

Article

An Improved Hidden Markov Model for Monitoring the Process with Autocorrelated Observations

Yaping Li ^{1,*}, Haiyan Li ¹, Zhen Chen ² and Ying Zhu ²

¹ College of Economics and Management, Nanjing Forestry University, Nanjing 210037, China; lihaiyan117@njfu.edu.cn

² Department of Industrial Engineering & Management, Shanghai Jiao Tong University, Shanghai 200240, China; chenzhendr@sjtu.edu.cn (Z.C.); zhuyingme@sjtu.edu.cn (Y.Z.)

* Correspondence: yapingli@njfu.edu.cn

Abstract: With the development of intelligent manufacturing, automated data acquisition techniques are widely used. The autocorrelations between data that are collected from production processes have become more common. Residual charts are a good approach to monitoring the process with data autocorrelation. An improved hidden Markov model (IHMM) for the prediction of autocorrelated observations and a new expectation maximization (EM) algorithm is proposed. A residual chart based on IHMM is employed to monitor the autocorrelated process. The numerical experiment shows that, in general, IHMMs outperform both conventional hidden Markov models (HMMs) and autoregressive (AR) models in quality shift diagnosis, decreasing the cost of missing alarms. Moreover, the times taken by IHMMs for training and prediction are found to be much less than those of HMMs.

Keywords: hidden Markov model (HMM); autocorrelation; residual chart



Citation: Li, Y.; Li, H.; Chen, Z.; Zhu, Y. An Improved Hidden Markov Model for Monitoring the Process with Autocorrelated Observations. *Energies* **2022**, *15*, 1685. <https://doi.org/10.3390/en15051685>

Academic Editor: Abu-Siada Ahmed

Received: 17 January 2022

Accepted: 16 February 2022

Published: 24 February 2022

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



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

1. Introduction

Process monitoring plays an essential role in intelligent manufacturing [1,2]. Statistical process control (SPC) is a quality control technique that uses statistical methods to monitor and control processes. The aim of SPC is to ensure that the process runs efficiently, producing more products that meet specifications, while reducing waste at the same time. Shewhart control charts are a key SPC tool used to determine whether a process is in control. If the observation value is within the upper and lower control limits, the process is in control; otherwise, the process is out of control. Shewhart charts often assume that the data collected from the process are independent. However, this assumption is not true in a variety of processes. For example, consecutive measurements on chemical and pharmaceutical processes or product characteristics are often highly autocorrelated, or the chronological measurements of every characteristic on every unit in automated test and inspection procedures are often autocorrelated. It has been shown in numerous studies that conventional charts do not work well, in the form of giving too many false alarms if the observations exhibit even a small amount of autocorrelation over time [3–9]. Clearly, better approaches are needed. In the following paragraphs, three common ways to solve the problems related to autocorrelation are introduced.

The first approach to solving an autocorrelation problem is simply to sample from the observation data stream at a lower frequency [10]. This seems to be an easy solution, although it has some shortcomings. Concretely speaking, it makes inefficient use of the available data. For example, if every tenth observation is sampled, approximately 90% of the information is discarded. In addition, since only every tenth datum is used, this approach may delay the discovery of a process shift.

The second approach is to re-estimate the real process variance, aiming to revise the upper and lower control limits. See, for example, [9,11–19].

The third approach is to use residual charts. The statistics in these charts are residuals; values are calculated by subtracting predicted values from observed values. In the implementation of residual charts, the key step is choosing a reasonable prediction model to obtain predicted values. Autoregressive integrated moving average (ARIMA) models are mostly used to model autocorrelated data. See, for example, [3,8,20–26]. In addition, multistage or multivariate autocorrelated processes are mostly studied by using Hotelling T^2 control charts, such as in [27–30]. Some exceptions include Pan et al., who proposed an overall run length (ORL) to replace T^2 charts [31], and S. Yang and C. Yang, who used a residual chart and cause-selecting control chart [32].

With the rapid development of artificial intelligence, machine-learning methods are becoming more and more popular in SPC in the case of observation autocorrelation [33–38]. Most of the related literature focuses on neural networks methods [39–42]. Some successful applications have also been introduced, for example, the failure diagnosis of wind turbine [43], and the prediction of the remaining useful life (RUL) of bearings [44]. Another machine-learning approach proposed in SPC is the hidden Markov model (HMM). Lee et al. proposed a modified HMM, combined with a Hotelling multivariate control chart to perform adaptive fault diagnosis [45]. The HMMs, whose training sets contain autocorrelated data, were employed to forecast observation values for residual charts in process monitoring [46]. However, although HMMs are supposed to deal with the autocorrelated processes, the essence of the model itself is inconsistent with the case of autocorrelations. This is because one key assumption in conventional HMMs is that the observations are independent of each other. Therefore, it is worth developing a modified HMM, by accounting for observation autocorrelations in models themselves, and observing whether it is better than a traditional HMM.

Therefore, to realize the goal of monitoring the autocorrelated process well, the residual chart based on an improved hidden Markov model (IHMM) with autocorrelated observations considered is developed. Due to the autocorrelations, the conventional expectation maximization (EM) algorithm for HMMs is not appropriate. A new EM algorithm is developed for the solutions. The Shewhart residual chart is employed in quality shift detection in conjunction with the IHMM. The residual is defined as the deviation of the predicted value by the IHMM and the current real observation value. Through the residual chart, we are able to see whether the process is in control. If the chart initiates an alarm, one running length is obtained. Thus, the average running lengths (ARLs) can be calculated with sufficient samples. The ARL is set as the comparison index for different models, including IHMM, HMM and AR.

The rest of this paper is organized as follows: Section 2 introduces the development of the IHMM model and its algorithm. In addition, the comparison of the prediction performances of different approaches is presented in Section 2. In Section 3, residual charts are introduced. In Section 4, numerical examples and interesting results are presented. The conclusions and possible areas for future research are given in Section 5.

2. Model Development

2.1. Hidden Markov Models

Denote a Markov chain with a finite state set $\{s_1, s_2, \dots, s_N\}$ by $\{S_n, n = 1, 2, \dots\}$. Let a_{ij} be the probabilities that the Markov chain enters state s_j from state s_i ($1 \leq i, j \leq N$) and $\pi_i = P\{S_1 = s_i\}$, $i = 1, 2, \dots, N$ be the initial state probabilities. Denote a finite set of signals by ζ , and suppose a signal from ζ is sent each time the Markov chain enters a state. Suppose that when the Markov chain enters state s_j independently of previous states and signals, the signal sent is o_r with probability $p(o_r|s_j)$ that meets $\sum_{o_r \in \zeta} p(o_r|s_j) = 1$. That is, if O_n represents the n th signal, then it can be written by

$$P\{O_1 = o_r | S_1 = s_j\} = p(o_r|s_j), \quad (1)$$

$$P\{O_n = o_r | S_1, O_1, \dots, S_{n-1}, O_{n-1}, S_n = s_j\} = p(o_r|s_j). \quad (2)$$

Such a model, in which the signal sequence $O_1, O_2, \dots$ can be observed while the underlying Markov chain state sequence $S_1, S_2, \dots$ cannot be observed, is called a hidden Markov model [47].

Normally, an HMM contains the following elements [48]:

- A finite hidden state set $s = \{s_1, s_2, \dots, s_N\}$, where N is the hidden state number.
- A set of possible observation values (signals) $q = \{q_1, \dots, q_K\}$, where K is the number of possible observation values. Note that if the values of the observations are continuous, K should be infinite.
- An observation sequence, $o = (o_1, \dots, o_T)$, where T is the number of observations, $o_t \in q, 1 \leq t \leq T$.
- A distribution of state transition probabilities, $A = \{a_{ij}\}$, where

$$a_{ij} = P\{S_{t+1} = s_j | S_t = s_i\}, 1 \leq i, j \leq N, \sum_j a_{ij} = 1. \quad (3)$$

- A distribution of initial state probabilities, $\pi = \{\pi_i\}$, where

$$\pi_i = P\{S_1 = s_i\}, i = 1, 2, \dots, N, \sum_i \pi_i = 1. \quad (4)$$

- A conditional probability distribution of the observations, given $S_t = s_i$, $B = \{b_i(q_k)\}$, where

$$b_i(q_k) = P\{o_t = q_k | S_t = s_i\}, 1 \leq i \leq N, 1 \leq t \leq T, \sum_{q_k} b_i(q_k) = 1. \quad (5)$$

In general, an HMM contains three elements, denoted by $\lambda = \{A, B, \pi\}$.

2.2. The Improved HMM with Autocorrelated Observations

2.2.1. The Selection for the Order of Autocorrelation

Different from the traditional HMMs, in which current observations are assumed to be independent of previous states and observations, autocorrelated observations are considered. The observations are assumed to follow a Gaussian distribution, and the current observations are required to be dependent, not only on the current hidden state, but also on the previous observations.

The autocorrelated data from production may be multi-order, and it seems not cost-effective for judging this detailed order by using engineering experience and professional knowledge. Therefore, it is hoped to find a suitable order that will keep the implementation of the IHMM feasible, efficient and cost-effective.

The numerical experiment on the order selection for autocorrelated data in SPC application is designed as follows:

Step 1. Generate stationary autocorrelated data by a d th-order autoregressive model, $AR(d), d \geq 2$;

Step 2. Use $AR(p), p = d, d-1, \dots, 1$ models to fit the data, respectively, and obtain the parameters that need to be estimated by least-squares estimation;

Step 3. Generate 2000 stationary data sequences with a same length of 2000 under the processes with a mean shift magnitude of δ ;

Step 4. Predict each sequence by using $AR(p), p = d, d-1, \dots, 1$ models, respectively, and calculate their residuals;

Step 5. Determine the central lines (CLs), upper control limits (UCLs) and lower control limits (LCLs) of p residual charts;

Step 6. Generate stationary autocorrelated data by a d th-order autoregressive model;

Step 7. Calculate the average running lengths (ARLs) of p residual charts.

In this study, twelve $AR(2)$ and eight $AR(3)$ models are used to generate data. The control limit coefficients of residual charts are 3. The shift magnitudes of the mean are set by 0, 1.5 and 3, respectively.

When data are generated by AR(2) models for an in-control process, the ARL of the residual charts is 371.3803 if the data are fitted by AR(2) models, and the ARL is 361.1688 if the data are fitted by AR(1) models. Thus, the probability of the type I error is increased by 2.83%. Similarly, when data are generated by AR(3) models, the ARL of the residual charts is 371.8250 if the data are fitted by AR(3) models, the ARL is 358.5401 if the data are fitted by AR(2) models, and the ARL is 345.8113 if the data are fitted by AR(1) models. Thus, the probabilities of the type I error are increased by 3.7% and 7.51%, respectively.

The ARLs of different situations are shown in Figure 1, in which the numbers on the x-axis represent the combinations of autocorrelation coefficients.

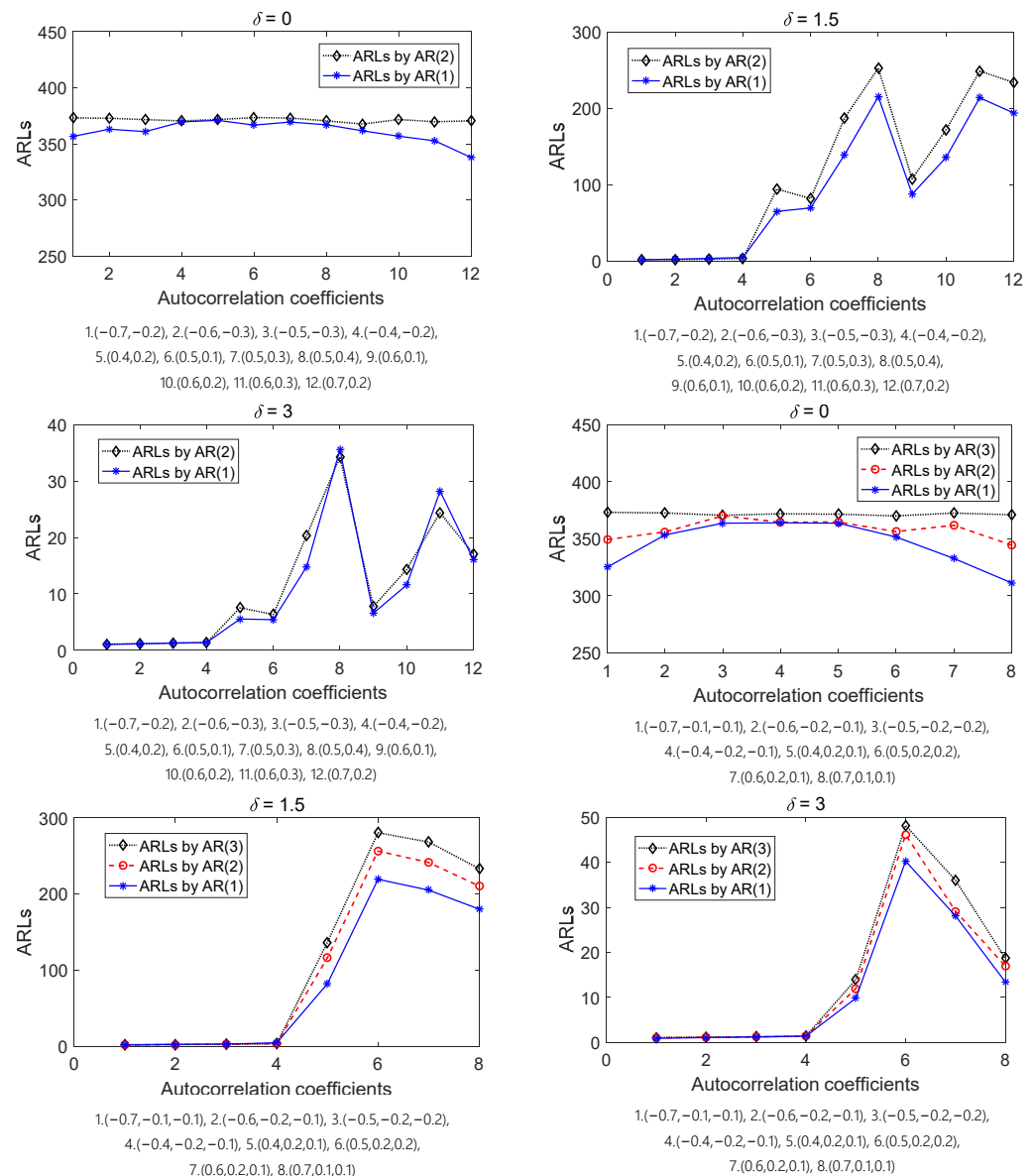


Figure 1. The ARL results of the order selection experiment.

From Figure 1, it is seen that generally, AR(1) models outperformed other AR(d), $d \geq 2$ models in shift detection regardless of the autocorrelation order. Although AR(1) models lead to a slight increase in type I error, it seems to be insignificant compared with their good performances in detecting quality shifts. Therefore, only the first-order autocorrelation is considered in this study. As a result, there is no need to judge what order the autocorrelation is so that the modeling cost can be saved.

2.2.2. The Development of the IHMM

According to the analysis in Section 2.2.1, the construction of the IHMM, in which the current observation is related to its previous observation, is shown in Figure 2.

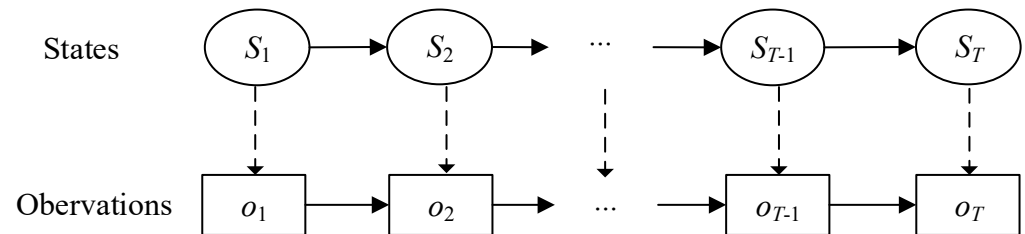


Figure 2. The construction of the IHMM.

In Figure 2, arrows pointing to the right indicate the correlation between neighboring states or observations. It is intuitive that the IHMM will become a traditional HMM if there is no correlation between neighboring observations.

Let $o_i(t)$ be the observation value at time t given state s_i ; $o_i(t)$ can be fitted with the following function:

$$o_i(t) = \zeta_i + c_i o_{t-1} + \varepsilon_i, 1 \leq i \leq N, 1 \leq t \leq T, \quad (6)$$

where c_i is the first-order autocorrelation coefficient, ζ_i is the constant term and ε_i is white noise, following a normal distribution with mean zero and variance σ_i^2 .

Let $\mu_i(t)$ be the mean of $o_i(t)$ given state s_i , $\mathbf{x}_t = (1, o_{t-1})'$, then Equation (6) can be written as:

$$\mu_i(t) = \mathbf{C}_i \mathbf{x}_t, \quad (7)$$

where $\mathbf{C}_i$ is a 1×2 matrix consisting of $\{\zeta_i\}$ and $\{c_i\}$.

Hence, we conclude that the states are a Markov process, and that the conditional observations given state s_i follows a Gaussian distribution with mean $\mathbf{C}_i \mathbf{x}_t$ and variance σ_i^2 , i.e.,

$$P\{S_{t+1}|S_1, \dots, S_t, o_1, \dots, o_t\} = p(S_{t+1}|S_t), \quad (8)$$

$$b_i(o_t|o_{t-1}) = P(o_t|o_{t-1}, S_t = s_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(o_t - \mathbf{C}_i \mathbf{x}_t)^2}{2\sigma_i^2}\right). \quad (9)$$

If the current observation is not related to its previous observations, $\mu_i(t)$ is the constant ζ_i . Thus, the IHMM becomes a traditional HMM.

2.3. Parameter Estimation

The aim of using an IHMM is to forecast observation values. Firstly, we estimate the parameters to obtain an optimal $\hat{\lambda}$ by maximizing the probability $P(o|\lambda)$.

The EM algorithm is a popular method to estimate the parameters for HMMs. However, since autocorrelation between observations is considered, the traditional algorithm could not be used directly. We redefine the parameters as $\lambda = \{A, C, \sigma^2, \pi\}$ where $C = \{\mathbf{C}_i\}$, $\sigma^2 = \{\sigma_i^2\}$, and the definitions of $\mathbf{C}_i$ and σ_i^2 are provided in Section 2.2.2. The flowchart demonstrating how to estimate the parameters in IHMM with an improved EM algorithm is shown in Figure 3.

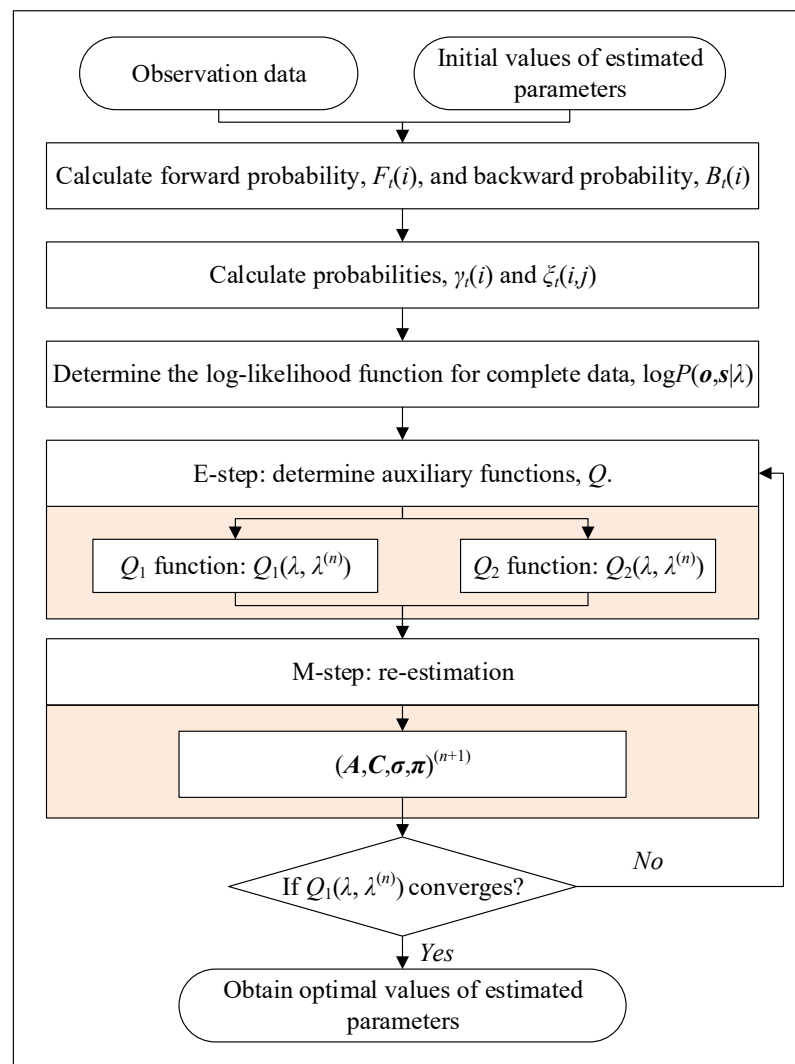


Figure 3. The flowchart of the improved EM algorithm.

Firstly, we introduce two types of probabilities: forward probability, $F_t(i)$, and backward probability, $B_t(i)$, defined as:

$$F_t(i) = P(o_1, \dots, o_t, S_t = s_i | \lambda), \quad (10)$$

$$B_t(i) = P(o_{t+1}, \dots, o_T | S_t = s_i, \lambda). \quad (11)$$

The calculation of the two probabilities here is similar to that of traditional HMMs. Slightly different from traditional HMMs, the initial values are defined by $F_{d+1}(i) = \pi_i b_i(o_{d+1} | o_1, \dots, o_d)$ and $B_T(i) = 1$, respectively.

Based on the forward and backward probabilities, some intermediate probabilities are computed. Given λ and o , denote the probability that the process is in state s_i at time t by $\gamma_t(i)$ and the probability that the process is in state s_i at time t and in state s_j at time $t + 1$ by $\xi_t(i, j)$. $\gamma_t(i)$ and $\xi_t(i, j)$ can be written as Equations (12) and (13). Please refer to reference [48] for the details of the derivations.

$$\gamma_t(i) = \frac{F_t(i) B_t(i)}{\sum_j F_t(i) B_t(i)}, \quad (12)$$

$$\xi_t(i, j) = \frac{F_t(i) a_{ij} b_j(o_{t+1}) B_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N F_t(i) a_{ij} b_j(o_{t+1}) B_{t+1}(j)}. \quad (13)$$

Next, we develop an improved EM algorithm for the IHMM, step by step.

Step 1. Determine the log-likelihood function for complete data.

The complete data are $(o, S) = (o_1, \dots, o_T, s_1, \dots, s_T)$, and its log-likelihood function is $\log P(o, s|\lambda)$.

Step 2. E-step: determine Q functions.

$$Q_1(\lambda, \lambda^{(n)}) = E_s [\log P(o, s|\lambda) | o, \lambda^{(n)}] = \sum_s \log P(o, s|\lambda) P(o, s | \lambda^{(n)}), \quad (14)$$

where λ is the parameter that will maximize the Q_1 function, while $\lambda^{(n)}$ is the current parameter value.

With $P(o, s|\lambda) = \pi_{s_1} b_{s_1}(o_1) a_{s_1 s_2} b_{s_2}(o_2) \cdots a_{s_{T-1} s_T} b_{s_T}(o_T)$, $Q_1(\lambda, \lambda^{(n)})$ can be written by

$$\begin{aligned} Q_1(\lambda, \lambda^{(n)}) &= \sum_s \pi_{s_1} \log P(o, s | \lambda^{(n)}) \\ &+ \sum_s \left(\sum_{t=1}^{T-1} \log a_{s_t s_{t+1}} \right) P(o, s | \lambda^{(n)}) \\ &+ \sum_s \left(\sum_{t=1}^T \log b_{s_t}(o_t) \right) P(o, s | \lambda^{(n)}). \end{aligned} \quad (15)$$

For the re-estimation of $b_i(q_k)$, one more auxiliary function Q_2 is proposed by taking the conditional expectation of the log-likelihood of the observation sequence:

$$Q_2(\lambda, \lambda^{(n)}) = \sum_s \sum_{t=1}^T \gamma_t(i) \left(\ln \left(\frac{1}{\sqrt{2\pi}} \right) + \ln \frac{1}{\sigma_i} - \frac{(o_t - c_i x_t)^2}{2\sigma_i^2} \right). \quad (16)$$

Step 3. M-step: re-estimation.

Re-estimate λ that maximizes $Q_1(\lambda, \lambda^{(n)})$, that is

$$\lambda^{(n+1)} = \operatorname{argmax}_{\lambda} Q_1(\lambda, \lambda^{(n)}). \quad (17)$$

The state transition probabilities are derived as:

$$a_{ij}^{(n+1)} = \frac{\sum_{t=1}^{T-1} \zeta_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}. \quad (18)$$

The initial state probabilities are derived as:

$$\pi_i^{(n+1)} = \gamma_1(i). \quad (19)$$

By re-estimating λ that maximizes $Q_2(\lambda, \lambda^{(n)})$, $c_{i,\tau}^{(n+1)}, \zeta_i^{(n+1)}, \sigma_i^{2(n+1)}$ can be written as:

$$c_i^{(n+1)} = \frac{\sum_{t=1}^T \gamma_t(i) o_{t-1} \left(o_t - \frac{\sum_{t=1}^T \gamma_t(i) o_t}{\sum_{t=1}^T \gamma_t(i)} \right)}{\sum_{t=1}^T \gamma_t(i) o_{t-1} \left(o_{t-1} - \frac{\sum_{t=1}^T \gamma_t(i) o_{t-1}}{\sum_{t=1}^T \gamma_t(i)} \right)}, \quad (20)$$

$$\zeta_i^{(n+1)} = \frac{\sum_{t=1}^T \gamma_t(i) o_t - \sum_{t=1}^T \gamma_t(i) o_{t-1} c_i^{(n+1)}}{\sum_{t=1}^T \gamma_t(i)}, \quad (21)$$

$$\sigma_i^{2(n+1)} = \frac{\sum_{t=1}^T \gamma_t(i) \left(o_t - c_i^{(n+1)} o_{t-1} - \zeta_i^{(n+1)} \right)^2}{\sum_{t=1}^T \gamma_t(i)}. \quad (22)$$

Repeat Step 2 and Step 3 until the log-likelihood function converges.

2.4. Prediction

Once the parameter is determined by the improved EM algorithm, the model can be employed to forecast the expected value of the next observation.

The conditional probability distribution of the observation o_{T+1} can be derived as

$$\begin{aligned} P(o_{T+1}|o_T, \dots, o_1, \lambda) &= \sum_{i=1}^N P(o_{T+1}|S_T = s_i, o_T, \dots, o_1, \lambda) P(S_T = s_i|o_T, \dots, o_1, \lambda) \\ &= \sum_{i=1}^N \sum_{j=1}^N P(o_{T+1}|S_{T+1} = s_j, o_T, \dots, o_1, \lambda) P(S_{T+1} = s_j|S_T = s_i, \lambda) \gamma_T(i) \\ &= \sum_{i=1}^N \sum_{j=1}^N \gamma_T(i) a_{ij} b_j(o_{T+1}|o_T, \dots, o_{T+1-d}). \end{aligned} \quad (23)$$

Therefore, the expectation of o_{T+1} is computed by

$$\hat{o}_{T+1} = \int o_{T+1} P(o_{T+1}|o_T, \dots, o_1, \lambda) do_{T+1}. \quad (24)$$

Given an observation sequence $\mathbf{o} = (o_1, \dots, o_T)$, $\hat{o}_t$ is predicted by

$$\hat{o}_t = \int o_t P(o_t|o_{t-1}, \lambda) do_t = \int o_t \sum_{i=1}^N \sum_{j=1}^N \gamma_{t-1}(i) b_j a_{ij} (o_t|o_{t-1}) do_t, t \geq d. \quad (25)$$

Thus, the predicted values of the observations can be denoted by $\hat{\mathbf{o}} = (o_1, \hat{o}_2, \dots, \hat{o}_T)$.

2.5. Performance Comparison

The mean squared error (MSE), absolute mean error (AME) and mean absolute percentage error (MAPE) are common tools for measuring, fitting and predicting accuracy [49]. Both MSE and AME values determine the average deviation between fitting values and original values, while MAPE provides a measurement for testing the relevant difference between them. In this study, we use MSE as the criterion to evaluate the models. The equation for MSE is:

$$\text{MSE} = \frac{1}{LT} \sum_{l=1}^L \sum_{t=1}^T (o_t - \hat{o}_t)^2, \quad (26)$$

where L is the number of predicted samples, and T is the length of each sample.

We assume that a variable from a production process follows a normal distribution with mean 100 and variance 25 when the process is under control, and that the observations are first-order autocorrelated with a correlation coefficient of 0.6. We use IHMM, HMM and AR(1) methods to predict observation values, respectively. The MSEs for the three models are 15.1276, 16.1867 and 15.7861, respectively. Since these MSEs are very close, we conclude that the predicted performances of the three approaches are similar. The predicted results of an observation sequence with a length of 50 from the in-control process are shown in Figure 4, from which we can see that the three models have close performances. However, the time taken for prediction using the three models are quite different. For the prediction of an observation sequence with a length of 50, IHMM is 14.6523 s, HMM is 93.6521 s, and AR(1) is almost instantaneous under the environment of win10 OS (Microsoft, Redmond, WA, USA) with CPU of Intel(R) Core(TM) i7-7500U (Santa Clara, CA, USA).

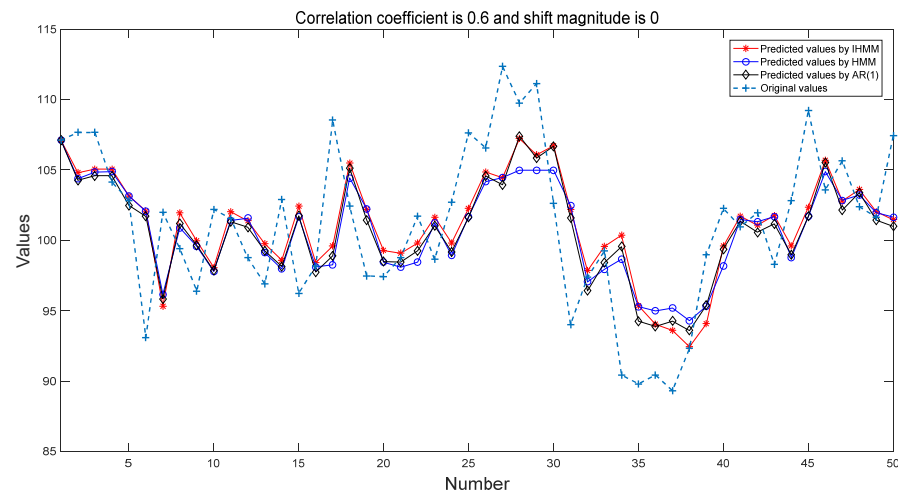


Figure 4. The predicted results of three models under an in-control process.

Then, we suppose the process has a shift magnitude of 3. We still use the three methods to predict observation values, respectively. The predicted results of an observation sequence with a length of 50 from the out-of-control process are shown in Figure 5, from which we can see distinctly different performances of the three models. By observing the distances between the lines with different colors, obviously, if the MSEs are calculated, the MSE from AR(1) is much less than that from IHMM, and the MSE from IHMM is much less than that from HMM. The IHMM only results in a medium-level performance in the prediction for the autocorrelated process. However, it is very interesting that the performances of corresponding residual charts have the best performances in detecting quality shifts. This seems to suggest that the residual charts integrating IHMM can achieve a surprising effect. This result is verified in Section 3.

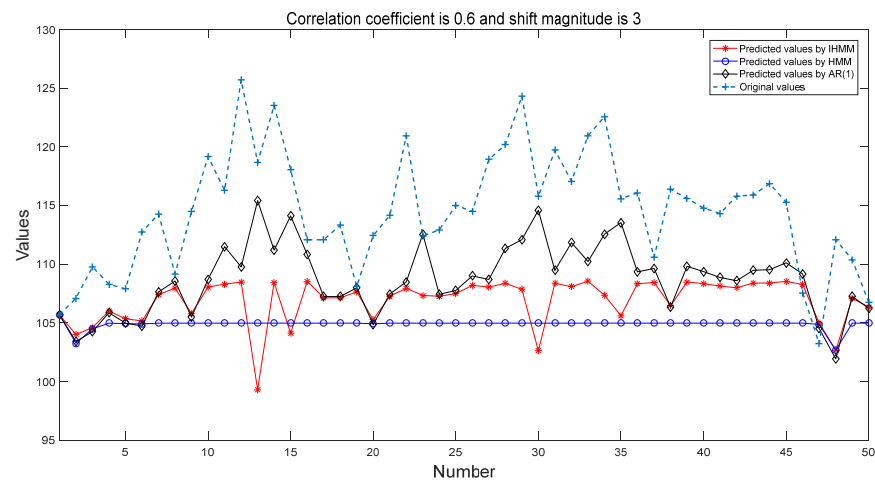


Figure 5. The predicted results of three models under an out-of-control process.

3. Statistical Process Control with Residual Chart

Residual control charts are an effective tool for online monitoring in the presence of autocorrelations. A residual chart called e chart is developed in our study.

Residuals are obtained by subtracting the predicted values of observations from the original values, that is $e = o - \hat{o} = (e_1, \dots, e_T)$. The control limits of the e chart are given by

$$UCL = \mu_e + k\sigma_e, \quad (27)$$

$$LCL = \mu_e - k\sigma_e. \quad (28)$$

where UCL represents the upper limit, while LCL represents the lower limit, k is the number of σ_e , μ_e represents the mean of e , and σ_e represents the standard deviation of e . μ_e and σ_e can be obtained by simulations based on sufficient samples.

If the value of e_t drops within UCL and LCL , the process is judged to be in control; otherwise, it is judged to be out of control.

4. Numerical Examples

We consider that the variable from a production process followed a normal distribution with mean 100 and variance 25 when the process is in control and that the observations were first-order autocorrelated. The correlation coefficient varies between -0.6 and 0.6 with increments of 0.3 . Two shift magnitudes of 1.5 and 3 are considered. According to the definition of residual charts, the ARLs of in-control processes for all predicted methods are 370 , so we focus our discussion on the out-of-control processes. By conducting multiple experiments, we find that it is appropriate to make the state number with 5 for both IHMM and HMM. The experimental results are shown in Figures 6 and 7.

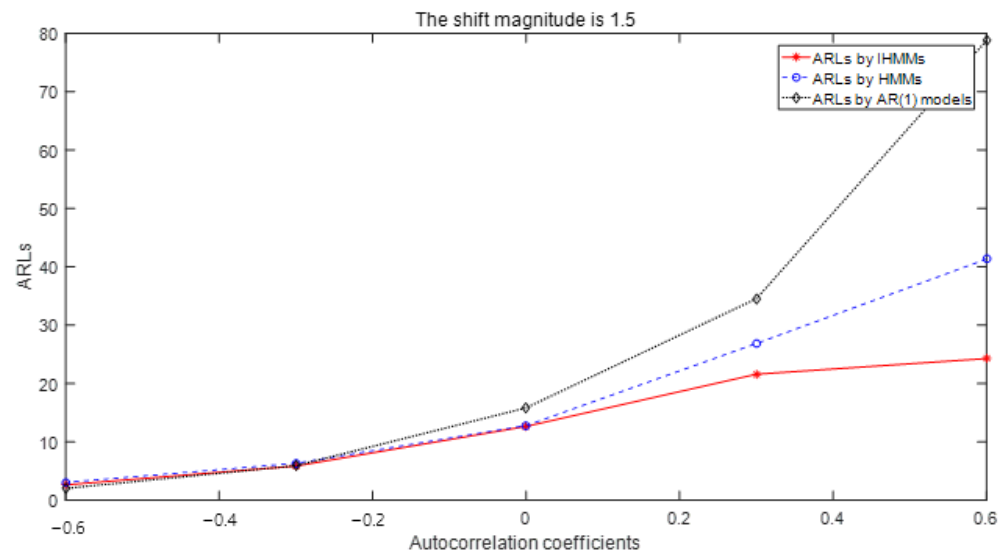


Figure 6. The ARLs of residual charts obtained by different models when the shift magnitude is 1.5.

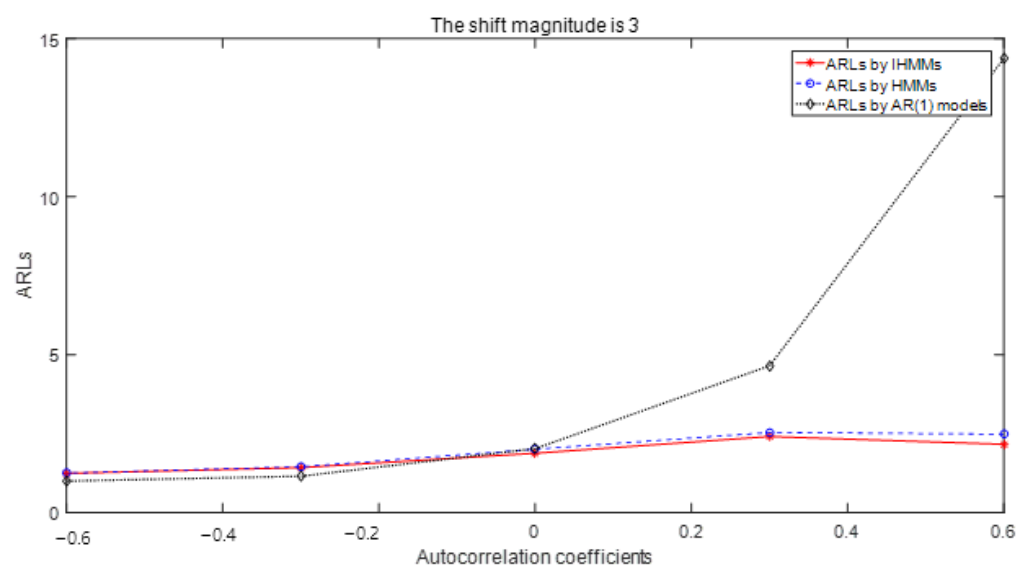


Figure 7. The ARLs of residual charts obtained by different models when the shift magnitude is 3.

As shown in Figures 6 and 7, when correlation coefficient changes from positive to negative, ARLs decrease dramatically, regardless of the approach used. Compared with

positive correlations, the ARLs of negative correlations are relatively very small, and ARLs obtained by different models are very close to each other. Thus, the following discussions focus on positive correlations.

As pointed out in Section 3, although the predictions of both the IHMM and HMM are inferior to AR(1) models, the performances of residual charts from the former models are much better than the latter ones. As seen in Figures 4 and 5, when the coefficients are larger than zero, the ARLs by IHMMs are shorter than those by HMMs, and by HMMs shorter than by AR(1) models.

As correlation coefficients increase, the ARLs generally increase, regardless of the approach used, especially as the shift magnitude decreases.

Generally speaking, when detecting quality shifts, the performances of IHMM, HMM and AR(1) models are ranked with IHMMs first, HMMs second and AR(1) last. Moreover, as pointed out in Section 2, the times taken by IHMMs are much shorter than HMMs under the same running environments.

5. Conclusions

In this paper, an IHMM with autocorrelated observations and a new EM algorithm are proposed. Residual charts in conjunction with the IHMM are employed for detecting quality shifts. The results demonstrate that: (1) the IHMM outperforms the HMM and AR(1) method with positive correlations; (2) the IHMM has similar performances with the HMM and AR(1) methods with negative correlations; (3) compared with positive correlations, the ARLs of the IHMM under negative correlations are relatively very small, as well as those of the HMMs and AR(1) models; (4) the IHMMs take a much shorter time than HMMs, for both training and prediction, but still longer than the AR(1) models.

Future research might focus on further experimental validations for the IHMM and its algorithm. The strict Gaussian distribution of observations could be extended to other probability distributions. Since multistage systems are commonplace in the manufacturing industry, it is worth extending this approach in this research direction.

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

Funding: This research was funded by the National Natural Science Foundation of China, grant number 72171120, 71701098, 72001138 and Qing Lan Project of Jiangsu Province in China, grant number 2021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Ouyang, L.; Chen, J.; Park, C.; Ma, Y.; Jin, J. Bayesian closed-loop robust process design considering model uncertainty and data quality. *IIE Trans.* **2020**, *52*, 288–300. [\[CrossRef\]](#)
2. Ouyang, L.; Zheng, W.; Zhu, Y.; Zhou, X. An interval probability-based FMEA model for risk assessment: A real-world case. *Qual. Reliab. Eng. Int.* **2020**, *36*, 125–143. [\[CrossRef\]](#)
3. Montgomery, D.C.; Mastrangelo, C.M. Some Statistical Process Control Methods for Autocorrelated Data. *J. Qual. Technol.* **1991**, *23*, 179–204. [\[CrossRef\]](#)
4. Maragah, H.D.; Woodall, W.H. The Effect of Autocorrelation on the Retrospective X-chart. *J. Stat. Comput. Simul.* **1992**, *40*, 29–42. [\[CrossRef\]](#)
5. Runger, G.C. Assignable Causes and Autocorrelation: Control Charts for Observations or Residuals? *J. Qual. Technol.* **2002**, *34*, 165–170. [\[CrossRef\]](#)

6. Franco, B.C.; Celano, G.; Castagliola, P.; Costa, A.F.B.; Machado, M.A.G. A New Sampling Strategy for the Shewhart Control Chart Monitoring A Process with Wandering Mean. *Int. J. Prod. Res.* **2015**, *53*, 4231–4248. [\[CrossRef\]](#)
7. Kim, J.; Jeong, M.K.; Elsayed, E.A. Monitoring Multistage Processes with Autocorrelated Observations. *Int. J. Prod. Res.* **2017**, *55*, 2385–2396. [\[CrossRef\]](#)
8. Yang, H.H.; Huang, M.L.; Lai, C.M.; Jin, J.R. An Approach Combining Data Mining and Control Charts-Based Model for Fault Detection in Wind Turbines. *Renew. Energy* **2018**, *115*, 808–816. [\[CrossRef\]](#)
9. Li, Y.; Pan, E.; Xiao, Y. On autoregressive model selection for the exponentially weighted moving average control chart of residuals in monitoring the mean of autocorrelated processes. *Qual. Reliab. Eng. Int.* **2020**, *36*, 2351–2369. [\[CrossRef\]](#)
10. Montgomery, D.C. *Statistical Quality Control: A Modern Introduction*, 6th ed.; John Wiley & Sons, Inc.: New York, NY, USA, 2009.
11. Vasilopoulos, A.V.; Stamboulis, A.P. Modification of Control Chart Limits in the Presence of Data Correlation. *J. Qual. Technol.* **1978**, *20*, 20–30. [\[CrossRef\]](#)
12. Wardell, D.G.; Moskowitz, H.; Plante, R.D. Control Charts in the Presence of Data Autocorrelation. *Manag. Sci.* **1992**, *38*, 1084–1105. [\[CrossRef\]](#)
13. Yashchin, E. Performance of CUSUM Control Schemes for Serially Correlated. *Technometrics* **1993**, *35*, 37–52. [\[CrossRef\]](#)
14. Schmid, W. On the Run Length of a Shewhart Chart for Correlated Data. *Stat. Pap.* **1995**, *36*, 111–130. [\[CrossRef\]](#)
15. Jiang, W.; Tsui, K.L.; Woodall, W.H. A New SPC Monitoring Method: The ARMA Chart. *Technometrics* **2000**, *42*, 399–410. [\[CrossRef\]](#)
16. Lu, C.W.; Reynolds, M.R., Jr. CUSUM Charts for Monitoring an Autocorrelated Process. *J. Qual. Technol.* **2001**, *33*, 316–334. [\[CrossRef\]](#)
17. Castagliola, P.; Tsung, F. Auto-correlated Statistical Process Control for Non-Normal Situations. *Qual. Reliab. Eng. Int.* **2005**, *21*, 131–161. [\[CrossRef\]](#)
18. Osei-Aning, R.; Abbasi, S.A.; Riaz, M. Optimization Design of the CUSUM and EWMA Charts for Autocorrelated Processes. *Qual. Reliab. Eng. Int.* **2017**, *33*, 1827–1841. [\[CrossRef\]](#)
19. Garza-Venegas, J.A.; Tercero-Gomez, V.G.; Ho, L.L.; Castagliola, P.; Celano, G. Effect of Autocorrelation Estimators on the Performance of the (X)over-bar Control Chart. *J. Stat. Comput. Simul.* **2018**, *88*, 2612–2630. [\[CrossRef\]](#)
20. Ryan, T.P. Discussion of Some Statistical Process Control Methods for Autocorrelated Data by D.C. Montgomery and C.M. Mastrangelo. *J. Qual. Technol.* **1991**, *23*, 200–202. [\[CrossRef\]](#)
21. Wardell, D.G.; Moskowitz, H.; Plante, R.D. Run-length Distributions of Special-cause Control Charts for Correlated Process. *Technometrics* **1994**, *36*, 3–27. [\[CrossRef\]](#)
22. Mastrangelo, C.M.; Montgomery, D.C. SPC with Correlated Observations for the Chemical and Process Industries. *Int. J. Reliab. Qual. Saf. Eng.* **1995**, *11*, 79–89. [\[CrossRef\]](#)
23. Zhang, N.F. Detection Capability of Residual Control Chart for Stationary Process Data. *J. Appl. Stat.* **1997**, *24*, 363–380. [\[CrossRef\]](#)
24. Lu, C.W.; Reynolds, M.R., Jr. EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes. *J. Qual. Technol.* **1999**, *31*, 166–188. [\[CrossRef\]](#)
25. Davoodi, M.; Niaki, S.T.A. Estimating the Step Change Time of the Location Parameter in Multistage Processes Using MLE. *Qual. Reliab. Eng. Int.* **2012**, *28*, 843–855. [\[CrossRef\]](#)
26. Perez-Rave, J.; Munoz-Giraldo, L.; Correa-Morales, J.C. Use of Control Charts with Regression Analysis for Autocorrelated Data in the Context of Logistic Financial Budgeting. *Comput. Ind. Eng.* **2017**, *112*, 71–83. [\[CrossRef\]](#)
27. Pan, X.; Jarrett, J. Using Vector Autoregressive Residuals to Monitor Multivariate Processes in the Presence of Serial Correlation. *Int. J. Prod. Econ.* **2007**, *106*, 204–216. [\[CrossRef\]](#)
28. Hwang, H.B.; Wang, Y. Shift Detection and Source Identification in Multivariate Autocorrelated Processes. *Int. J. Prod. Res.* **2010**, *48*, 835–859. [\[CrossRef\]](#)
29. Vanhatalo, E.; Kulahci, M. The Effect of Autocorrelation on the Hotelling T-2 Control Chart. *Qual. Reliab. Eng. Int.* **2015**, *31*, 1779–1796. [\[CrossRef\]](#)
30. Leoni, R.C.; Machado, G.; Aparecida, M.; Costa, B.; Fernando, A. The T-2 Chart with Mixed Samples to Control Bivariate Autocorrelated Processes. *Int. J. Prod. Res.* **2016**, *54*, 3294–3310. [\[CrossRef\]](#)
31. Pan, J.N.; Li, C.I.; Wu, J.J. A New Approach to Detecting the Process Changes for Multistage Systems. *Expert Syst. Appl.* **2016**, *62*, 293–301. [\[CrossRef\]](#)
32. Yang, S.F.; Yang, C.M. An Approach to Controlling Two Dependent Process Steps with Autocorrelated Observations. *Int. J. Adv. Manuf. Technol.* **2006**, *29*, 170–177. [\[CrossRef\]](#)
33. Li, Y.; Zio, E.; Pan, E. An MEWMA-Based Segmental Multivariate Hidden Markov Model for Degradation Assessment and Prediction. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2021**, *235*, 831–844. [\[CrossRef\]](#)
34. Xia, T.; Sun, B.; Chen, Z.; Pan, E.; Wang, H.; Xi, L. Opportunistic maintenance policy integrating leasing profit and capacity balancing in service-oriented manufacturing. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107233. [\[CrossRef\]](#)
35. Xia, T.; Dong, Y.; Pan, E.; Zheng, M.; Wang, H.; Xi, L. Fleet-level opportunistic maintenance for large-scale wind farms integrating real-time prognostic updating. *Renew. Energy* **2021**, *163*, 1444–1454. [\[CrossRef\]](#)
36. Xia, T.; Zhuo, P.; Xiao, L.; Du, S.; Wang, D.; Xi, L. Multi-stage Fault Diagnosis Framework for Rolling Bearing Based on OHF Elman AdaBoost-Bagging Algorithm. *Neurocomputing* **2021**, *433*, 237–251. [\[CrossRef\]](#)

37. Tang, D.; Yu, J.; Chen, X.; Makis, V. An optimal condition-based maintenance policy for a degrading system subject to the competing risks of soft and hard failure. *Comput. Ind. Eng.* **2015**, *83*, 100–110. [\[CrossRef\]](#)
38. Tang, D.; Makis, V.; Jafari, L.; Yu, J. Optimal maintenance policy and residual life estimation for a slowly degrading system subject to condition monitoring. *Reliab. Eng. Syst. Saf.* **2015**, *134*, 198–207. [\[CrossRef\]](#)
39. Chiu, C.C.; Chen, M.K.; Lee, K.M.S. Shifts Recognition in Correlated Process Data Using a Neural Network. *Int. J. Syst. Sci.* **2001**, *32*, 137–143. [\[CrossRef\]](#)
40. Arkat, J.; Niaki, S.T.A.; Abbasi, B. Artificial Neural Networks in Applying MCUSUM Residuals Charts for AR(1) Processes. *Appl. Math. Comput.* **2007**, *189*, 1889–1901. [\[CrossRef\]](#)
41. Pacella, M.; Semeraro, Q. Using Recurrent Neural Networks to Detect Changes in Autocorrelated Processes for Quality Monitoring. *Comput. Ind. Eng.* **2007**, *52*, 502–520. [\[CrossRef\]](#)
42. Camargo, M.E.; Priesnitz, W.; Russo, S.L.; Dullius, A.I.D. Control Charts for Monitoring Autocorrelated Processes Based on Neural Networks Model. In Proceedings of the International Conference on Computers and Industrial Engineering, Troyes, France, 6–9 July 2009; pp. 1881–1884.
43. Yang, H.H.; Huang, M.L.; Yang, S.W. Integrating Auto-Associative Neural Networks with Hotelling T-2 Control Charts for Wind Turbine Fault Detection. *Energies* **2015**, *8*, 12100–12115. [\[CrossRef\]](#)
44. Rai, A.; Upadhyay, S.H. The Use of MD-CUMSUM and NARX Neural Network for Anticipating the Remaining Useful Life of Bearings. *Measurement* **2017**, *111*, 397–410. [\[CrossRef\]](#)
45. Lee, S.; Li, L.; Ni, J. Online Degradation Assessment and Adaptive Fault Detection Using Modified Hidden Markov Model. *J. Manuf. Sci. Eng.-Trans. ASME* **2010**, *132*, 021010. [\[CrossRef\]](#)
46. Alshraideh, H.; Runger, G. Process Monitoring Using Hidden Markov Models. *Qual. Reliab. Eng. Int.* **2014**, *30*, 1379–1387. [\[CrossRef\]](#)
47. Ross, S.M. *Introduction to Probability Models*, 11th ed.; Posts & Telecom Press: Beijing, China, 2015.
48. Chen, Z.; Xia, T.B.; Li, Y.; Pan, E.S. Degradation Modeling and Classification of Mixed Populations Using Segmental Continuous Hidden Markov Models. *Qual. Reliab. Eng. Int.* **2018**, *34*, 807–823. [\[CrossRef\]](#)
49. Guo, X.J.; Liu, S.F.; Wu, L.F.; Gao, Y.B.; Yang, Y.J. A Multi-variable Grey Model with a Self-memory Component and Its Application on Engineering Prediction. *Eng. Appl. Artif. Intell.* **2015**, *42*, 82–93. [\[CrossRef\]](#)