spatial distribution of latent variables, so as to solve the difficult problem of calculating latent variable statistics.

Under the framework of probability, the model assumed that the latent variables at the initial moment obeyed a Gaussian distribution with mean $\mathbf{u}_0$ and variance $\mathbf{V}_0$, that is, $\mathbf{z}_0, \mathbf{z}_{-1}, \cdots, \mathbf{z}_{1-L} \sim N(\mathbf{u}_0, \mathbf{V}_0)$. From the knowledge of probability theory [16], it is easy to get that the distribution of the latent variable $\mathbf{z}_t$, the process variable $\mathbf{x}_t$ and the quality variable $\mathbf{y}_t$ obey the Gaussian distribution, as shown in (3).

$$
\begin{aligned}
\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \cdots, \mathbf{z}_{t-L} &\sim N(\mathbf{A}_1 \mathbf{z}_{t-1} + \mathbf{A}_2 \mathbf{z}_{t-2} + \cdots + \mathbf{A}_L \mathbf{z}_{t-L}, \Sigma_{\mathbf{z}}), \\
\mathbf{x}_t | \mathbf{z}_t &\sim N(\mathbf{B}_{\mathbf{x}} \mathbf{z}_t, \Sigma_{\mathbf{x}}), \\
\mathbf{y}_t | \mathbf{z}_t &\sim N(\mathbf{B}_{\mathbf{y}} \mathbf{z}_t, \Sigma_{\mathbf{y}}).
\end{aligned}
\tag{3}
$$

The parameters that needed to be identified were denoted as $\mathbf{\Theta} = \{\mathbf{A}, \mathbf{B}_{\mathbf{x}}, \mathbf{B}_{\mathbf{y}}, \mathbf{u}_0, \mathbf{V}_0, \Sigma_{\mathbf{z}}, \Sigma_{\mathbf{x}}, \Sigma_{\mathbf{y}}\}$, of which $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_L]$. According to the naive what you see is what you get thought, the parameter identification problem was transformed into the maximum observation data $\mathbf{x}_{1:T}$, $\mathbf{y}_{1:T}$. The log-likelihood function on the parameter $\mathbf{\Theta}$ is shown in (4), where $\mathbf{x}_{1:T}$ represents the observation sequence of the process variable $\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots$, $\mathbf{x}_T$, $\mathbf{y}_{1:T}$ represents the observation sequence $\mathbf{y}_1$, $\mathbf{y}_2$, $\cdots$, $\mathbf{y}_T$ of the quality variable, where T is the total number of training samples, namely,

$$
\mathbf{\Theta}^{new} = \arg \max_{\mathbf{\Theta}} \log P(\mathbf{x}_{1:T}, \mathbf{y}_{1:T} | \mathbf{\Theta}).
\tag{4}
$$

The EM algorithm [29] was used to solve the optimization problem of Equation (4). In the E step of the EM algorithm, the log-likelihood function $\log P(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}_{1-L:T} | \mathbf{\Theta})$ of the complete data had to be calculated with respect to the conditional expectation of the latent variable $\mathbf{z}_{1-L:T}$ to obtain the objective cost function (Q function), as shown in (5).

$$
Q\left(\mathbf{\Theta} | \mathbf{\Theta}^{old}\right) = E_{\mathbf{z}_{1-L:T} | (\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{\Theta}^{old})} \{\log P(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}_{1-L:T} | \mathbf{\Theta})\}.
\tag{5}
$$

Actually, the likelihood function can be formulated by the application of the product rule of probability. From the model structure, the log-likelihood function of the complete data was expanded, as expressed by (6).

$$
\begin{aligned}
&\log P\left(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}_{(-L+1):T} | \mathbf{\Theta}\right) \\
&= \log\left\{ P(\mathbf{z}_0, \mathbf{z}_{-1}, \cdots, \mathbf{z}_{-L+1}) \prod_{t=1}^{T} P(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \cdots, \mathbf{z}_{t-L}) P(\mathbf{x}_t | \mathbf{z}_t) P(\mathbf{y}_t | \mathbf{z}_t) \right\} \\
&= \log P(\mathbf{z}_0, \mathbf{z}_{-1}, \cdots, \mathbf{z}_{-L+1}) + \sum_{t=1}^{T} \log P(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \cdots, \mathbf{z}_{t-L}) + \sum_{t=1}^{T} \log P(\mathbf{x}_t | \mathbf{z}_t) \\
&+ \sum_{t=1}^{T} \log P(\mathbf{y}_t | \mathbf{z}_t).
\end{aligned}
\tag{6}
$$

For clear writing, we denote $E_{\mathbf{z}_T}(f(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)) = E_{\mathbf{z}_{1-L:T} | (\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{\Theta}^{old})}(f(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t))$, then the expectation of the complete data likelihood function $\log P(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}_{(-L+1):T} | \mathbf{\Theta})$ with respect to the latent variable distribution $P(\mathbf{z}_{(-L+1):T} | \mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{\Theta}^{old})$ is shown in (7).

$$Q\left(\Theta|\Theta^{old}\right) = E_{\mathbf{z}_T}\{\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}_{1-L:T}|\Theta)\}$$

$$= -\frac{1}{2}\left\{\log|\mathbf{V}_0| + E_{\mathbf{z}_T}\left(\begin{bmatrix}\mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1}\end{bmatrix}^T \mathbf{V}_0^{-1}\begin{bmatrix}\mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1}\end{bmatrix}\right) - 2E_{\mathbf{z}_T}\left(\begin{bmatrix}\mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1}\end{bmatrix}^T\right)\mathbf{V}_0^{-1}\mathbf{u}_0 + \mathbf{u}_0^T\mathbf{V}_0^{-1}\mathbf{u}_0\right\}$$

$$-\frac{1}{2}\left\{T\log|\Sigma_{\mathbf{z}}| + \sum_{t=1}^{T}\left\{E_{\mathbf{z}_T}\left(\mathbf{z}_t^T\Sigma_{\mathbf{z}}^{-1}\mathbf{z}_t\right) - 2E_{\mathbf{z}_T}\left(\begin{bmatrix}\mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L}\end{bmatrix}^T\mathbf{A}^T\Sigma_{\mathbf{z}}^{-1}\mathbf{z}_t\right) + E_{\mathbf{z}_T}\left(\begin{bmatrix}\mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L}\end{bmatrix}^T\mathbf{A}^T\Sigma_{\mathbf{z}}^{-1}\mathbf{A}\begin{bmatrix}\mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L}\end{bmatrix}\right)\right\}\right\} \quad (7)$$

$$-\frac{1}{2}\left\{T\log|\Sigma_{\mathbf{x}}| + \sum_{t=1}^{T}\left\{\mathbf{x}_t^T\Sigma_{\mathbf{x}}^{-1}\mathbf{x}_t - 2E_{\mathbf{z}_T}\left(\mathbf{z}_t^T\right)\mathbf{B}_{\mathbf{x}}^T\Sigma_{\mathbf{x}}^{-1}\mathbf{x}_t + E_{\mathbf{z}_T}\left(\mathbf{z}_t^T\mathbf{B}_{\mathbf{x}}^T\Sigma_{\mathbf{x}}^{-1}\mathbf{B}_{\mathbf{x}}\mathbf{z}_t\right)\right\}\right\}$$

$$-\frac{1}{2}\left\{T\log|\Sigma_{\mathbf{y}}| + \sum_{t=1}^{T}\left\{\mathbf{y}_t^T\Sigma_{\mathbf{y}}^{-1}\mathbf{y}_t - 2E_{\mathbf{z}_T}\left(\mathbf{z}_t^T\right)\mathbf{B}_{\mathbf{y}}^T\Sigma_{\mathbf{y}}^{-1}\mathbf{y}_t + E_{\mathbf{z}_T}\left(\mathbf{z}_t^T\mathbf{B}_{\mathbf{y}}^T\Sigma_{\mathbf{y}}^{-1}\mathbf{B}_{\mathbf{y}}\mathbf{z}_t\right)\right\}\right\} + cons\tan t.$$

Appendix A provides a detailed update of all parameters at step M. From (7), the related statistics of latent variables in the Q function include $E_{\mathbf{z}_T}(\mathbf{z}_t)$, $E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_t^T\right)$ and $E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_{t-i}^T\right)$, where $t = 0, 1, \cdots, T$, $i = 1, 2, \cdots, L$, in fact, these statistics can be passed. The posterior probability distribution of the latent variables obtained in the E step of the EM algorithm was obtained. The calculation results of these statistics are shown in (8). The detailed derivation process is shown in Appendix B.

$$\begin{cases} E_{\mathbf{z}_T}(\mathbf{z}_t) = E\left(\mathbf{z}_t|\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \Theta^{old}\right) = \mathbf{m}_t^1 \\ E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_t^T\right) = E\left(\mathbf{z}_t\mathbf{z}_t^T|\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \Theta^{old}\right) = \mathbf{M}_t^{11} + \mathbf{m}_t^1\left(\mathbf{m}_t^1\right)^T \\ E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_{t-i}^T\right) = E\left(\mathbf{z}_t\mathbf{z}_{t-i}^T|\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \Theta^{old}\right) = \mathbf{M}_t^{1(i+1)} + \mathbf{m}_t^1\left(\mathbf{m}_t^{(i+1)}\right)^T \\ E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_{t-L}^T\right) = E\left(\mathbf{z}_t\mathbf{z}_{t-L}^T|\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \Theta^{old}\right) = \sum_{i=1}^{L}\mathbf{A}_i\left(\mathbf{M}_{t-1}^{iL} + \mathbf{m}_{t-1}^i\left(\mathbf{m}_{t-1}^L\right)^T\right) \end{cases}, \quad (8)$$

where $\mathbf{m}_t$ and $\mathbf{M}_t$ are the mean and covariance of the posterior probability distribution of the latent variable. Therefore, through E step and M step iterative update until the parameters converge, the optimized parameter set $\Theta^{opt} = \{\mathbf{A}, \mathbf{B}_{\mathbf{x}}, \mathbf{B}_{\mathbf{y}}, \mathbf{u}_0, \mathbf{V}_0, \Sigma_{\mathbf{z}}, \Sigma_{\mathbf{x}}, \Sigma_{\mathbf{y}}\}$ can be obtained.

3. Process Monitoring Method Based on Dynamic Autoregressive Latent Variable Model

In this section, the established dynamic autoregressive latent variable model is used for industrial process monitoring. At first, DALM was used to model the process data so that the current state information was reflected in the latent variables, and then the latent space at the current time was obtained by filtering the process data distribution, constructing statistics and monitoring them. Let us introduce the monitoring process in detail below.

Although the latent space was unobservable, the establishment of a data-driven DALM model based on the characteristics of the process data extracted the information of the process variables to the spatial distribution of the latent variables. The process input $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$ and output $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]$ needed to be pre-processed by the normalization method, as shown in (9).

$$\begin{aligned} \mathbf{X}^q &= (\mathbf{x} - \mathbf{u}_{\mathbf{x}})\cdot\mathbf{std}_{\mathbf{x}}^{-1}, \\ \mathbf{Y}^q &= (\mathbf{y} - \mathbf{u}_{\mathbf{y}})\cdot\mathbf{std}_{\mathbf{y}}^{-1}, \end{aligned} \quad (9)$$

where $\mathbf{u}_{\mathbf{x}}$ and $\mathbf{u}_{\mathbf{y}}$ are the means of the variables $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{std}_{\mathbf{x}}$ and $\mathbf{std}_{\mathbf{y}}$ are the variances of the variables $\mathbf{X}$ and $\mathbf{Y}$. Preprocessed data were filtered through the filtering algorithm to obtain the spatial distribution of the latent variables, as shown in (10).

$$\mathbf{z}_t^q, \mathbf{z}_{t-1}^q, \cdots, \mathbf{z}_{t-L+1}^q|\mathbf{x}_{1:t}^q, \mathbf{y}_{1:t}^q \sim N(\mathbf{u}_t^q, \mathbf{V}_t^q) = N\left(\begin{bmatrix}\mathbf{u}_t^{1q} \\ \vdots \\ \mathbf{u}_t^{Lq}\end{bmatrix}, \begin{bmatrix}\mathbf{V}_t^{11q} & \cdots & \mathbf{V}_t^{1Lq} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_t^{(L)1q} & \cdots & \mathbf{V}_t^{LLq}\end{bmatrix}\right). \quad (10)$$

Among them, $\mathbf{u}_t{}^{1q}$ and $\mathbf{V}_t{}^{11q}$ are the mean and variance of the latent space distribution, respectively, which were obtained from (11). The detailed derivation process is shown in (A12)–(A16).

$$\mathbf{u}_t{}^{1q} = \mathbf{g}_t^1 + \begin{bmatrix} \mathbf{G}_t^{1(L+1)} & \mathbf{G}_t^{1(L+2)} \end{bmatrix} \begin{bmatrix} \mathbf{G}_t^{(L+1)(L+1)} & \mathbf{G}_t^{(L+1)(L+2)} \\ \mathbf{G}_t^{(L+2)(L+1)} & \mathbf{G}_t^{(L+2)(L+2)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t - \mathbf{g}_t^{L+1} \\ \mathbf{y}_t - \mathbf{g}_t^{L+2} \end{bmatrix},$$

$$\mathbf{V}_t{}^{11q} = \begin{bmatrix} \mathbf{G}_t^{11} & \cdots & \mathbf{G}_t^{1L} \end{bmatrix} - \begin{bmatrix} \mathbf{G}_t^{1(L+1)} & \mathbf{G}_t^{1(L+2)} \end{bmatrix} \begin{bmatrix} \mathbf{G}_t^{(L+1)(L+1)} & \mathbf{G}_t^{(L+1)(L+2)} \\ \mathbf{G}_t^{(L+2)(L+1)} & \mathbf{G}_t^{(L+2)(L+2)} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{G}_t^{1(L+1)} & \mathbf{G}_t^{1(L+2)} \\ \vdots & \vdots \\ \mathbf{G}_t^{L(L+1)} & \mathbf{G}_t^{L(L+2)} \end{bmatrix}^T. \tag{11}$$

It can be seen from (A24) that the information of the data $\mathbf{X}_t{}^q = [\mathbf{x}_1^q, \mathbf{x}_2^q, \cdots, \mathbf{x}_t^q]$ and $\mathbf{Y}_t{}^q = [\mathbf{y}_1^q, \mathbf{y}_2^q, \cdots, \mathbf{y}_t^q]$ at the current and previous moments was filtered into the current latent variable, and the latent variable distribution at the current moment is shown in (12).

$$\mathbf{z}_t{}^q | \mathbf{x}_{1:t}{}^q, \mathbf{y}_{1:t}{}^q \sim N(\mathbf{u}_t^{1q}, \mathbf{V}_t^{11q}). \tag{12}$$

Because the latent space contains the current state of the process dynamics and variable time lag information, the process statistic T^2 was constructed for the current latent variable at time t, as shown in (13).

$$T_{t,q}^2 = E(\mathbf{z}_t{}^q | \mathbf{x}_{1:t}{}^q, \mathbf{y}_{1:t}{}^q)^T covariance(\mathbf{z}_t{}^q | \mathbf{x}_{1:t}{}^q, \mathbf{y}_{1:t}{}^q)^{-1} E(\mathbf{z}_t{}^q | \mathbf{x}_{1:t}{}^q, \mathbf{y}_{1:t}{}^q). \tag{13}$$

Among them, the mathematical expectation and variance of the latent variables on the observation data at the current moment are shown in (14).

$$\begin{aligned} E(\mathbf{z}_t{}^q | \mathbf{x}_{1:t}{}^q, \mathbf{y}_{1:t}{}^q) &= \mathbf{u}_t^{1q}, \\ covariance(\mathbf{z}_t{}^q | \mathbf{x}_{1:t}{}^q, \mathbf{y}_{1:t}{}^q) &= \mathbf{V}_t^{11q}. \end{aligned} \tag{14}$$

The probability of the latent variable obeyed the Gaussian distribution. Therefore, according to the definition of chi-square distribution, this statistic obeyed the chi-square distribution $\chi_\alpha^2(d)$ after data preprocessing. Then, combining to the latent variable dimension d of the model and the significance level α required by the industry, the control threshold $T_{\lim}^2$ of the process monitoring method was obtained, and then the statistics of each time data were calculated online and compared with the control threshold, to determine whether the process deviated from the normal state. The process monitoring logic is determined by (15).

$$T_{t,q}^2 < T_{\lim}^2 = \chi_\alpha^2(d). \tag{15}$$

Too large an α value will lead to a high false alarm rate, and too low an α will lead to a high false alarm rate; therefore, in practice, it is a balance between false alarms and missed alarms. This paper chose α as 0.01, which means that the false positive rate of normal data was 0.01. If $T_{t,q}^2 < T_{\lim}^2$, the system was in a normal state. Otherwise, the process located in a fault state, and further diagnosis and identification of the fault was required for process maintenance. The process of DALM modeling and online process monitoring is shown in Figure 2.

The main steps of the process monitoring method based on the DALM model were as follows:

Step 1: Collect process data, divide the training and test data sets and standardize them.

Step 2: Use the training data set to learn the parameters of the DALM model.

Step 3: Build the model and determine the control threshold.

Step 4: Filter the process data online to get the latent space distribution at the current moment.

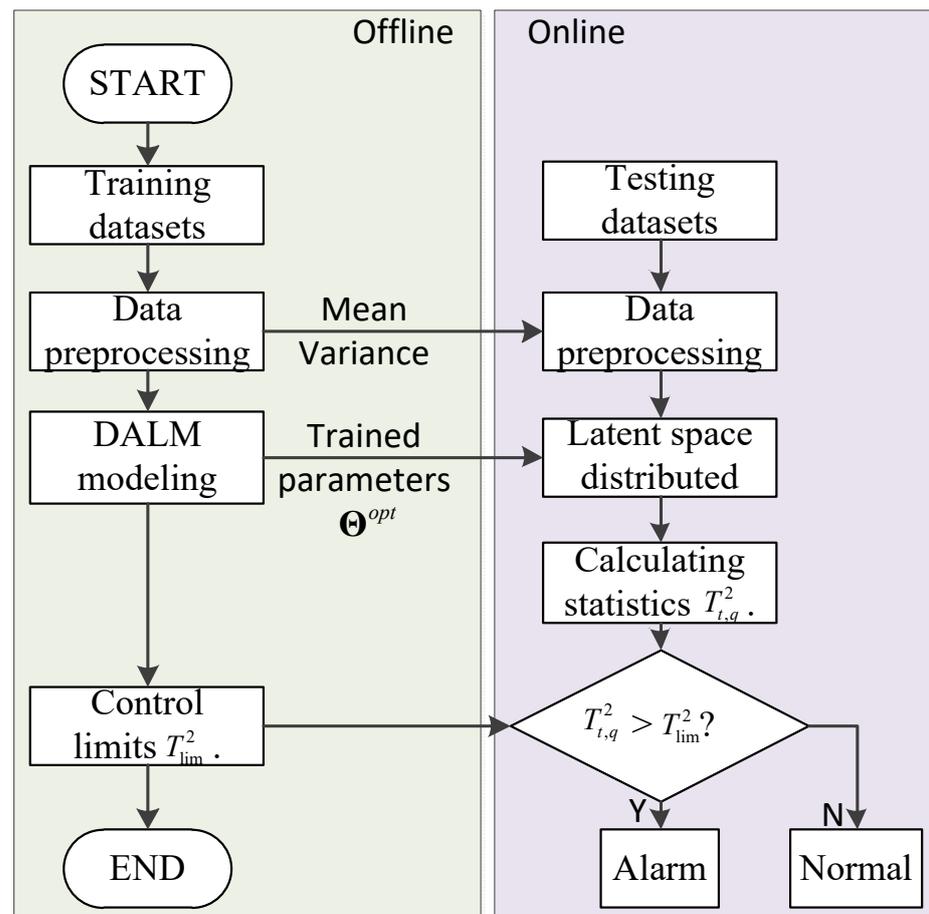Step 5: Calculate statistics and compare with the control threshold to determine whether the process is abnormal.

Figure 2. Flowchart based on DALM process monitor.

4. Case Study on the Sintering Process of Ternary Cathode Materials

In this section, the proposed process monitoring method based on the dynamic autoregressive latent variable model is used to monitor the sintering process of ternary cathode materials to verify the effectiveness of the method. First the sintering process technology of the ternary cathode material was introduced, then the model structure and parameter determination were introduced in detail, and finally the performance of the model was evaluated.

4.1. Introduction to the Sintering Process of Ternary Cathode Materials

The rapid development of the new energy industry has led to an extremely urgent demand for high-quality ternary cathode materials, and the sintering process of battery materials is the core and key process of battery preparation. This process consists of a series connection of a heating section, a constant temperature section and a cooling section, as shown in Figure 3. The optimal production state of a single temperature section cannot guarantee that the product performance indicators of the entire sintering process are within the optimal range; at the same time, changes in the sintering process, such as environmental humidity or temperature, also affect the stability of product performance indicators. In order to ensure the stability of product performance indicators as much as possible, while reducing energy consumption and material consumption, it is necessary to adjust the sintering parameters of the kiln according to the sintering state in real time, which leads to many variables in each temperature zone and series coupling, which makes the process data present complex process characteristics [29].
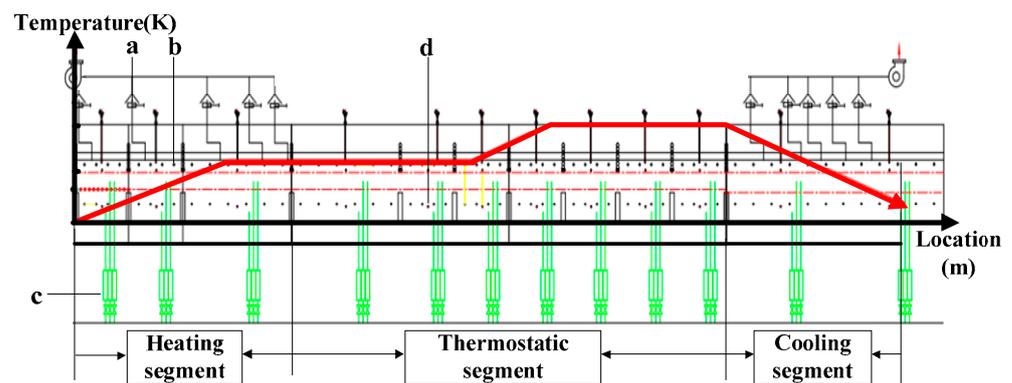
Figure 3. Structure diagram of sintering furnace for battery sintering.

The temperature field in the sintering process has a significant effect on the material properties. Over-firing will cause changes in the material morphology and internal structure, and under firing will not provide sufficient activation energy for chemical reactions. However, the decomposition reaction that occurs in the heating section is an endothermic process and requires sufficient heat supply, otherwise a reverse reaction will occur, resulting in inefficient water removal, which will affect the subsequent oxidation reaction. Therefore, the state of the heating section is very important to the sintering process. At the same time, the residual lithium content can directly reflect the quality of the product. In order to monitor the process status in real time, a monitoring model is established for the temperature and residual lithium content of the heating section.

Huang et al. [30] established a temperature field monitoring model based on the PBF equipment equation to monitor the dynamic sintering process of parts, but this method requires precise grinding tool structure parameters and can only monitor uniformly distributed temperature fields. Egorova et al. [31] tried to combine neural networks and PCA diagnosis method monitor and diagnose the sintering process. This method can locate the fault and diagnose the cause of the fault. However, the introduction of neural networks increases the time and space complexity of the system and ignores the system dynamic and time lag problems.

Due to the severe temperature interval coupling, the process variables exhibit complex characteristics, making the traditional static monitoring methods unable to achieve accurate monitoring results. The dynamic autoregressive latent variable model proposed in this section considers the dynamic and time lag information of the process at the same time, so it is more in line with the sintering process.

4.2. Determination of Model Parameters

This section establishes a monitoring model for the temperature and product quality in the heating section of the sintering process. The heating section contained seven temperature zones, and each temperature zone had two upper and lower temperature measuring points, but the temperature changes in the 4th to 7th temperature zones were not obvious. The temperature of the first three temperature zones was selected as the process variable $\mathbf{x}_t$ of the model. At the same time, the residual lithium content of the product reflects the quality of the battery, as does the quality variable $\mathbf{y}_t$ of the model, Table 1 lists the physical meaning of these variables.

Table 1. Selected variables in the sintering process.

No.	Measured Variables
1	Below temperature of 1st zone
2	Upper temperature of 1st zone
3	Below temperature of 2nd zone
4	Upper temperature of 2nd zone
5	Below temperature of 3rd zone
6	Upper temperature of 3rd zone
7	Lithium loss coefficient

To test the monitoring effect of the model under different faults, a total of 2200 continuous time data samples were collected on site with a sampling period of five minutes. The process included

a total of three types of faults such as over-temperature, under-temperature and shutdown. For detailed status information, see Table 2.

Table 2. Process data description.

Date Types	Data Description	Time Durations
Normal	Normal data	1st–1000th
Fault 1	Normal samples in 1001st–1200th and abnormal samples of 3rd zone temperature rise in 1201st–1400th	1001st–1400th
Fault 2	Normal samples in 1401st–1600th and abnormal samples of 3rd zone temperature drop in 1601st–1800th	1401st–1800th
Fault 3	Normal samples in 1801st–2000th and abnormal samples of downtime fault in 2001st–2200th	1801st–2200th

First, analyze the dynamics of the data and the time lag characteristics of the variables from the data point of view. Figure 4 shows the autocorrelation and cross-correlation diagrams of process data.
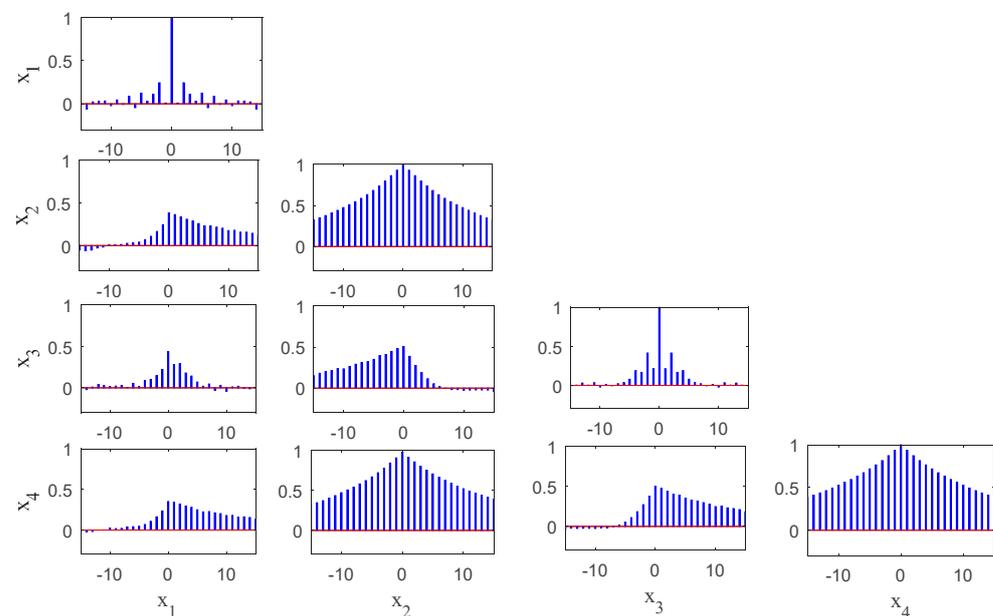

Figure 4. Correlation plots for the first four process variables.

Figure 4 shows the correlation and cross-correlation between the first four process variables. The value at time 0 in each figure represents the cross-correlation between variables; the value at non-zero time shows the autocorrelation between variables under different time lags. It is worth mentioning that the cross-correlation index can measure the redundancy of variable information, and the autocorrelation index can indirectly measure the dynamic and time delay information between variables. It can be seen that the cross-correlation performance between the variables was above 0.5, indicating that there was strong redundant information between the variables. At the same time, even if there was a difference of 10 sampling times, the autocorrelation between the variables was still very high, indicating that there were time lags and dynamic characteristics between the variables. Therefore, the establishment of a DALM model for the process can be considered. The emission equation of the model extracts the redundant information of the data, and the autoregressive equation of the model extracts the dynamic and time lag information of the variables. This paper uses the trend similarity algorithm, which constructs the trend similarity function according to the time lag feature and solves it, to determine the time lag coefficient, that is, L = 3.

To verify the rationality of the time lag coefficient, under different time lag coefficients, a dynamic autoregressive latent variable monitoring model was established respectively. Note: In order to avoid the latent variable dimension from interfering with the selection of the time lag coefficient the latent variable dimension selected by Akaike information criterion (AIC) was temporarily used [32].

The false alarm rate (false alarm rate, FAR) and fault detection rate (fault detection rate, FDR) were defined to evaluate and monitor performance indicators, as defined in (16).

$$FAR = \frac{N_{FAR}}{N_n} \quad FDR = \frac{N_{FDR}}{N_f}. \tag{16}$$

N_{FAR} represents the number of normal samples that were mistakenly detected as abnormal by the monitoring method, and N_n is the number of all normal samples. N_{FDR} represents the number of fault samples correctly monitored by the monitoring method, N_f is the number of all abnormal samples. Therefore, the closer the FAR is to the significance level, the better, and the closer the FDR is to 1, the better. The significance level of this work was set to 0.01.

The first 1000 normal samples were selected to train the model, and the data type fault 1 was used to test the monitoring effect of the model. Table 3 shows the indicators of the monitoring results of the new method under different time lag coefficients.

Table 3. FAR and FDR under different time lag.

Time Lag	2	3	4	5
FAR	0.165	0.050	0.270	\
FDR	0.665	1.000	0.905	\

The model did not converge when the time lag coefficient was 5, and when the model time lag coefficient was 3, the error and false alarm rate of the model were the best. Therefore, when the time lag coefficient was 3, the model gave the best performance. In order to visually see the monitoring results of the model, Figure 5 shows the monitoring T^2 diagram when the model's time lag coefficient was 2, 3 and 4.
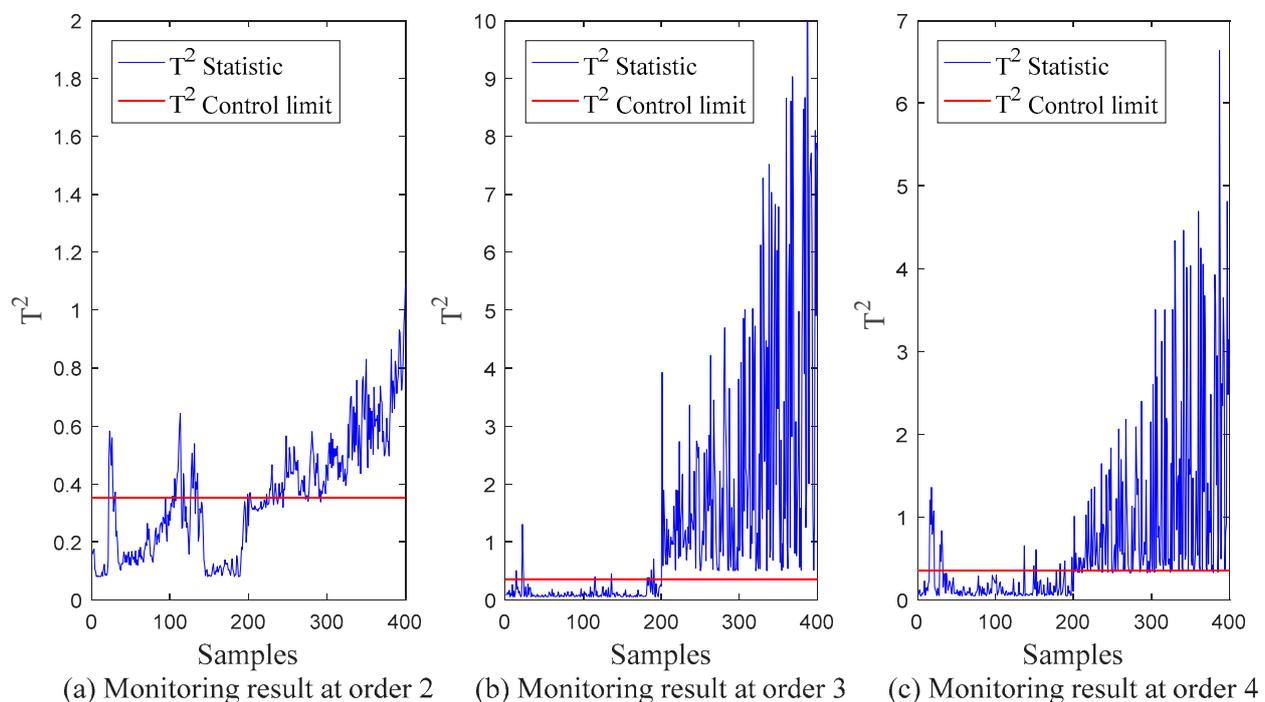

(a) Monitoring result at order 2　　(b) Monitoring result at order 3　　(c) Monitoring result at order 4

Figure 5. Monitoring performance under different time lag. (**a**) Monitoring result at order 2; (**b**) monitoring result at order 3; (**c**) monitoring result at order 4.

It can be seen from Figure 5 that when the model time lag coefficient was 2 and 4, it was easy to misclassify the sample. Especially in the fault interval of 201st–400th: the divided normal samples and abnormal samples were close to the monitoring threshold, which shows that the robustness of the model with this time lag is low; when the model had a time lag coefficient of 3, it is insensitive to the noise and the false alarms are the smallest. Hence, its N_{FAR} and N_{FDR} were the best. Therefore,

the time lag coefficient obtained by the trend similarity identification algorithm enabled the model to obtain a better monitoring effect.

Next, the latent variable dimension was determined. The latent variable dimension is the result of comprehensively considering the complexity and accuracy of the model. The root mean square error (RMSE) is an indicator to measure the accuracy of the model. The expression is shown in (17).

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{|\mathbf{y}_i - \hat{\mathbf{y}}_i|}{N}}, \tag{17}$$

where N is the number of test samples, $\hat{\mathbf{y}}_i$ is the prediction of the true value $\mathbf{y}_i$ and $\bar{\mathbf{y}}_i$ is the mean value of the true value of the test sample. Samples from the 1st to the 600th were used to train the model, and samples from the 601st to the 1000th were used as the test set. Table 4 shows the root mean square error of model prediction under different latent variable dimensions.

Table 4. RMSE under different latent variable dimensions.

Number of Latent Variables	1	2	3	4	5
RMSE	0.123	0.08	0.045	0.042	0.042

Table 4 shows that the prediction performance of the model tends to be stable after the latent variable dimension increased to 3, which was the balance point between model complexity and accuracy. It is worth mentioning that under the time lag coefficient, the latent variable dimension selected by the AIC algorithm was also 3, so the latent variable dimension was determined to be 3.

4.3. Model Performance Test

This section verifies the effect of the proposed monitoring method, and constructs a first-order dynamic process monitoring method: DPLVM [18] and static process monitoring method: PPLSR [33], which were used to compare with the proposed method. The latent variable dimensions of the model were adjusted to 3.

The first 1000 normal samples were used to train the parameters of the model, and the trained model was monitored for three types of different fault samples. In order to distinguish between normal and abnormal samples, the first 200 samples of each type of failure test set were normal samples, and the last 200 samples were their respective failure samples. Table 5 shows the FAR and FDR of different monitoring methods under different failure test sets, and the last line calculates the average value of different indicators.

Table 5. FAR and FDR of the three methods under different fault cases.

Faults	PPLSR		DPLVM		DALM	
	FAR	**FDR**	**FAR**	**FDR**	**FAR**	**FDR**
Fault 1	0.240	0.120	0.210	0.795	0.050	1.000
Fault 2	0.100	0.350	0.155	0.810	0.045	1.000
Fault 3	0.315	0.980	0.080	0.770	0.045	1.000
Average	0.218	0.483	0.148	0.792	0.047	1.000

It can be seen from Table 5 that the monitoring performance of the proposed method was better than that of the static model PPLSR and the first-order dynamic model DPLVM. Therefore, the detection performance was greatly improved after the autoregressive equation was added to the model to extract the dynamic and time lag information. Compared with the basic first-order dynamic DPLVM fault detection method, DALM considered the time lag characteristics, so the model performance was further improved. The detailed monitoring results of the three methods for the three types of faults are shown in Figures 6–8.
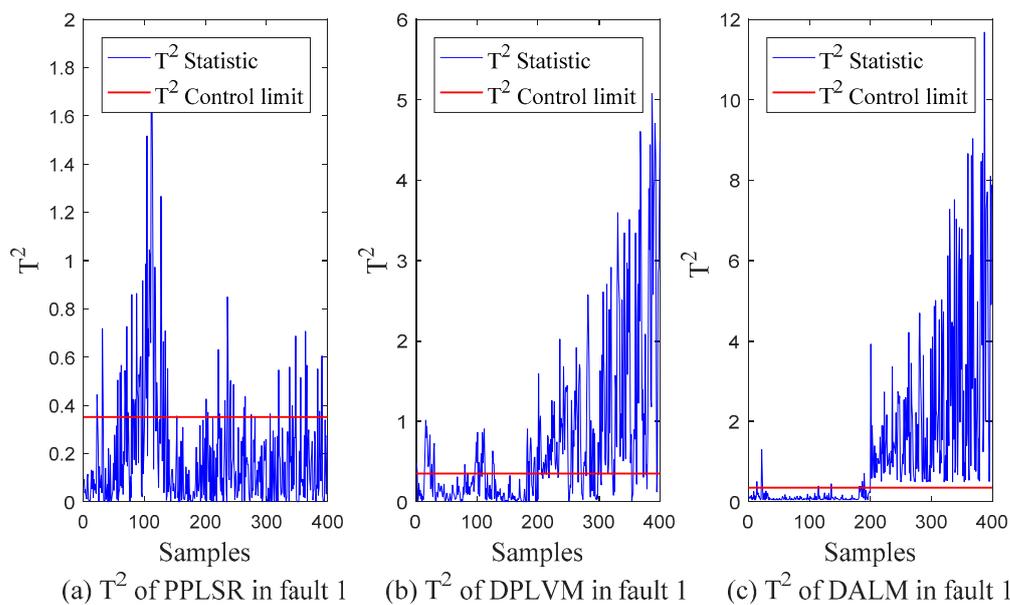
(a) T^2 of PPLSR in fault 1   (b) T^2 of DPLVM in fault 1   (c) T^2 of DALM in fault 1

Figure 6. Monitoring results of fault 1. (**a**) T^2 of PPLSR in fault 1; (**b**) T^2 of DPLVM in fault 1; (**c**) T^2 of DALM in fault 1.

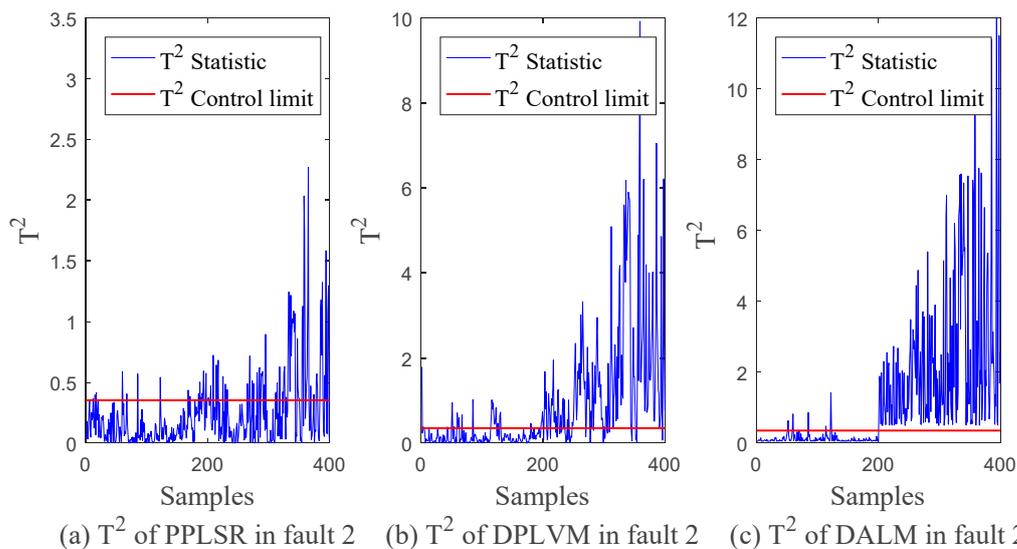(a) T^2 of PPLSR in fault 2   (b) T^2 of DPLVM in fault 2   (c) T^2 of DALM in fault 2

Figure 7. Monitoring results of fault 2. (**a**) T^2 of PPLSR in fault 2; (**b**) T^2 of DPLVM in fault 2; (**c**) T^2 of DALM in fault 2.

For each type of fault test set, the first 200 samples were in a normal state, and the last 200 samples were fault samples. It can be seen from Figure 8 that the static model PPLSR easily mistakenly classified normal samples into faulty samples, and it also easily classified faulty samples into normal samples. The error rate of the first-order dynamic model DPLVM was reduced a lot. Furthermore, the FAR based on the DALM fault detection method proposed in this paper was close to the significance level and the FDR was close to 1, verifying that its monitoring performance was greatly improved.
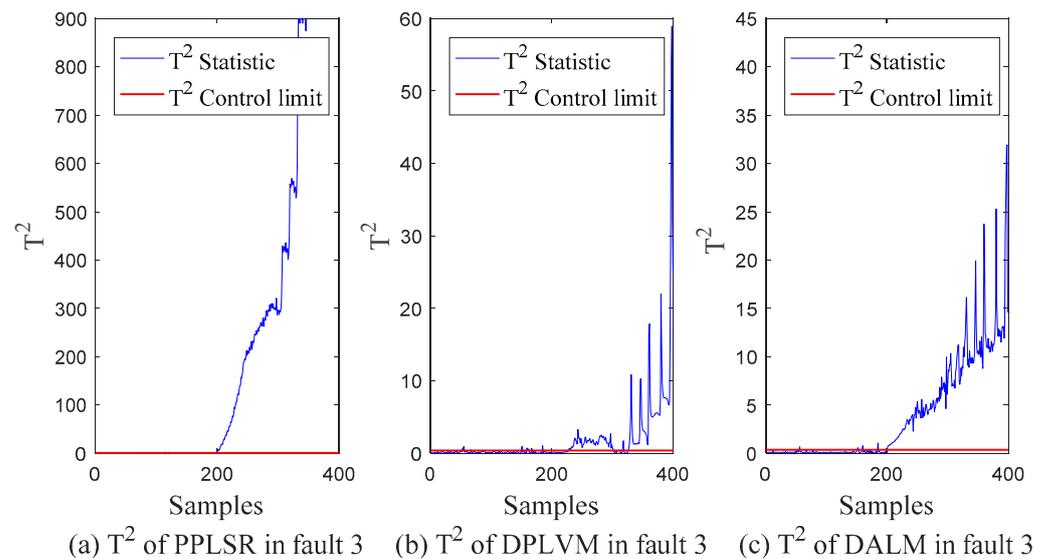
(a) T^2 of PPLSR in fault 3   (b) T^2 of DPLVM in fault 3   (c) T^2 of DALM in fault 3

Figure 8. Monitoring results of fault 3. (**a**) T^2 of PPLSR in fault 3; (**b**) T^2 of DPLVM in fault 3; (**c**) T^2 of DALM in fault 3.

5. Conclusions

A process monitoring method based on the dynamic autoregressive latent variable model was proposed in this paper. Compared with the traditional DPLVM monitoring method, this method not only considered the dynamic characteristics of the process but also considered the complex time lag characteristics, integrated the time lag information into the model, and greatly improved the monitoring performance of the model in the time lag process. First, from the point of data, this method established a dynamic autoregressive latent variable model to adopt the characteristics of dynamics and variable time lag. Then a fusion Bayesian filtering, smoothing and expectation maximization algorithm was used to identify model parameters. Then, on the basis of the identified model, the improved Bayesian filtering technique was used to infer the latent variable distribution of the process state, and the T^2 statistic was constructed for the latent space and online monitoring is performed. Finally, the proposed method was applied to the monitoring of the sintering process of ternary cathode materials. Through industrial case studies, the modeling and monitoring results of the proposed method show that the DALM model was better than the static and first-order dynamic modeling process monitoring methods.

An important issue for process monitoring application in industrial processes is the multi-sampling rate problem. The method proposed in this paper assumed that the input and output data had the same sampling rate. If the sampling rate was inconsistent, some data were deleted by down-sampling. However, a more worthwhile way to try would be to combine semi-supervised learning methods, which can train data on unbalanced input and output data, thereby improving data utilization. Another practical problem is the non-linear relationship between process data, which is very common in industrial processes. How to effectively deal with this problem is worthy of further research in the near future to make the monitoring method more applicable.

Author Contributions: Conceptualization, N.C. and W.G.; methodology, F.H. and J.C.; software, F.H. and Z.C.; validation, J.C.; writing—review and editing, N.C., F.H., J.C., Z.C. and X.L.; supervision, N.C., Z.C. and W.G.; project administration, W.G. and X.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Appendix A. Detailed Derivation of the M-Step

According to the EM algorithm, all the parameters of the DALM model can be updated in M steps. By maximizing the cost function $Q\left(\Theta|\Theta^{old}\right)$, the estimated value Θ^{new} of the next iteration parameter was determined, which is shown in (A1).

$$\Theta^{new} = \underset{\Theta}{\text{argmax}} Q(\Theta|\Theta_{old}). \tag{A1}$$

The Q function was applied to the partial derivative of the model parameters and the derivative was set to zero.

$$\frac{\partial Q(\Theta|\Theta_{old})}{\partial \Theta} = 0. \tag{A2}$$

The updated value of the model parameter Θ^{new} was obtained, as shown in (A3)–(A10).

$$\mathbf{u}_0^{new} = E_{\mathbf{z}_T}\left(\begin{bmatrix} \mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1} \end{bmatrix}\right), \tag{A3}$$

$$\mathbf{V}_0^{new} = E_{\mathbf{z}_T}\left(\begin{bmatrix} \mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1} \end{bmatrix}\begin{bmatrix} \mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1} \end{bmatrix}^T\right) - E_{\mathbf{z}_T}\left(\begin{bmatrix} \mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1} \end{bmatrix}\right) E_{\mathbf{z}_T}\left(\begin{bmatrix} \mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{-L+1} \end{bmatrix}\right)^T, \tag{A4}$$

$$\mathbf{A}^{new} = \sum_{t=1}^{T} E_{\mathbf{z}_T}\left(\mathbf{z}_t\begin{bmatrix} \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L} \end{bmatrix}^T\right)\left[\sum_{t=1}^{T} E_{\mathbf{z}_T}\left(\begin{bmatrix} \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L} \end{bmatrix}\begin{bmatrix} \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L} \end{bmatrix}^T\right)\right]^{-1}, \tag{A5}$$

$$\Sigma_{\mathbf{z}}^{new} = \frac{1}{T}\sum_{t=1}^{T}\left(E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_t^T\right) - 2\mathbf{A}^{new}E_{\mathbf{z}_T}\left(\begin{bmatrix} \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L} \end{bmatrix}\mathbf{z}_t^T\right) + \mathbf{A}^{new}E_{\mathbf{z}_T}\left(\begin{bmatrix} \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L} \end{bmatrix}\begin{bmatrix} \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t-L} \end{bmatrix}^T\right)\mathbf{A}^{newT}\right), \tag{A6}$$

$$\mathbf{B}_{\mathbf{x}}^{new} = \sum_{t=1}^{T}\mathbf{x}_t E_{\mathbf{z}_T}\left(\mathbf{z}_t^T\right)\left[\sum_{t=1}^{T} E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_t^T\right)\right]^{-1}, \tag{A7}$$

$$\mathbf{B}_{\mathbf{y}}^{new} = \sum_{t=1}^{T}\mathbf{y}_t E_{\mathbf{z}_T}\left(\mathbf{z}_t^T\right)\left[\sum_{t=1}^{T} E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_t^T\right)\right]^{-1}, \tag{A8}$$

$$\Sigma_{\mathbf{z}}^{new} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_t\mathbf{x}_t^T - 2\mathbf{B}_{\mathbf{x}}^{new}E_{\mathbf{z}_T}(\mathbf{z}_t)\mathbf{x}_t^T + \mathbf{B}_{\mathbf{x}}^{new}E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_t^T\right)\mathbf{B}_{\mathbf{x}}^{newT}\right), \tag{A9}$$

$$\Sigma_{\mathbf{y}}^{new} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{y}_t\mathbf{y}_t^T - 2\mathbf{B}_{\mathbf{y}}^{new}E_{\mathbf{z}_T}(\mathbf{z}_t)\mathbf{y}_t^T + \mathbf{B}_{\mathbf{y}}^{new}E_{\mathbf{z}_T}\left(\mathbf{z}_t\mathbf{z}_t^T\right)\mathbf{B}_{\mathbf{y}}^{newT}\right). \tag{A10}$$

The updated parameter set $\Theta^{new} = \left\{\mathbf{A}^{new}, \mathbf{B}_{\mathbf{x}}^{new}, \mathbf{B}_{\mathbf{y}}^{new}, \mathbf{u}_0^{new}, \mathbf{V}_0^{new}, \Sigma_{\mathbf{z}}^{new}, \Sigma_{\mathbf{x}}^{new}, \Sigma_{\mathbf{y}}^{new}\right\}$, E steps and M steps were iterated until the parameter Θ matrix converged, that is, satisfied (A11), where ς is a sufficiently small constant, and the model parameter identification was completed.

$$||\Theta^{new} - \Theta^{old}|| < \varsigma. \tag{A11}$$

Among them, $\boldsymbol{\Theta}^{old}$ is the parameter of the last iteration, and $\boldsymbol{\Theta}^{new}$ is the parameter after this round of iteration. Only when the parameters obtained by two adjacent identifications converged did the algorithm stop calculating. Therefore, the parameter convergence can be guaranteed by the EM algorithm itself.

Appendix B. Detailed Derivation of the E-Step

In order to determine the statistics $E_{\mathbf{z}_T}(\mathbf{z}_t)$, $E_{\mathbf{z}_T}(\mathbf{z}_t \mathbf{z}_t^T)$ and $E_{\mathbf{z}_T}(\mathbf{z}_t \mathbf{z}_{t-i}^T)$, the forward and backward algorithm were employed. This is an iterative calculation method, which includes the forward filtering and backward correction step.

In the Bayesian filtering stage, the goal was to calculate the posterior probability of the latent variable $[\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1}]$ with respect to the variable $\mathbf{x}_{1:t}, \mathbf{y}_{1:t}$ at time t, given the posterior distribution $\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \cdots, \mathbf{z}_{t-L} | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1} \sim N(\mathbf{u}_{t-1}, \mathbf{V}_{t-1})$ of the latent variable $[\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \cdots, \mathbf{z}_{t-L}]$ at the previous time $t-1$ on the variable $\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}$, as shown in (A12) where $1 \le t \le T$,

$$
\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1} | \mathbf{x}_{1:t}, \mathbf{y}_{1:t} \sim N(\mathbf{u}_t, \mathbf{V}_t) = N\left(\begin{bmatrix} \mathbf{u}_t^1 \\ \vdots \\ \mathbf{u}_t^L \end{bmatrix}, \begin{bmatrix} \mathbf{V}_t^{11} & \cdots & \mathbf{V}_t^{1L} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_t^{L1} & \cdots & \mathbf{V}_t^{LL} \end{bmatrix} \right). \tag{A12}
$$

The joint probability distribution of the latent variables $[\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1}]$ and $\mathbf{x}_t, \mathbf{y}_t$ with respect to the variable $\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}$ is shown in (A13).

$$
\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1}, \mathbf{x}_t, \mathbf{y}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1} \sim N(\mathbf{g}_t, \mathbf{G}_t)
$$
$$
= N\left(\begin{bmatrix} \mathbf{g}_t^1 \\ \vdots \\ \mathbf{g}_t^{L+2} \end{bmatrix}, \begin{bmatrix} \mathbf{G}_t^{11} & \cdots & \mathbf{G}_t^{1(L+2)} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_t^{(L+2)1} & \cdots & \mathbf{G}_t^{(L+2)(L+2)} \end{bmatrix} \right). \tag{A13}
$$

The parameters of (A13) can be calculated by (A15)

$$
\begin{cases}
\mathbf{g}_t^1 = E(\mathbf{z}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{A}\mathbf{u}_{t-1} \\
\mathbf{g}_t^i = E(\mathbf{z}_{t-i+1} | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{u}_{t-1}^{i-1} \\
\mathbf{g}_t^{L+1} = E(\mathbf{x}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{C}\mathbf{g}_t^1 \\
\mathbf{g}_t^{L+2} = E(\mathbf{y}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{P}\mathbf{g}_t^1
\end{cases}, \tag{A14}
$$

$$
\begin{cases}
\mathbf{G}_t^{11} = \text{cov}(\mathbf{z}_t, \mathbf{z}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{A}\mathbf{u}_{t-1}\mathbf{A}^T + \Sigma_{\mathbf{z}} \\
\mathbf{G}_t^{1i} = \text{cov}(\mathbf{z}_t, \mathbf{z}_{t-i+1} | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{A} \begin{bmatrix} \mathbf{V}_{t-1}^{1(i-1)} \\ \vdots \\ \mathbf{V}_{t-1}^{L(i-1)} \end{bmatrix} \\
\mathbf{G}_t^{i1} = \left(\mathbf{G}_t^{1i}\right)^T \\
\mathbf{G}_t^{ij} = \text{cov}(\mathbf{z}_{t-i+1}, \mathbf{z}_{t-j+1} | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{V}_{t-1}^{(i-1)(j-1)} \\
\mathbf{G}_t^{(L+1)k} = \text{cov}(\mathbf{x}_t, \mathbf{z}_{t-k+1} | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{B}_{\mathbf{x}}\mathbf{G}_t^{1k} \\
\mathbf{G}_t^{(L+1)k} = \left(\mathbf{G}_t^{k(L+1)}\right)^T \\
\mathbf{G}_t^{(L+2)k} = \text{cov}(\mathbf{y}_t, \mathbf{z}_{t-k+1} | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{B}_{\mathbf{y}}\mathbf{G}_t^{1k} \\
\mathbf{G}_t^{(L+2)k} = \left(\mathbf{G}_t^{k(L+2)}\right)^T \\
\mathbf{G}_t^{(L+1)(L+1)} = \text{cov}(\mathbf{x}_t, \mathbf{x}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{B}_{\mathbf{x}}\mathbf{G}_t^{11}\mathbf{B}_{\mathbf{x}}^T + \Sigma_{\mathbf{x}} \\
\mathbf{G}_t^{(L+1)(L+2)} = \text{cov}(\mathbf{x}_t, \mathbf{y}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{B}_{\mathbf{x}}\mathbf{G}_t^{11}\mathbf{B}_{\mathbf{y}}^T \\
\mathbf{G}_t^{(L+2)(L+1)} = \left(\mathbf{G}_t^{(L+2)(L+1)}\right)^T \\
\mathbf{G}_t^{(L+2)(L+2)} = \text{cov}(\mathbf{y}_t, \mathbf{y}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = \mathbf{B}_{\mathbf{y}}\mathbf{G}_t^{11}\mathbf{B}_{\mathbf{y}}^T + \Sigma_{\mathbf{y}}
\end{cases}, \tag{A15}
$$

where $i = 2, 3, \cdots, L; j = 2, 3, \cdots, L; k = 1, 2, \cdots, L$, therefore, according to the knowledge of conditional probability [34] and Appendix C, the mean value and variance of the latent variable filter distribution $\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1}|\mathbf{x}_{1:t}, \mathbf{y}_{1:t} \sim N(\mathbf{u}_t, \mathbf{V}_t)$ were calculated as shown in (A16).

$$
\mathbf{u}_t = \begin{bmatrix} \mathbf{g}_t^1 \\ \vdots \\ \mathbf{g}_t^L \end{bmatrix} + \begin{bmatrix} \mathbf{G}_t^{1(L+1)} & \mathbf{G}_t^{1(L+2)} \\ \vdots & \vdots \\ \mathbf{G}_t^{L(L+1)} & \mathbf{G}_t^{L(L+2)} \end{bmatrix} \begin{bmatrix} \mathbf{G}_t^{(L+1)(L+1)} & \mathbf{G}_t^{(L+1)(L+2)} \\ \mathbf{G}_t^{(L+2)(L+1)} & \mathbf{G}_t^{(L+2)(L+2)} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_t - \mathbf{g}_t^{L+1} \\ \mathbf{y}_t - \mathbf{g}_t^{L+2} \end{bmatrix},
$$

$$
\mathbf{u}_t = \begin{bmatrix} \mathbf{G}_t^{11} & \cdots & \mathbf{G}_t^{1L} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_t^{L1} & \cdots & \mathbf{G}_t^{LL} \end{bmatrix} - \begin{bmatrix} \mathbf{G}_t^{1(L+1)} & \mathbf{G}_t^{1(L+2)} \\ \vdots & \vdots \\ \mathbf{G}_t^{L(L+1)} & \mathbf{G}_t^{L(L+2)} \end{bmatrix} \begin{bmatrix} \mathbf{G}_t^{(L+1)(L+1)} & \mathbf{G}_t^{(L+1)(L+2)} \\ \mathbf{G}_t^{(L+2)(L+1)} & \mathbf{G}_t^{(L+2)(L+2)} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{G}_t^{1(L+1)} & \mathbf{G}_t^{1(L+2)} \\ \vdots & \vdots \\ \mathbf{G}_t^{L(L+1)} & \mathbf{G}_t^{L(L+2)} \end{bmatrix}^{T}.
$$

(A16)

In the Bayesian smoothing stage, the goal was to calculate the posterior probability $\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+2}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T} \sim N(\mathbf{m}_{t+1}, \mathbf{M}_{t+1})$ of the latent variable $[\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+2}]$ with respect to the variable $\mathbf{x}_{1:T}, \mathbf{y}_{1:T}$ at time $t + 1$ to calculate the posterior probability of the latent variable $[\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1}]$ with respect to the variable $\mathbf{x}_{1:T}, \mathbf{y}_{1:T}$ at time t, as shown in (A17).

$$
\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T} \sim N(\mathbf{m}_t, \mathbf{M}_t) = N\left(\begin{bmatrix} \mathbf{m}_t^1 \\ \vdots \\ \mathbf{m}_t^L \end{bmatrix}, \begin{bmatrix} \mathbf{M}_t^{11} & \cdots & \mathbf{M}_t^{1L} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_t^{L1} & \cdots & \mathbf{M}_t^{LL} \end{bmatrix} \right),
$$

(A17)

where $0 \leq t \leq T$, T, there is $\mathbf{m}_T = \mathbf{u}_T, \mathbf{M}_T = \mathbf{V}_T$. In order to calculate the distribution, first, the posterior distribution of the latent variable $[\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+1}]$ was calculated with respect to the variable $\mathbf{x}_{1:t}, \mathbf{y}_{1:t}$, as shown in (A18).

$$
\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+1}|\mathbf{x}_{1:t}, \mathbf{y}_{1:t} \sim N(\mathbf{d}_t, \mathbf{D}_t) = N\left(\begin{bmatrix} \mathbf{d}_t^1 \\ \vdots \\ \mathbf{d}_t^{L+1} \end{bmatrix}, \begin{bmatrix} \mathbf{D}_t^{11} & \cdots & \mathbf{D}_t^{1(L+1)} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_t^{(L+1)1} & \cdots & \mathbf{D}_t^{(L+1)(L+1)} \end{bmatrix} \right).
$$

(A18)

The parameter calculation of (A18) is shown in (A19)–(A20).

$$
\begin{cases}
\mathbf{d}_t^1 = E(\mathbf{z}_{t+1}|\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = \mathbf{A}\mathbf{u}_t \\
\mathbf{d}_t^i = E(\mathbf{z}_{t-i+2}|\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = \mathbf{u}_t^{i-1} \\
\mathbf{D}_t^{11} = \mathrm{cov}(\mathbf{z}_{t+1}, \mathbf{z}_{t+1}|\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = \mathbf{A}\mathbf{V}_t\mathbf{A}^T + \Sigma_{\mathbf{Q}} \\
\mathbf{D}_t^{1i} = \mathrm{cov}(\mathbf{z}_{t+1}, \mathbf{z}_{t-i+2}|\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = \mathbf{A}\begin{bmatrix} \mathbf{V}_t^{1(i-1)} \\ \vdots \\ \mathbf{V}_t^{L(i-1)} \end{bmatrix}, \\
\mathbf{D}_t^{i1} = \left(\mathbf{D}_t^{1i}\right)^T \\
\mathbf{D}_t^{ij} = \mathrm{cov}\left(\mathbf{z}_{t-i+2}, \mathbf{z}_{t-j+2}|\mathbf{x}_{1:t}, \mathbf{y}_{1:t}\right) = \mathbf{V}_t^{(i-1)(j-1)}
\end{cases}
$$

(A19)

where i = 2, 3, $\cdots$, $L + 1$; j = 2, 3, $\cdots$, $L + 1$, and then the following distribution was calculated.

$$
P(\mathbf{z}_{t-L+1}|\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+2}, \mathbf{x}_{1:T}, \mathbf{y}_{1:T}) = P(\mathbf{z}_{t-L+1}|\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+2}, \mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = N(\mathbf{r}_t, \mathbf{R}_t).
$$

(A20)

According to the knowledge of conditional probability [34] and Appendix C. The calculation of its mean and variance is shown in (A21).

$$
\mathbf{r}_t = \mathbf{d}_t^{L+1} + \begin{bmatrix} \mathbf{D}_t^{1(L+1)} \\ \vdots \\ \mathbf{D}_t^{L(L+1)} \end{bmatrix}^T \begin{bmatrix} \mathbf{D}_t^{11} & \cdots & \mathbf{D}_t^{11} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_t^{L1} & \vdots & \mathbf{D}_t^{LL} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{z}_{t+1} - \mathbf{d}_t^1 \\ \vdots \\ \mathbf{z}_{t-L+2} - \mathbf{d}_t^L \end{bmatrix},
$$

$$
R_t = \mathbf{D}_t^{(L+1)(L+1)} - \begin{bmatrix} \mathbf{D}_t^{1(L+1)} \\ \vdots \\ \mathbf{D}_t^{L(L+1)} \end{bmatrix}^T \begin{bmatrix} \mathbf{D}_t^{11} & \cdots & \mathbf{D}_t^{11} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_t^{L1} & \cdots & \mathbf{D}_t^{LL} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}_t^{1(L+1)} \\ \vdots \\ \mathbf{D}_t^{L(L+1)} \end{bmatrix}. \tag{A21}
$$

The posterior probability of the latent variable $[\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+1}]$ was obtained with respect to the variable $\mathbf{x}_{1:T}, \mathbf{y}_{1:T}$ as shown in (A22).

$$
\begin{aligned}
&P(\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+1} | \mathbf{x}_{1:T}, \mathbf{y}_{1:T}) \\
&= P(\mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+2} | \mathbf{x}_{1:T}, \mathbf{y}_{1:T}) P(\mathbf{z}_{t-1} | \mathbf{z}_{t+1}, \mathbf{z}_t, \cdots, \mathbf{z}_{t-L+2}, \mathbf{x}_{1:T}, \mathbf{y}_{1:T}) \\
&= N(\mathbf{h}_t, \mathbf{H}_t) = N\left(\begin{bmatrix} \mathbf{h}_t^1 \\ \vdots \\ \mathbf{h}_t^{L+1} \end{bmatrix}, \begin{bmatrix} \mathbf{H}_t^{11} & \cdots & \mathbf{H}_t^{1(L+1)} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_t^{(L+1)1} & \cdots & \mathbf{H}_t^{(L+1)(L+1)} \end{bmatrix} \right).
\end{aligned} \tag{A22}
$$

The mean and variance of the distribution were calculated as shown in (A23).

$$
\mathbf{h}_t = \begin{bmatrix} \mathbf{m}_{t+1} \\ \mathbf{K} \begin{bmatrix} \mathbf{m}_{t+1}^1 - \mathbf{d}_t^1 \\ \vdots \\ \mathbf{m}_{t+1}^L - \mathbf{d}_t^L \end{bmatrix} + \mathbf{d}_t^{L+1} \end{bmatrix},
$$

$$
\mathbf{H}_t = \begin{bmatrix} \mathbf{M}_{t+1} & \mathbf{M}_{t+1}\mathbf{K}^T \\ \mathbf{KM}_{t+1} & \mathbf{KM}_{t+1}\mathbf{K}^T + R_t \end{bmatrix} \text{ where } \mathbf{K} = \begin{bmatrix} \mathbf{D}_t^{1(L+1)} \\ \vdots \\ \mathbf{D}_t^{L(L+1)} \end{bmatrix}^T \begin{bmatrix} \mathbf{D}_t^{11} & \cdots & \mathbf{D}_t^{11} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_t^{L1} & \vdots & \mathbf{D}_t^{LL} \end{bmatrix}^{-1}. \tag{A23}
$$

According to the Bayesian smoothing rule [31], the smooth distribution of the latent variable was obtained, as shown in (A24).

$$
\mathbf{z}_t, \mathbf{z}_{t-1}, \cdots, \mathbf{z}_{t-L+1} | \mathbf{x}_{1:T}, \mathbf{y}_{1:T} \sim N(\mathbf{m}_t, \mathbf{M}_t). \tag{A24}
$$

The calculation of its mean and variance is shown in (A25).

$$
\mathbf{m}_t = \begin{bmatrix} \mathbf{h}_t^2 \\ \vdots \\ \mathbf{h}_t^{L+1} \end{bmatrix} \quad \mathbf{M}_t = \begin{bmatrix} \mathbf{H}_t^{22} & \cdots & \mathbf{H}_t^{2(L+1)} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_t^{(L+1)2} & \cdots & \mathbf{H}_t^{(L+1)(L+1)} \end{bmatrix}. \tag{A25}
$$

Appendix C. Properties of Gaussian Distribution

Definition A1. *(Gaussian distribution) A random variable* $\mathbf{x} \in R^n$ *has a Gaussian distribution with mean* $\mathbf{m} \in R^n$ *and covariance* $\mathbf{P} \in R^{n \times n}$ *if its probability density has the form.*

$$
N(\mathbf{x}|\mathbf{m}, \mathbf{P}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{P}^T (\mathbf{x} - \mathbf{m}) \right), \tag{A26}
$$

where $|\mathbf{P}|$ *is the determinant of the matrix* $\mathbf{P}$.

Lemma A1. *(Joint distribution of Gaussian variables) If random variables* $\mathbf{x} \in R^n$ *and* $\mathbf{y} \in R^{\mathbf{m}}$ *have the Gaussian probability distributions.*

$$
\begin{aligned}
\mathbf{x} &\sim \mathbf{N}(\mathbf{m}, \mathbf{P}), \\
\mathbf{y}|\mathbf{x} &\sim \mathbf{N}(\mathbf{Hx} + \mathbf{u}, \mathbf{R}).
\end{aligned} \tag{A27}
$$

then the joint distribution of **x**,**y** *and the marginal distribution of* **y** *are given as (A28).*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{m} \\ \mathbf{Hm}+\mathbf{u} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{PH^T} \\ \mathbf{HP} & \mathbf{HPH^T}+\mathbf{R} \end{pmatrix} \right),$$
$$\mathbf{y} \sim \mathbf{N}(\mathbf{Hm}+\mathbf{u}, \mathbf{HPH^T}+\mathbf{R}). \tag{A28}$$

Lemma A2. *(Conditional distribution of Gaussian variables) If the random variables* **x** *and* **y** *have the joint Gaussian probability distribution.*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C^T} & \mathbf{B} \end{pmatrix} \right). \tag{A29}$$

then the marginal and conditional distributions of **x** *and* **y** *are given as follows:*

$$\begin{aligned} \mathbf{x} &\sim \mathbf{N}(\mathbf{a}, \mathbf{A}), \\ \mathbf{y} &\sim \mathbf{N}(\mathbf{b}, \mathbf{B}), \\ \mathbf{x}|\mathbf{y} &\sim \mathbf{N}(\mathbf{a}+\mathbf{CB}^{-1}(\mathbf{y}-\mathbf{b}), \mathbf{A}-\mathbf{CB}^{-1}\mathbf{C}^T), \\ \mathbf{y}|\mathbf{x} &\sim \mathbf{N}(\mathbf{b}+\mathbf{C}^T\mathbf{A}^{-1}(\mathbf{x}-\mathbf{a}), \mathbf{B}-\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}). \end{aligned} \tag{A30}$$

References

1. Chen, Z.W.; Cao, Y.; Ding, S.X.; Zhang, K.; Koenings, T.; Peng, T.; Yang, C.H.; Gui, W.H. A Distributed Canonical Correlation Analysis-Based Fault Detection Method for Plant-Wide Process Monitoring. *IEEE Trans. Ind. Inf.* **2019**, *15*, 2710–2720. [CrossRef]
2. Qin, S.J. Survey on data-driven industrial process monitoring and diagnosis. *Annu. Rev. Control* **2012**, *36*, 220–234. [CrossRef]
3. Venkatsubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S.N. A review of process fault detection and diagnosis Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, *27*, 293–311. [CrossRef]
4. Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.N. A review of process fault detection and diagnosis Part II: Quantitative model and search strategies. *Comput. Chem. Eng.* **2003**, *27*, 313–326. [CrossRef]
5. Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.N.; Yin, K. A review of process fault detection and diagnosis Part III: Process history based methods. *Comput. Chem. Eng.* **2003**, *27*, 327–346. [CrossRef]
6. Chen, H.; Jiang, B.; Ding, S.; Huang, B.S. Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Trans. Intell. Transp. Syst.* **2020**. aarly access. [CrossRef]
7. Kim, D.S.; Lee, I.B. Process monitoring based on probabilistic PCA. *Chemometr. Intell. Lab. Syst.* **2003**, *67*, 109–123. [CrossRef]
8. Kong, X.Y.; Cao, Z.H.; An, Q.S.; Gao, Y.B.; Du, B.Y. Quality-Related and Process-Related Fault Monitoring With Online Monitoring Dynamic Concurrent PLS. *IEEE Access* **2018**, *6*, 59074–59086. [CrossRef]
9. Chen, H.; Jiang, B.; Chen, W.; Yi, H. Data-driven Detection and Diagnosis of Incipient Faults in Electrical Drives of High-Speed Trains. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4716–4725. [CrossRef]
10. Ge, Z.Q. Process Data Analytics via Probabilistic Latent Variable Models: A Tutorial Review. *Ind. Eng. Chem. Res.* **2018**, *57*, 12646–12661. [CrossRef]
11. Si, Y.B.; Wang, Y.Q.; Zhou, D.H. Key-Performance-Indicator-Related Process Monitoring Based on Improved Kernel Partial Least Squares. *IEEE Trans. Ind. Electron.* **2021**, *68*, 2626–2636. [CrossRef]
12. Jiang, Y.; Fan, J.L.; Chai, T.; Lewis, F.L.; Li, J.N. Tracking Control for Linear Discrete-Time Networked Control Systems With Unknown Dynamics and Dropout. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 4607–4620. [CrossRef] [PubMed]
13. Ge, Z.Q.; Chen, X.R. Dynamic Probabilistic Latent Variable Model for Process Data Modeling and Regression Application. *IEEE Trans. Control Syst. Technol.* **2019**, *27*, 323–331. [CrossRef]
14. Kruger, U.; Zhou, Y.Q.; Irwin, G.W. Improved principal component monitoring of large-scale processes. *J. Process Control* **2004**, *14*, 879–888. [CrossRef]
15. Chen, J.H.; Liu, K.C. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chem. Eng. Sci.* **2002**, *57*, 63–75. [CrossRef]
16. Li, G.; Qin, S.J.; Zhou, D.H. A New Method of Dynamic Latent-Variable Modeling for Process Monitoring. *IEEE Trans. Ind. Electron.* **2014**, *61*, 6438–6445. [CrossRef]
17. Ge, Z.Q.; Song, Z.H.; Gao, F.R. Review of Recent Research on Data-Based Process Monitoring. *Ind. Eng. Chem. Res.* **2013**, *52*, 3543–3562. [CrossRef]
18. Ge, Z.Q.; Chen, X.R. Supervised linear dynamic system model for quality related fault detection in dynamic processes. *J. Process Control* **2016**, *44*, 224–235. [CrossRef]
19. Zhurabok, A.N.; Shumsky, A.E.; Pavlov, S.V. Diagnosis of Linear Dynamic Systems by the Nonparametric Method. *Autom. Remote Control* **2017**, *78*, 1173–1188. [CrossRef]

20. Ren, X.M.; Rad, A.B.; Chan, P.T.; Lo, W.L. Online identification of continuous-time systems with unknown time delay. *IEEE Trans. Autom. Control* **2005**, *50*, 1418–1422. [CrossRef]

21. Yang, Z.J.; Iemura, H.; Kanae, S.; Wada, K. Identification of continuous-time systems with multiple unknown time delays by global nonlinear least-squares and instrumental variable methods. *Automatica* **2007**, *43*, 1257–1264. [CrossRef]

22. Yalin, W.; Haibing, X.; Xiaofeng, Y. Multi-delay identification by trend-similarity analysis and its application to hydrocracking process. *Chem. Eng. News* **2018**, *69*, 1149–1157.

23. Drakunov, S.V.; Perruquetti, W.; Richard, J.P.; Belkoura, L. Delay identification in time-delay systems using variable structure observers. *Annu. Rev. Control* **2006**, *30*, 143–158. [CrossRef]

24. Shen, B.B.; Ge, Z.Q. Supervised Nonlinear Dynamic System for Soft Sensor Application Aided by Variational Auto-Encoder. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 6132–6142. [CrossRef]

25. Prosper, H.B. Deep Learning and Bayesian Methods. In Proceedings of the 12th Conference on Quark Confinement and the Hadron Spectrum, Thessaloniki, Greece, 29 August–3 September 2016.

26. Alcala, C.F.; Qin, S.J. Reconstruction-based contribution for process monitoring. *Automatica* **2009**, *45*, 1593–1600. [CrossRef]

27. Lataire, J.; Chen, T. Transfer function and transient estimation by Gaussian process regression in the frequency domain. *Automatica* **2016**, *72*, 217–229. [CrossRef]

28. Panic, B.; Klemenc, J.; Nagode, M. Improved Initialization of the EM Algorithm for Mixture Model Parameter Estimation. *Mathematics* **2020**, *8*, 373. [CrossRef]

29. Chen, J.Y.; Gui, W.H.; Dai, J.Y.; Jiang, Z.H.; Chen, N.; Li, X. A hybrid model combining mechanism with semi-supervised learning and its application for temperature prediction in roller hearth kiln. *J. Process Control* **2021**, *98*, 18–29. [CrossRef]

30. Huang, X.K.; Tian, X.Y.; Zhong, Q.; He, S.W.; Huo, C.B.; Cao, Y.; Tong, Z.Q.; Li, D.C. Real-time process control of powder bed fusion by monitoring dynamic temperature field. *Adv. Manuf.* **2020**, *8*, 380–391. [CrossRef]

31. Egorova, E.G.; Rudakova, I.V.; Rusinov, L.A.; Vorobjev, N.V. Diagnostics of sintering processes on the basis of PCA and two-level neural network model. *J. Chemom.* **2018**, *32*, 2. [CrossRef]

32. De Waele, S.; Broersen, P.M.T. Order selection for vector autoregressive models. *IEEE Trans. Signal Process.* **2003**, *51*, 427–433. [CrossRef]

33. Wang, Z.Y.; Liang, J. A JITL-Based Probabilistic Principal Component Analysis for Online Monitoring of Nonlinear Processes. *J. Chem. Eng. Jpn.* **2018**, *51*, 874–889. [CrossRef]

34. Särkkä, S. *Bayesian Filtering and Smoothing*; Cambridge University Press: London, UK, 2013.