


# A deep mixed-effects modeling approach for real-time monitoring of metal additive manufacturing process

Ruiyu Xu, Song Huang, Zheren Song, Yuanyuan Gao & Jianguo Wu


To cite this article: Ruiyu Xu, Song Huang, Zheren Song, Yuanyuan Gao & Jianguo Wu (2023): A deep mixed-effects modeling approach for real-time monitoring of metal additive manufacturing process, IIE Transactions, DOI: [10.1080/24725854.2023.2192252](https://doi.org/10.1080/24725854.2023.2192252)



To link to this article: <https://doi.org/10.1080/24725854.2023.2192252>

 View supplementary material 

 Published online: 17 Apr 2023.

 Submit your article to this journal 



 Article views: 167

 View related articles 

 View Crossmark data 



# A deep mixed-effects modeling approach for real-time monitoring of metal additive manufacturing process

Ruiyu Xu , Song Huang, Zheren Song, Yuanyuan Gao, and Jianguo Wu 

Department of Industrial Engineering and Management, Peking University, Beijing, China

## ABSTRACT

*In-situ* process monitoring of additive manufacturing (AM) is critically important and has been attracting significant efforts to achieve desirable process reliability and quality consistency. Despite enormous progress in embedding various sensors into the AM system, effectively fusing these sensor data for anomaly detection is very challenging, due to complex relationship among sensing signals, process condition and environment. In this article we propose a deep mixed-effects modeling approach to monitor the melt pool temperature for anomaly detection. It consists of a deep neural network (DNN) capturing the relationship between the temperature and other sensing data, a random-effect term accounting for the random variation of mean temperature, and a residual term modeling potential autocorrelations. To train the model (Phase I), an efficient hard expectation-maximization (EM) algorithm is developed, which iteratively optimizes the DNN parameters and the ones for the random-effect and residual terms. In online monitoring (Phase II), a  $T^2$ -based and a generalized likelihood ratio (GLR) test-based control charts are developed to timely detect the process anomalies. The asymptotic behaviors of both the  $T^2$  and GLR statistics are further established. The effectiveness of the proposed approach is demonstrated through thorough simulation study and real case study of an AM process.

## ARTICLE HISTORY

Received 30 June 2022  
Accepted 10 March 2023

## KEYWORDS

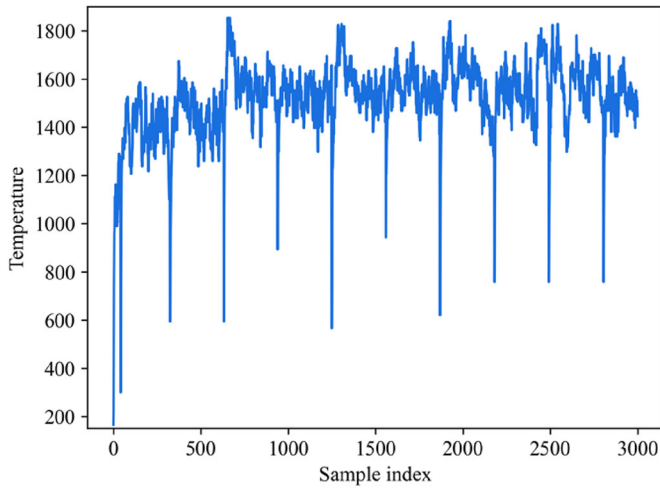
Additive manufacturing;  
mixed-effects model; deep  
neural network; hard EM;  
profile monitoring

## 1. Introduction

Additive Manufacturing (AM) is one of the most promising emerging manufacturing technologies and has gained significant application in many industrial sectors, such as aviation, aerospace and die manufacturing. Compared with traditional manufacturing technologies, AM technology offers numerous advantages, including the capability of producing sophisticated and customized components, reduced production time and production costs (Ko *et al.*, 2015; Chua *et al.*, 2017). Despite these advances, the lack of quality assurance procedures remains one of the most significant barriers to the widespread adoption of AM technology (Everton *et al.*, 2016). Since post-processing operations, such as surface finishing and heat treatment, cannot remove internal defects, significant efforts have been expended by funding agencies and researchers on developing *in-situ* process monitoring for defect detection and mitigation in the early processing stage (Song and Mazumder, 2011; Tapia and Elwany, 2014; Waller *et al.*, 2014; Rao *et al.*, 2015). With a well-established monitoring system, closed-loop control systems can be developed for defect mitigation by the adjustment of process parameters (Kruth *et al.*, 2007; Renken *et al.*, 2019; Vasileška *et al.* 2020), layer re-melting (Demir and Previtali, 2017; Heeling and Wegener, 2018; Jalalahmadi *et al.*, 2019) and grinding (Colosimo *et al.*, 2020).

Laser Powder Bed Fusion (L-PBF) processes are popular metal AM processes that utilize a high-energy laser beam to

fuse metal particles in the powder bed. They have attracted most of the research attention in the field of *in-situ* AM process monitoring (Grasso *et al.*, 2021). In L-PBF processes, various quality issues exist, such as pores, cracks, unfused powder and the balling of molten materials. The potential causes include improper process parameters, non-homogeneous powder deposition, out-of-control heat exchanges, etc. (Everton *et al.*, 2016; Grasso and Colosimo, 2017; Grasso and Colosimo, 2019). Many factors are involved in the formation of defects, such as energy density, build geometry, powder properties and chamber atmosphere (Grasso *et al.*, 2021). The mechanism of how these factors cause the defects is very complicated, due to complex physical reactions at the microscopic level. To provide real-time information of the manufacturing process, various sensors are nowadays embedded in the AM monitoring system, including light, heat, sound, and force signals (Mani *et al.*, 2015; Wu *et al.*, 2021). Since all metal-based AM processes utilize thermal energy to fuse metal powders, most of the existing studies have focused on thermal or temperature monitoring of the process using pyrometers (e.g., photodiodes, digital cameras) or thermocouples (Tapia and Elwany, 2014; Colosimo *et al.*, 2018). Some studies used off-axis mounted photodiodes to measure the radiation intensity of the entire build area (Bisht *et al.*, 2018; Nadipalli *et al.*, 2019). In most studies, photodiodes and cameras are placed co-axially to the laser



**Figure 1.** The melt pool temperature of the first 3000 measurements in a normal layer.

path to gather the temperature information, such as the intensity and spectrum of emitted radiation.

Despite great progress in collecting *in-situ* sensing signals, there is still a lack of effective modeling and monitoring approaches for real-time anomaly detection, especially for the photodiode signals. Most of the existing methods are either for quality prediction of the whole build instead of *in-situ* anomaly detection, or simply using photodiode signals alone to detect anomalies. For example, Alberts *et al.* (2017) constructed a monitoring system based on two different photodiodes and the ratio of the two photodiode signals is used as a quality assurance tool to control the building density. Okaro *et al.* (2019) used Random Singular Value Decomposition (R-SVD) to process photodiode signals, and adopted a semi-supervised learning method with a small number of labeled samples to identify the product quality. Renken *et al.* (2019) combined a closed-loop control strategy with a feedforward approach to fast control the melt pool temperature, based on the modeling of the relationship between the laser power and the photodiode signals. Forien *et al.* (2020) correlated pore locations measured by X-ray radiography with pyrometer signals to predict the probability of defect formation. Jayasinghe *et al.* (2022) extracted features from photodiode signals via SVD and employed Gaussian Process regression to predict build density.

The modeling and monitoring of the melt pool temperature is still very challenging, due to the following reasons. First, the melt pool temperature is highly dynamic, showing complex nonlinear relationships with other signals that may continuously change. Figure 1 shows a typical melt pool temperature profile of a normal layer showing high randomness. It periodically varies in the range from 200 to 1800 due to shifts in the scanning path. A sharp temperature rise is seen at the beginning of each scan trace, which then decreases without a clear pattern. The existing linear profile monitoring methods (Zhang *et al.*, 2009) is not capable of modeling such nonlinear relationships. The common nonlinear profile monitoring methods, such as local linear kernel estimation (Zou *et al.* 2009; Qiu and Zou, 2010; Qiu *et al.*, 2010), b-spline method (Chuang *et al.*, 2013), and p-spline

method (Chen *et al.*, 2015), are mainly designed for univariate functions, i.e., one explanatory variable (time  $t$  or other physical variable). Although these methods can be applied to multivariate cases, they are not flexible enough to capture the complex nonlinear relationship in AM processes. The existing regression-adjusted charts (or risk-adjusted charts in healthcare), which often employ Poisson regression or logistic regression (Sachlas *et al.*, 2019; Wen *et al.*, 2021), also face the challenge of modeling the arbitrarily nonlinear relationship between profile and exogenous variables.

Second, the photodiode signals are highly auto-correlated. Due to some unobserved or unknown factors, the residuals after regression adjustment may still be auto-correlated. The mixed-effects models can model autocorrelation within the profile. However, this type of autocorrelation is due to the mean, not the residuals. In their models, the residuals are assumed to be independent and identically distributed random errors, which might not be applicable to photodiode signals. Third, to timely detect the anomalies, the temperature profile should be monitored incrementally, i.e., once a new measurement is collected, the monitoring method should judge whether this profile is out of control (OC). Therefore, it is a sequential test problem (Wald, 1992). However, almost all the existing methods are for monitoring the whole profile, which cannot be directly applied to our case. Besides, as the scanning is very fast, a computationally efficient monitoring algorithm whose computation time does not depend on the profile length is highly desirable.

To fill this gap, we propose a deep mixed-effects modeling approach to monitor the melt pool temperature. Specifically, a Deep Neural Network (DNN) is proposed to better capture the nonlinear relationship between the temperature, process parameters and other sensing data. Compared with the existing nonlinear methods, DNN has a stronger capability of modeling complex nonlinearity. Besides, a random-effect term is added to model the random variation of the mean temperature that cannot be explained by the explanatory variables among different printing layers. Moreover, a correlated residual term is used to model the potential correlations. However, the coupling of deep learning and mixed-effects modeling makes the model training very difficult. To solve this problem, an efficient hard EM algorithm is developed, which iteratively optimizes the DNN parameters and the ones for random-effects term and residual term. In Phase II, two control charts, i.e., a  $T^2$ -control chart and a Generalized Likelihood Ratio (GLR) control chart are developed to detect the process anomalies. The asymptotic behaviors of the  $T^2$  and GLR statistics with and without anomalies are further established, based on which the dynamic control limits are designed. A dynamic updating strategy is proposed to reduce the computational cost. The effectiveness of the proposed approach is demonstrated through simulation studies and a real case study of a Selective Laser Melting (SLM)-based AM process.

The rest of this article is organized as follows. The model formulation and Phase I optimization algorithm are provided in Section 2. In Section 3, two Phase II control charts are developed and the asymptotic properties of the test

statistics are established. The comprehensive performance of each of the proposed methods is investigated through simulation in Section 4. A real case study of an AM process demonstrates the effectiveness of our proposed methods in Section 5. Section 6 provides a brief conclusion and discussion of this study.

## 2. Deep mixed-effects modeling

In this section, we first introduce the proposed deep mixed-effects model for the melt pool temperature in the normal working condition, and then develop an efficient hard EM algorithm for optimizing the model parameters.

### 2.1. Model formulation

In AM processes, the observed photodiode signals in different printing layers can be regarded as longitudinal data, and the temperature monitoring can be treated as a profile monitoring problem with temperature as the response variable and other observed data as the explanatory variables. In what follows, we use the term “profile” to represent the measured photodiode signals of each layer.

Suppose that there are  $m$  profiles and the  $i$ th profile has  $n_i$  observations for  $i = 1, 2, \dots, m$ . Let  $y_{ij}$  be the  $j$ th observation in the  $i$ th profile and  $\mathbf{x}_{ij}$  be the corresponding explanatory variables for  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n_i$ . These observations are in a temporal order, where profile  $i$  is started at time  $t_i$ ,  $t_1 < t_2 < \dots < t_m$ , and  $(\mathbf{x}_{ij}, y_{ij})$  is the  $j$ th observation at time  $t_{ij}$ ,  $t_i = t_{i1} < t_{i2} < \dots < t_{in_i}$ . We propose the following deep mixed-effects model

$$y_{ij} = g(\mathbf{x}_{ij}) + \xi_i + \gamma_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, n_i, \quad (1)$$

where  $g(\cdot)$  is the population profile function (the fixed-effects term) capturing the complex relation between profile data and other explanatory variables,  $\xi_i$  is the random-effects term modeling the between-profile variation of the  $i$ th individual profile from  $g(\cdot)$ , and  $\gamma_{ij}$  is the residual term considering the within-profile correlation. It is assumed that  $\xi_i$  and  $\gamma_{ij}$  are independent of each other and  $\xi_i$ 's follow an independent and identically distributed (i.i.d.) Gaussian distribution  $N(0, \sigma_\xi^2)$ . The correlation (covariance) between  $\gamma_{ij}$  and  $\gamma_{ik}$  is modeled by a correlation function  $\text{Corr}(\mathbf{x}_{ij}, t_{ij}; \mathbf{x}_{ik}, t_{ik})$  and a variance term  $\sigma_\gamma^2$  as follows:

$$\text{Cov}(\gamma_{ij}, \gamma_{ik}) = \sigma_\gamma^2 \text{Corr}(\mathbf{x}_{ij}, t_{ij}; \mathbf{x}_{ik}, t_{ik}). \quad (2)$$

The above correlation function is flexible to model various types of correlations. For instance, when data are spatially correlated, the correlation function can be set as  $\text{Corr}(\mathbf{x}_{ij}, t_{ij}; \mathbf{x}_{ik}, t_{ik}) = \rho(|\mathbf{x}_{ij} - \mathbf{x}_{ik}|, \boldsymbol{\beta})$  with a correlation function  $\rho(\cdot)$  and a coefficient vector  $\boldsymbol{\beta}$ . The i.i.d. measurement error is a special case of our framework where  $\text{Corr}(\mathbf{x}_{ij}, t_{ij}; \mathbf{x}_{ik}, t_{ik}) = 0$ . In the literature, there are some reported spatial-temporal modeling works for AM processes. Guo *et al.* (2020) proposed a multivariate first-order autoregressive process with a spatially correlated precision matrix

to recognize implicit anomalies in AM processes. Mahmoudi *et al.* (2019) and Yang *et al.* (2022) used Gaussian process models to capture spatial-temporal correlations in adjacent regions for characterizing and monitoring melt-pool variations. Liu *et al.* (2019) proposed an augmented layer-wise spatial log Gaussian Cox process to model and quantify the layer-wise spatial evolution of porosity. However, in the current article, only the average temperature of the melt pool is measured, which moves along the scan path. The temperatures of the adjacent regions on both sides of the scan path are unknown. Therefore, we only focus on the situation where data are temporally correlated, which is also a spatial correlation on the scan path. We set  $\text{Corr}(\mathbf{x}_{ij}, t_{ij}; \mathbf{x}_{ik}, t_{ik}) = \rho(|t_{ij} - t_{ik}|, \omega)$ , where the function  $\rho(\cdot)$  is set to an exponential function  $\rho(t, \omega) = \omega^t$ . Note that this is exactly the AutoCorrelation Function (ACF) of a first-order autoregressive model AR(1).

For  $g(\cdot)$ , although the existing local smoothing methods can be applied to a multivariate case, they are not as flexible as a DNN to capture complex a nonlinear relationship between profile data and other variables. To address this problem, we propose to use a DNN to model the population profile  $g(\cdot)$  due to its capability of approximating any nonlinear functions. For the AM process monitoring case, the structure of the DNN model is carefully customized by considering the balance between model accuracy and model complexity. We refer to the overall model as a Deep Mixed-Effects Model (DMEM).

### 2.2. Model optimization via hard EM algorithm

The model parameters to be optimized include the DNN parameters, denoted by  $\boldsymbol{\theta}$ , the random-effects variance  $\sigma_\xi^2$ , the noise parameters  $\sigma_\gamma^2$  and  $\omega$ . Let  $\boldsymbol{\Phi} = (\boldsymbol{\theta}, \sigma_\xi^2, \sigma_\gamma^2, \omega)$ . In this subsection, we propose a hard EM algorithm to optimize the model parameters.

Suppose the In Control (IC) historical dataset contains  $m_{IC}$  profiles. Based on the DMEM framework and the assumptions above, the observed data follow a Gaussian distribution

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix} \sim N \left( \begin{pmatrix} g(\mathbf{x}_{i1}) \\ \vdots \\ g(\mathbf{x}_{in_i}) \end{pmatrix}, \sigma_\xi^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}' + \sigma_\gamma^2 \mathbf{R}_i \right),$$

where  $\mathbf{1}_{n_i}$  is a  $n_i \times 1$  vector with all elements equal to one, and  $\mathbf{R}_i$  is the  $n_i \times n_i$  correlation matrix with element  $\rho_{jk} = \omega^{|j-k|}$ . Therefore, we can use the Maximum Likelihood Estimation (MLE) approach to estimate the model parameters, which is equivalent to minimizing the negative log-likelihood function

$$\min_{\boldsymbol{\Phi}} \sum_{i=1}^{m_{IC}} \frac{1}{2} (\boldsymbol{\zeta}_i' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\zeta}_i + \log |\boldsymbol{\Sigma}_i|), \quad (3)$$

where  $\boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{in_i})'$ ,  $\zeta_{ij} = y_{ij} - g(\mathbf{x}_{ij})$ , and  $\boldsymbol{\Sigma}_i = \sigma_\xi^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}' + \sigma_\gamma^2 \mathbf{R}_i$  is the covariance matrix.

To get optimal estimates of  $\theta, \sigma_\xi^2, \sigma_\gamma^2$  and  $\omega$ , a widely used approach is to iteratively optimize each of the parameters via coordinate descent algorithms. However, the complexity of  $\Sigma_i^{-1}$ , DNN and their coupling make the above optimization problem extremely difficult to solve. Another natural approach for a mixed-effects model is the Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977), where the random effects  $\xi = \{\xi_i, i = 1, \dots, m_{IC}\}$  can be treated as latent or unobserved variables. Let  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_{m_{IC}}\}$  denote all the historical IC profiles with  $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})'$ , and  $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_{m_{IC}}\}$  denote the data of explanatory variables with  $\mathbf{X}_i = \{\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}\}'$ . The E-step and M-step at  $l$ th iteration can be expressed as

$$\begin{aligned} \text{E-step: } Q(\Phi, \Phi^{(l-1)}) &= E_{\xi|\mathbf{Y}, \mathbf{X}, \Phi^{(l-1)}}(\log L(\Phi|\mathbf{Y}, \mathbf{X}, \xi)), \\ \text{M-step: } \Phi^{(l)} &= \underset{\Phi}{\operatorname{argmax}} \left\{ Q(\Phi, \Phi^{(l-1)}) \right\}, \end{aligned} \quad (4)$$

where  $\log L(\Phi|\mathbf{Y}, \mathbf{X}, \xi)$  is the complete data log-likelihood,  $Q(\Phi, \Phi^{(l-1)})$  is the expected  $\log L(\Phi|\mathbf{Y}, \mathbf{X}, \xi)$  with respect to posterior density  $p(\xi|\mathbf{Y}, \mathbf{X}, \Phi^{(l-1)})$ . Nevertheless,  $Q(\Phi, \Phi^{(l-1)})$  is very complex and intractable analytically. Besides, due to the introduction of the DNN model, the Monte Carlo EM algorithm (Wei and Tanner, 1990), where the Q-function is computed numerically through Monte Carlo simulation, still faces a significant challenge due to its high computational cost. To overcome this issue, we propose to use a “hard” version of the EM algorithm (Neal and Hinton, 1998; Ruggieri *et al.*, 2020), where the latent variables in the complete data log-likelihood are imputed by the maximum a posteriori (MAP) estimate. The rationale of using the hard EM algorithm is that the posterior probability of the latent variables  $\xi$  lies within a narrow range around the peak value, due to large sample sizes in IC profiles. With this strategy, the parameters can be effectively decoupled and optimized iteratively. Mathematically, the hard EM algorithm's  $l$ th iteration can be expressed as

$$\begin{aligned} \text{E-step: } \xi^{(l)} &= \underset{\xi}{\operatorname{argmax}} \left\{ \log p(\xi|\mathbf{Y}, \mathbf{X}, \Phi^{(l-1)}) \right\}, \\ \text{M-step: } \Phi^{(l)} &= \underset{\Phi}{\operatorname{argmax}} \left\{ \log L(\Phi|\mathbf{Y}, \mathbf{X}, \xi^{(l)}) \right\}. \end{aligned} \quad (5)$$

In the E-step, the posterior probability of latent variables can be obtained by

$$\begin{aligned} p(\xi|\mathbf{Y}, \mathbf{X}, \Phi^{(l-1)}) &\propto p(\xi, \mathbf{Y}|\mathbf{X}, \Phi^{(l-1)}) \\ &= p(\xi|\Phi^{(l-1)})p(\mathbf{Y}|\xi, \mathbf{X}, \Phi^{(l-1)}) \\ &= \left(2\pi\sigma_\xi^2\right)^{-\frac{m_{IC}}{2}} \exp\left(-\frac{\sum_{i=1}^{m_{IC}} \xi_i^2}{2\sigma_\xi^2}\right) \\ &\quad \times \prod_{i=1}^{m_{IC}} \left( (2\pi)^{-\frac{n_i}{2}} \left| \sigma_\gamma^2 \mathbf{R}_i^{(l-1)} \right|^{-\frac{1}{2}} \exp\left(-\frac{\hat{\mathbf{y}}_i' \mathbf{R}_i^{(l-1)-1} \hat{\mathbf{y}}_i}{2\sigma_\gamma^2}\right) \right), \end{aligned} \quad (6)$$

where  $\hat{\mathbf{y}}_i = (\hat{y}_{i1}, \dots, \hat{y}_{in_i})$ ,  $\hat{y}_{ij} = y_{ij} - g^{(l-1)}(\mathbf{x}_{ij}) - \xi_i$  is the estimation of the residuals. Therefore, the optimization problem of  $\xi^{(l)}$  can be expressed as

$$\xi^{(l)} = \underset{\xi}{\operatorname{argmin}} \frac{1}{\sigma_\xi^2} \sum_{i=1}^{m_{IC}} \xi_i^2 + \frac{1}{\sigma_\gamma^2} \sum_{i=1}^{m_{IC}} \hat{\mathbf{y}}_i' \mathbf{R}_i^{(l-1)-1} \hat{\mathbf{y}}_i. \quad (7)$$

The above optimization problem has an analytical solution

$$\xi_i^{(l)} = \frac{\frac{1}{\sigma_\gamma^2} \sum_j \sum_k (\mathbf{R}_i^{-1})_{ik}^{(l-1)} (y_{ij} - g^{(l-1)}(\mathbf{x}_{ij}))}{\frac{1}{\sigma_\gamma^2} \sum_j \sum_k (\mathbf{R}_i^{-1})_{ik}^{(l-1)} + \frac{1}{\sigma_\xi^2}}, \quad (8)$$

where  $(\mathbf{R}_i^{-1})_{ik}$  is the  $(j, k)$ th element of the  $\mathbf{R}_i^{-1}$ . In the M-step, the complete data log-likelihood is

$$\begin{aligned} L(\Phi|\mathbf{Y}, \mathbf{X}, \xi^{(l)}) &= p(\xi^{(l)}|\Phi) p(\mathbf{Y}|\Phi, \xi^{(l)}, \mathbf{X}) \\ &= \left(2\pi\sigma_\xi^2\right)^{-\frac{m_{IC}}{2}} \exp\left(-\frac{\sum_{i=1}^{m_{IC}} \xi_i^{(l)2}}{2\sigma_\xi^2}\right) \\ &\quad \times \prod_{i=1}^{m_{IC}} \left( (2\pi)^{-\frac{n_i}{2}} \left| \sigma_\gamma^2 \mathbf{R}_i \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_\gamma^2} \hat{\mathbf{y}}_i' \mathbf{R}_i^{-1} \hat{\mathbf{y}}_i \right) \right), \end{aligned} \quad (9)$$

where  $\hat{\mathbf{y}}_i = (\hat{y}_{i1}, \dots, \hat{y}_{in_i})$ ,  $\hat{y}_{ij} = y_{ij} - g(\mathbf{x}_{ij}) - \xi_i^{(l)}$ . The optimization problem of  $\Phi^{(l)}$  can be reformulated as

$$\Phi^{(l)} = \underset{\Phi}{\operatorname{argmin}} \sum_{i=1}^{m_{IC}} \left( \frac{1}{\sigma_\gamma^2} \hat{\mathbf{y}}_i' \mathbf{R}_i^{-1} \hat{\mathbf{y}}_i + \log \left| \sigma_\gamma^2 \mathbf{R}_i \right| + \frac{1}{\sigma_\xi^2} \xi_i^{(l)2} + \log \sigma_\xi^2 \right). \quad (10)$$

In the above problem, the four sets of parameters  $\theta, \sigma_\xi^2, \sigma_\gamma^2, \omega$  can be optimized iteratively via the block coordinate descent algorithm (Tseng, 2001).

In the first term which contains the DNN parameter  $\theta$ ,  $\mathbf{R}_i^{-1}$  is sparse and can be represented by only three parameters, say  $a, b, c$ :

$$\begin{aligned} \mathbf{R}_i &= \begin{pmatrix} 1 & \omega & \omega^2 & \dots & \omega^{n_i-1} \\ \omega & 1 & \omega & \dots & \omega^{n_i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^{n_i-1} & \omega^{n_i-2} & \omega^{n_i-3} & \dots & 1 \end{pmatrix}, \\ \mathbf{R}_i^{-1} &= \begin{pmatrix} a & b & 0 & 0 & \dots & 0 \\ b & c & b & 0 & \dots & 0 \\ 0 & b & c & b & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & b & a \end{pmatrix}, \end{aligned}$$

where

$$a = \frac{1}{1 - \omega^2}, \quad b = -\frac{\omega}{1 - \omega^2}, \quad c = \frac{1 + \omega^2}{1 - \omega^2}.$$

Due to this sparsity, the term  $\hat{\mathbf{y}}_i' \mathbf{R}_i^{-1} \hat{\mathbf{y}}_i$  can be calculated by  $\sum_{j=1}^{n_i} (c_{ij} \hat{y}_{ij}^2 + 2b \hat{y}_{ij} \hat{y}_{i,j-1})$  where  $\hat{y}_{i0}$  is defined as  $\hat{y}_{i0} = 0$ ,  $c_{ij} = a$  for  $j = 1, n_i$  and  $c$  for  $j = 2, \dots, n_i - 1$ . To avoid



cross-multiplication  $\hat{\gamma}_{ij}\hat{\gamma}_{i,j-1}$  and make the loss function of DNN independent among samples, a simple trick is proposed to replace  $\hat{\gamma}_{i,j-1}$  with an optimized value at the previous step,  $\hat{\gamma}_{i,j-1}^{(l-1)}$ , where  $\hat{\gamma}_{i,j-1}^{(l-1)} = y_{ij} - g^{(l-1)}(\mathbf{x}_{i,j-1}) - \xi_i^{(l-1)}$ . Therefore,  $\sum_{i=1}^{m_{IC}} \sum_{j=1}^{n_i} (c_{ij}\hat{\gamma}_{ij}^2 + 2b\hat{\gamma}_{ij}\hat{\gamma}_{i,j-1}^{(l-1)})$  is taken as the loss function of DNN to simply the optimization process for  $\theta^{(l)}$ .

For  $\sigma_{\xi}^{2(l)}$  and  $\sigma_{\gamma}^{2(l)}$ , they can be analytically updated:

$$\sigma_{\xi}^{2(l)} = \frac{1}{m_{IC}} \sum_{i=1}^{m_{IC}} \xi_i^{(l)^2}, \sigma_{\gamma}^{2(l)} = \frac{1}{\sum_i n_i} \left( \sum_i \hat{\gamma}_i' \mathbf{R}_i^{-1} \hat{\gamma}_i \right), \quad (11)$$

where  $\hat{\gamma}_i = (\hat{\gamma}_{i1}, \dots, \hat{\gamma}_{ini})$ ,  $\hat{\gamma}_{ij} = y_{ij} - g^{(l)}(\mathbf{x}_{ij}) - \xi_i^{(l)}$ . In the optimization of  $\omega^{(l)}$ , it is difficult to obtain an analytical solution. However, since the determinant  $|\mathbf{R}_i| = (1 - \omega^2)^{n_i-1}$ , we can easily update the estimates in a numerical way. To judge whether the algorithm converges, we can use the relative difference of the successive estimates of  $\xi, \sigma_{\gamma}^2, \omega$  as the convergence criterion. The stop conditions of the proposed hard EM algorithm are set as

$$\max_i \left\{ \frac{|\xi_i^{(l)} - \xi_i^{(l-1)}|}{|\xi_i^{(l)}|} \leq \epsilon_1, \frac{|\sigma_{\gamma}^{2(l)} - \sigma_{\gamma}^{2(l-1)}|}{|\sigma_{\gamma}^{2(l)}|} \leq \epsilon_2, \frac{|\omega^{(l)} - \omega^{(l-1)}|}{|\omega^{(l)}|} \leq \epsilon_3 \right\}, \quad (12)$$

where  $\epsilon_1, \epsilon_2, \epsilon_3$  are pre-specified small positive numbers (e.g.,  $10^{-2}$ ). The hard EM optimization algorithm is summarized in Algorithm 1.

---

**Algorithm 1:** The hard EM algorithm for DMEM model training

---

Input: IC Dataset  $(\mathbf{x}_{ij}, y_{ij})$ ,  $i = 1, \dots, m_{IC}, j = 1, \dots, n_i$

Initialize:  $\xi^{(0)}$  and  $\Phi^{(0)}$

Iterate for  $l = 1, \dots$

E-step:

Update  $\xi^{(l)}$  by Eq. (8).

M-step:

Optimize  $\theta^{(l)}$  by training the DNN model.

Update  $\sigma_{\xi}^{2(l)}, \sigma_{\gamma}^{2(l)}$  and  $\omega^{(l)}$  by Equations (11).

Stop if (12) are satisfied.

---

### 3. Online monitoring for anomaly detection

In this section, we develop an online monitoring scheme to detect anomalies under the conventional assumption that the IC model parameters are known in Phase II analysis. In the past few decades, various control charts have been developed for profile monitoring. These profile monitoring methods are mainly for monitoring the change of population profile function. Generally, they can be classified into three categories based on the quantity used to construct the control charts: (i) using the estimated model parameters, e.g., coefficients of a linear profile (Kim *et al.*, 2003; Zou *et al.*, 2007; Jensen *et al.*, 2008; Zhang *et al.*, 2009), (ii) using the estimated confidence band of observed data via resampling methods (Hung *et al.*, 2012; Chuang *et al.*, 2013), and (iii) using the residuals of the profile

or deviations from normal profile at a fixed set of design points (Qiu *et al.*, 2010; Wang *et al.*, 2018). However, all the three types are not for incremental monitoring. In almost all of these methods, the whole profile is treated as an instance, and the control charts are designed to test whether this profile is OC. However, in our case, the length of the profile dynamically increases, and the objective is to detect the anomalous profile as soon as possible based on the observations of the profile at hand. In addition, it is easy to see that type (i) and (ii) cannot be applied to our DMEM model, since they are for simple univariate profiles. Therefore, we turn to monitor the residuals of the profile  $\xi_{i,1:t} = (\xi_{i1}, \dots, \xi_{it})$ ,  $\xi_{ij} = y_{ij} - g^{(IC)}(\mathbf{x}_{ij})$ , using the fixed IC parameters  $\theta^{(IC)}$  of DNN. In a conventional profile monitoring setting, it is often assumed that only the population profile function changes, while the other parameters are fixed in the OC process (Qiu *et al.*, 2010). In such a case, for any profile  $i$  in monitoring, the detection of anomalies is equivalent to the following testing problem

$$H_0 : g_i = g^{(IC)} \text{ and } H_1 : g_i = g^{(IC)} + g_{\delta}, g_{\delta} \neq 0.$$

The residuals  $\xi_{i,1:t} \sim N(g(\mathbf{x}_{i,1:t}), \Sigma_{1:t}^{(IC)})$ , where  $\Sigma_{1:t}^{(IC)} = \sigma_{\xi}^{2(IC)} \mathbf{1}_t \mathbf{1}_t' + \sigma_{\gamma}^{2(IC)} \mathbf{R}_{1:t}^{(IC)}$ ,  $g_{\delta}(\mathbf{x}_{i,1:t}) = 0$  for IC profiles and  $g_{\delta}(\mathbf{x}_{i,1:t}) = [g_{\delta}(\mathbf{x}_{i1}), \dots, g_{\delta}(\mathbf{x}_{it})]' \neq 0$  for OC profiles. Naturally, we can construct the Hotelling  $T^2$  control chart or  $\chi^2$  control chart to monitor the change of  $g(\cdot)$ . It is directionally invariant, i.e., the detection power only depends on the shift magnitude (measured by Mahalanobis distance), not in its directions (Montgomery, 2007). However, in practical applications, the model parameters for  $\xi_i + \gamma_{ij}$  in (1) may also change, leading to an unsatisfactory detection power for the  $T^2$  control chart. Therefore, in addition to the  $T^2$  control chart, we also establish a GLR-based control chart to detect the shift of other parameters.

#### 3.1. $T^2$ control chart

For any IC profile  $i$ , at time  $t$  with  $\mathbf{x}_{i,1:t} = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{it})$  and  $y_{i,1:t} = (y_{i1}, \dots, y_{it})$  observed,

$$y_{i,1:t} \sim N(g^{(IC)}(\mathbf{x}_{i,1:t}), \Sigma_{1:t}^{(IC)}),$$

where  $\Sigma_{1:t}^{(IC)} = \sigma_{\xi}^{2(IC)} \mathbf{1}_t \mathbf{1}_t' + \sigma_{\gamma}^{2(IC)} \mathbf{R}_{1:t}^{(IC)}$  and

$$\mathbf{R}_{1:t}^{(IC)} = \begin{pmatrix} 1 & \omega^{(IC)} & \omega^{(IC)^2} & \dots & \omega^{(IC)^{t-1}} \\ \omega^{(IC)} & 1 & \omega^{(IC)} & \dots & \omega^{(IC)^{t-2}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^{(IC)^{t-1}} & \omega^{(IC)^{t-2}} & \omega^{(IC)^{t-3}} & \dots & 1 \end{pmatrix}.$$

A multivariate  $T^2$  chart can thus be designed, where the  $T^2$  statistic is

$$T^2(t) = \frac{\xi_{i,1:t}' \left( \Sigma_{1:t}^{(IC)} \right)^{-1} \xi_{i,1:t}}{t}. \quad (13)$$

Note that in  $T^2(t)$ , the Mahalanobis distance is divided by the dimension  $t$  to avoid a dynamically increasing control

limit. **Theorem 1** shows some obvious results regarding the distribution of  $T^2(t)$  and its asymptotic property. The proof of **Theorem 1** is provided in Appendix A.

**Theorem 1.** For any IC profile  $i$  in monitoring,

$$\zeta'_{i,1:t} \left( \Sigma_{1:t}^{(IC)} \right)^{-1} \zeta_{i,1:t} \sim \chi^2(t), T^2(t) \sim \frac{\chi^2(t)}{t}.$$

As time  $t$  goes to infinity,

$$\sqrt{t/2} (T^2(t) - 1) \xrightarrow{d} N(0, 1).$$

For online monitoring, we need to set control limit for the  $T^2(t)$  statistic. Different from traditional SPC methods where samples are independent, the proposed one is a sequential test with auto-correlated test statistics. Therefore, setting a common type I error  $\alpha$  for all time steps will not guarantee that the overall type I error is  $\alpha$ . To achieve a desired global  $\alpha$  error or IC average run length  $ARL_0$ , we propose to use an upper control limit  $UCL_{T^2}(t) = \frac{\chi_{c_\alpha}^2(t)}{t}$ , where  $\chi_{c_\alpha}^2(t)$  is the  $\alpha$ -upper quantile of  $\chi^2(t)$ , and  $c_\alpha$  is a scaling coefficient that can be obtained through Monte Carlo simulations.

As mentioned earlier, the Hotelling  $T^2$  control chart has equal detection power regardless of the mean shift directions. The detection power only depends on the shift magnitude. In AM, the average temperature of a scanning path often has a shift when anomaly occurs; that is, the intercept of the DNN function  $g(\cdot)$  changes. Suppose only the intercept of  $g(\cdot)$  can change, e.g.,  $\zeta_{1:t}$  may have a mean shift in the direction of  $\mathbf{1}_t$ , which is equivalent to the case that only  $\mu_\xi$  can change while  $g(\cdot)$  and other parameters remain unchanged. In **Corollary 1**(1), we show that the shift magnitude (Mahalanobis distance) converges to a constant value as  $t$  increases. Due to the dimensionality of the control chart dynamically increasing, the detection power decreases continuously. This phenomenon is not surprising, because all observations in a profile share the same parameter  $\xi$ , and thus increasing the length of the profile will not increase the information for  $\mu_\xi$ . The conclusion can also be made from **Corollary 1**(2). As  $t$  increases, the asymptotic distribution of  $T^2(t)$  is identical to the one without any changes, i.e., the one shown in **Theorem 1**. The proof of **Corollary 1** is provided in Appendix B.

**Corollary 1.** For any OC profile  $i$  with only  $\mu_\xi$  changed, i.e.,  $\mu_\xi \neq 0$ , as time  $t$  goes to infinity,

(1) The shift magnitude

$$(\mu_\xi \mathbf{1}_t)' \left( \Sigma_{1:t}^{(IC)} \right)^{-1} (\mu_\xi \mathbf{1}_t) \rightarrow \frac{\mu_\xi^2}{\sigma_\xi^2},$$

(2)

$$\sqrt{t/2} (T^2(t) - 1) \xrightarrow{d} N(0, 1).$$

Note that in process monitoring, we can equivalently assume that  $g(\cdot)$  is fixed, and any change of the profile is attributed to either model changes or parameter changes of the remaining two terms  $\xi_i$  and  $\gamma_i$ . For model changes,  $T^2(t)$  may work well. However, if only the distributional parameters of  $\xi_i$  and  $\gamma_i$  change,  $T^2(t)$  may perform poorly. **Corollary 1** is just a representative example that  $T^2(t)$  is ineffective in detecting the change of  $\mu_\xi$ . Therefore, in the following subsection, we propose a GLR-based control chart for monitoring the other parameters.

### 3.2. GLR-based control chart

Based on the discussion in previous subsection, the testing problem is formulated to

$$H_0 : \mu_\xi = 0, \sigma_\gamma^2 = \sigma_\gamma^{2(IC)}, \omega = \omega^{(IC)},$$

$$H_1 : \mu_\xi \neq 0 \text{ or } \sigma_\gamma^2 \neq \sigma_\gamma^{2(IC)} \text{ or } \omega \neq \omega^{(IC)}.$$

Note that we are not able to test the change of  $\sigma_\xi^2$  with only one profile, and thus, we assume it is fixed. For any new profile  $i$ , at time  $t$ , the log likelihood of the observed data with IC parameters is defined as

$$L_1(t) = -\frac{1}{2} \left( \zeta'_{i,1:t} \left( \Sigma_{1:t}^{(IC)} \right)^{-1} \zeta_{i,1:t} + \ln |\Sigma_{1:t}^{(IC)}| \right). \quad (14)$$

The log likelihood of the observed data evaluated at the MLE is expressed as:

$$L_2(t) = \max_{\mu_\xi, \sigma_\gamma^2, \omega} -\frac{1}{2} \left( (\zeta_{i,1:t} - \mu_\xi \mathbf{1}_t)' \Sigma_{i,1:t}^{-1} (\zeta_{i,1:t} - \mu_\xi \mathbf{1}_t) + \ln |\Sigma_{i,1:t}| \right), \quad (15)$$

where  $\Sigma_{i,1:t} = \sigma_\xi^{2(IC)} \mathbf{1}_t \mathbf{1}_t' + \sigma_\gamma^2 \mathbf{R}_{i,1:t}(\omega)$ . The GLR statistics is then calculated as

$$GLR(t) = L_2(t) - L_1(t).$$

In (15), the optimization problem has no analytical solution and needs to be solved numerically. However, the iterative algorithms may make it unrealistic for applications requiring a fast calculation. In such a case, we assume that the parameter  $\omega$  is fixed, with which the computational cost can be significantly reduced. **Theorem 2** gives the optimal solution of  $\hat{\mu}_\xi(t)$  and  $\hat{\sigma}_\gamma^2(t)$ , based on which the GLR statistics can be efficiently calculated. The proof is provided in Appendix C. Note that in most cases,  $\omega$  can still be effectively monitored with the GLR test, since its change the variance of observations, and thus cause the change of the optimized  $\sigma_\gamma^2$ .

**Theorem 2.** The optimal  $\mu_\xi, \sigma_\gamma^2$  of optimization problem (14) given  $\omega$  is

$$\hat{\mu}_\xi(t) = \frac{\mathbf{1}_t' \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \zeta_{i,1:t}}{\mathbf{1}_t' \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t},$$

$$\hat{\sigma}_\gamma^2(t) = \frac{1}{t-1} \left( \zeta'_{i,1:t} \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \zeta_{i,1:t} - \frac{\left( \mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \zeta_{i,1:t} \right)^2}{\mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t} \right) - \frac{\left( \hat{\sigma}_\gamma^2(t) \right)^2}{(t-1) \left( \hat{\sigma}_\gamma^2(t) + \sigma_\xi^2 \right) \mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t}.$$

and the GLR statistics can be calculated by

**Corollary 2.** For any OC profile  $i$  in monitoring, as time  $t$  goes to infinity,

(1) If  $\mu_\xi \neq 0$  and  $\sigma_\gamma^2 = \sigma_\gamma^{2(IC)}$ ,

$$GLR(t) \xrightarrow{d} \frac{1}{2} \left( \chi^2(1) + \frac{(\mu_\xi + \xi_i)^2}{\sigma_\xi^2} \right).$$

(2) If  $\mu_\xi = 0$  and  $\sigma_\gamma^2 = (1 + \delta_\gamma) \sigma_\gamma^{2(IC)}$ , ( $\delta_\gamma \neq 0$ ,  $\delta_\gamma > -1$ ),

$$GLR(t) = \frac{1}{2} (t-1) (\delta_\gamma - \ln(1 + \delta_\gamma)) + o(t).$$

$$GLR(t) = \frac{1}{2} (t-1) \left( \frac{\hat{\sigma}_\gamma^2(t)}{\sigma_\gamma^{2(IC)}} - 1 - \ln \frac{\hat{\sigma}_\gamma^2(t)}{\sigma_\gamma^{2(IC)}} \right) + \frac{1}{2} \left( \frac{\hat{\sigma}_\gamma^2(t) + \sigma_\xi^2 \mathbf{1}'_t \mathbf{R}_{1:t}^{(IC)-1} \mathbf{1}_t}{\sigma_\gamma^{2(IC)} + \sigma_\xi^2 \mathbf{1}'_t \mathbf{R}_{1:t}^{(IC)-1} \mathbf{1}_t} - 1 - \ln \frac{\hat{\sigma}_\gamma^2(t) + \sigma_\xi^2 \mathbf{1}'_t \mathbf{R}_{1:t}^{(IC)-1} \mathbf{1}_t}{\sigma_\gamma^{2(IC)} + \sigma_\xi^2 \mathbf{1}'_t \mathbf{R}_{1:t}^{(IC)-1} \mathbf{1}_t} \right) + \frac{1}{2} \frac{\hat{\mu}_\xi^2(t) \left( \mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t \right)}{\left( \sigma_\gamma^{2(IC)} + \sigma_\xi^2 \mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t \right)} + \frac{1}{2} \frac{\sigma_\xi^2 \left( \hat{\sigma}_\gamma^2(t) - \sigma_\gamma^{2(IC)} \right)^2 \mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t}{\sigma_\gamma^{2(IC)} \left( \hat{\sigma}_\gamma^2(t) + \sigma_\xi^2 \mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t \right) \left( \sigma_\gamma^{2(IC)} + \sigma_\xi^2 \mathbf{1}'_t \left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t \right)}. \quad (17)$$

The asymptotic property of  $\hat{\mu}_\xi(t)$ ,  $\hat{\sigma}_\gamma^2(t)$  and the GLR statistics are further investigated, as given in **Theorem 3** (see Appendix D for the proof).

**Theorem 3.** For any IC profile  $i$  in monitoring, as time  $t$  goes to infinity,

$$\hat{\mu}_\xi(t) \rightarrow \xi_i, \hat{\sigma}_\gamma^2(t) \rightarrow \sigma_\gamma^{2(IC)},$$

$$GLR(t) \xrightarrow{d} \frac{1}{2} \left( \chi^2(1) + \frac{\xi_i^2}{\sigma_\xi^2} \right) \sim \frac{1}{2} \chi^2(2), \quad t \rightarrow \infty.$$

**Theorem 3** demonstrates that as time  $t$  goes to infinity, the GLR statistics converges in distribution to a Chi-square distribution plus a positive value. Note that  $\xi_i$  is unknown in the monitoring process, and  $\xi_i \sim N(0, \sigma_\xi^2)$  for IC profiles. This positive value follows  $\chi^2(1)$  considering the randomness of  $\xi_i$  and then  $GLR(t)$  can be regarded as following  $\frac{1}{2} \chi^2(2)$ . Although the exact distribution of  $GLR(t)$  at finite time  $t$  is hard to get, it is found that  $GLR(t)$  also approximately follows  $\frac{1}{2} \chi^2(2)$  in simulation studies when  $t$  is small. Therefore,  $\frac{1}{2} \chi^2(2)$  can be used as an upper control limit in the proposed GLR control chart. Similar to the  $T^2$  control chart, the upper control limit is set to  $UCL_{GLR}(t) = \frac{1}{2} \chi_{c_\alpha, \alpha}^2(2)$  to achieve a desired global  $\alpha$  error or IC average run length  $ARL_0$ . The scaling coefficient  $c_\alpha$  can be obtained through Monte Carlo simulations.

**Corollary 2** gives the convergence property of the GLR statistics for OC profiles with shifted  $\mu_\xi$  and  $\sigma_\gamma^2$ . The proof is provided in Appendix E.

**Corollary 2** shows that the asymptotic property of the GLR statistic is influenced by both  $\mu_\xi$  and  $\sigma_\gamma^2$ . The shift of  $\mu_\xi$  will lead to an increase of GLR statistics and the change of  $\sigma_\gamma^2$  will make the statistic diverge to infinity. Clearly, the GLR test can effectively detect the changes as the OC distributions have obvious shifted. Indeed, taking  $\mu_\xi$  for example, if only  $\mu_\xi$  can change, we can easily show that the GLR test is exactly a directional  $\chi^2$  control chart monitoring the shift of  $g(\cdot)$  in the specific direction of  $\mathbf{1}_t$ , i.e.,

$$GLR(t) = \frac{1}{2} \frac{\left( \mathbf{1}'_t \left( \Sigma_{1:t}^{(IC)} \right)^{-1} \zeta_{i,1:t} \right)^2}{\mathbf{1}'_t \left( \Sigma_{1:t}^{(IC)} \right)^{-1} \mathbf{1}_t} = \frac{1}{2} \frac{\left( (\zeta_{1:t}^*)' (\mathbf{1}_t^*) \right)^2}{\text{var}[(\zeta_{1:t}^*)' (\mathbf{1}_t^*)]} \sim \frac{1}{2} \chi^2(1) \text{ under IC,}$$

where

$$\zeta_{1:t}^* = \left( \Sigma_{1:t}^{(IC)} \right)^{-1/2} \zeta_{1:t} \text{ and } \mathbf{1}_t^* = \left( \Sigma_{1:t}^{(IC)} \right)^{-1/2} \mathbf{1}_t.$$

Therefore, GLR is expected to be more powerful than the proposed  $T^2$  control chart in detecting the parameter change for  $\xi_i$  and  $\gamma_i$ .

### 3.3. Dynamic updating strategy for computational cost reduction

For both control charts, storing and computing  $\left( \mathbf{R}_{1:t}^{(IC)} \right)^{-1}$  requires a substantial amount of storage space and computational time when  $t$  is large, which may hinder its practical applications. In this section, we propose an efficient dynamic updating strategy that can significantly reduce the



computation as well as storage requirement when computing  $T^2(t)$  and  $GLR(t)$ .

The basic idea of the dynamic updating strategy is to take advantage of the sparsity of  $\mathbf{R}_{1:t}^{(IC)-1}$ , and update  $(\zeta_{i,1:t})'(\mathbf{R}_{1:t}^{(IC)})^{-1}(\zeta_{i,1:t})$ ,  $\mathbf{1}_t'(\mathbf{R}_{1:t}^{(IC)})^{-1}(\zeta_{i,1:t})$  and  $\mathbf{1}_t'\mathbf{R}_{1:t}^{(IC)-1}\mathbf{1}_t$  recursively. Denote the three terms as  $M_1(t)$ ,  $M_2(t)$  and  $M_3(t)$  respectively. It can be shown that

$$\begin{aligned} M_1(t) &= (\zeta_{i,1:t})'(\mathbf{R}_{1:t}^{(IC)})^{-1}(\zeta_{i,1:t}) \\ &= a_{\zeta_{i1}}^2 + b_{\zeta_{i1}}r_{i2} + \sum_{j=2}^{t-1} (c_{\zeta_{ij}}^2 + b_{\zeta_{ij}}\zeta_{i,j-1} + b_{\zeta_{ij}}\zeta_{i,j+1}) + a_{\zeta_{it}}^2 + b_{\zeta_{it}}r_{it} \\ &= M_1(t-1) + (c-a)\zeta_{i,t-1}^2 + a_{\zeta_{it}}^2 + b_{\zeta_{it}}r_{it}, t \geq 3, \end{aligned}$$

and  $M_1(1) = \zeta_{i1}^2$ ,  $M_1(2) = a_{\zeta_{i1}}^2 + 2b_{\zeta_{i1}}\zeta_{i2} + a_{\zeta_{i2}}^2$ . Similarly,

$$\begin{aligned} M_2(t) &= \mathbf{1}_t'(\mathbf{R}_{1:t}^{(IC)})^{-1}(\zeta_{i,1:t}) \\ &= M_2(t-1) + (c+b-a)\zeta_{i,t-1} + (a+b)\zeta_{it}, \end{aligned}$$

$$M_3(t) = \mathbf{1}_t'\mathbf{R}_{1:t}^{(IC)-1}\mathbf{1}_t = M_3(t-1) + (c+2b), t \geq 3,$$

and

$$\begin{aligned} M_2(1) &= r_{i1}, M_2(2) = (a+b)(\zeta_{i1} + \zeta_{i2}), M_3(1) = 1, M_3(2) \\ &= 2(a+b). \end{aligned}$$

Clearly, in the monitoring process, only few values, such as  $\zeta_{it}$ ,  $\zeta_{i,t-1}$ ,  $M_1(t)$ ,  $M_2(t)$ ,  $M_3(t)$ , have to be stored for updating, which greatly reduces the storage requirement. More importantly, the computation at each time step is very simple, with only several addition, subtraction and multiplication operations. It is worth mentioning that such a strategy can also be used in the Phase I optimization process to reduce the memory cost.

#### 4. Simulation study

In this section, we conduct a simulation study to demonstrate the performance of the proposed modeling and monitoring approach. All experiments are done using Python on an i7-6800K 3.40GHz Intel processor with 16GB RAM. The number of IC profiles is set to  $m_{IC} = 100$  for Phase I optimization and each profile has a length of  $1 \times 10^5$ . Suppose that there are 10 explanatory variables. The design points  $\mathbf{x}_{ij}$  of the profiles are uniformly distributed in the range of  $(-1, 1)$ . The IC parameters  $\sigma_{\zeta}^2$ ,  $\sigma_{\gamma}^2$  and  $\omega$  are set to 1, 0.25 and 0.5 respectively.

##### 4.1. Phase I optimization

In this subsection, the effectiveness of the proposed DMEM model and hard EM optimization algorithm is demonstrated. To show the advantage of the DMEM model, we compare it with some simplified models (SMs) without considering the random effects  $\zeta_i$  or the within-profile correlations. Specifically, define model SM1 and SM2 as

**Table 1.** The structure of the MLP network.

	Layer	Nodes	Number of parameters
MLP model	Dense layer 1	32	
	Dense layer 2	128	4224
	Dense layer 3	1024	132,096
	Dense layer 4	128	1312,00
	Dense layer 5	1	129

**Table 2.** The optimization results in Phase I optimization.

	IC	$\sigma_{\zeta}^2$	$\sigma_{\gamma}^2$	$\omega$	MSE
Profile 1	DMEM	<b>1.0017</b>	<b>0.2567</b>	<b>0.5119</b>	<b>0.0076</b>
	DMEM-TS	1.0130	0.2624	0.5227	0.0144
	SM1	/	0.5282	0.6197	0.1019
	SM2	0.9806	0.2588	/	0.0101
Profile 2	DMEM	<b>1.0010</b>	<b>0.2557</b>	<b>0.5101</b>	<b>0.0067</b>
	DMEM-TS	1.0124	0.2664	0.5297	0.0187
	SM1	/	0.5287	0.6127	0.0949
	SM2	0.9783	0.2572	/	0.0085
Profile 3	DMEM	<b>1.0040</b>	<b>0.2703</b>	<b>0.5358</b>	<b>0.0219</b>
	DMEM-TS	1.0092	0.2886	0.5647	0.0415
	SM1	/	0.5391	0.6430	0.1224
	SM2	0.9788	0.2706	/	0.0224

$y_{ij} = g(\mathbf{x}_{ij}) + \gamma_{ij}$  and  $y_{ij} = g(\mathbf{x}_{ij}) + \zeta_i + \epsilon_{ij}$  respectively, where  $\epsilon_{ij}$  is i.i.d. noise. In addition, the proposed hard EM algorithm is compared with a two-stage optimization algorithm (DMEM-TS), which first estimates the parameters of the DNN model, and then uses the residuals to estimate the parameters of the mixed-effects model. This two-stage algorithm is often used to roughly estimate the mixed-effects models or Bayesian hierarchical models (Wen *et al.*, 2018) due to its simplicity and low computational cost. It is actually used to initialize the parameters in our algorithm. The model fitting performance is evaluated by the accuracy of optimized model parameters and the Mean Squared Error (MSE) of  $g^{(IC)}(\mathbf{x}) - \hat{g}(\mathbf{x})$ , defined as

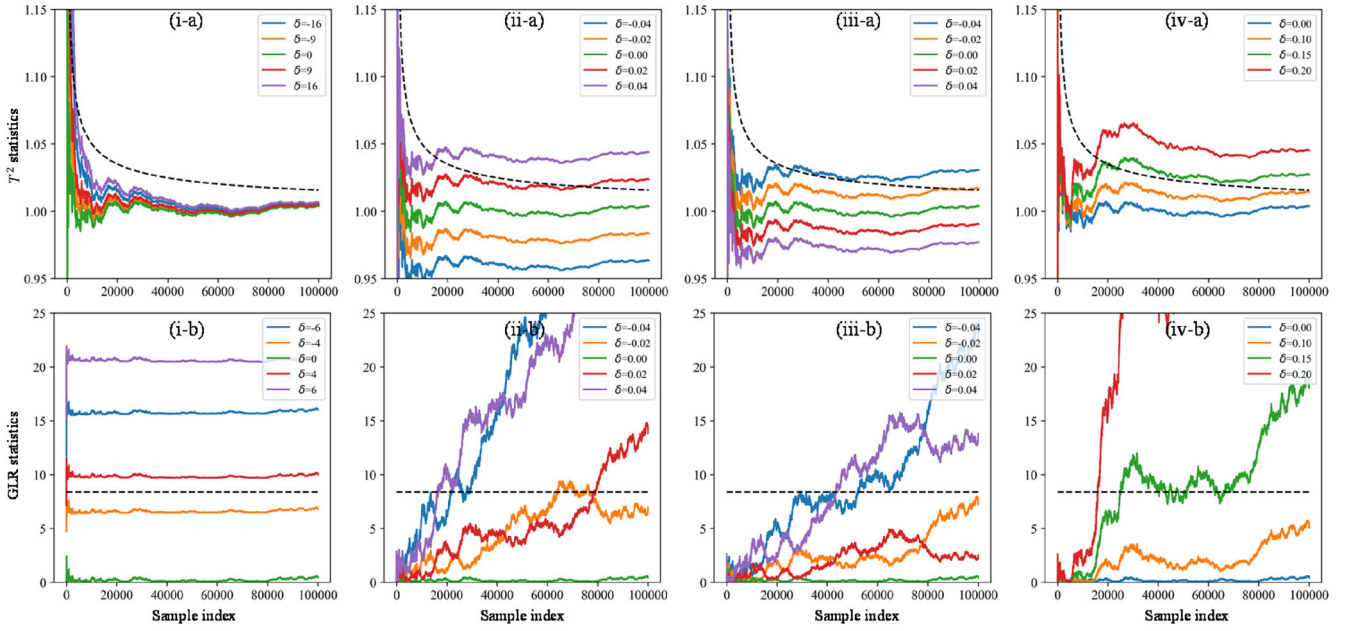
$$\text{MSE} = \frac{1}{\sum_{i=1}^{m_{IC}} n_i} \left( \sum_{i=1}^{m_{IC}} \sum_{j=1}^{n_i} (g^{(IC)}(\mathbf{x}_{ij}) - \hat{g}(\mathbf{x}_{ij}))^2 \right).$$

For the IC population profile function  $g^{(IC)}(\mathbf{x})$ , we consider the following three cases:

1.  $g^{(IC)}(\mathbf{x}) = \beta_1'\mathbf{x}$ .
2.  $g^{(IC)}(\mathbf{x}_{ij}) = \beta_1'\mathbf{x}_{ij} + \sin(\beta_2'\mathbf{x}_{ij})$ .
3.  $g^{(IC)}(\mathbf{x}_{ij}) = \beta_1'\mathbf{x}_{ij} + \sin(\beta_2'\mathbf{x}_{ij}) + \mathbf{x}_{ij}'\mathbf{B}_3\mathbf{x}_{ij}$ .

The IC parameters  $\beta_1, \beta_2, \mathbf{B}_3$  are randomly generated in the range of  $(-1, 1)$ . The sparsity of  $\mathbf{B}_3$  is set to 0.8. Obviously, from profile (1) to (3), the nonlinearity of the IC profile is increasing.

For DMEM, we choose a five-layer Multilayer Perceptron (MLP) network as the DNN network. The structure of this network is presented in Table 1. The optimization results of these models are provided in Table 2. The estimated parameters closest to the true value and the lowest MSE value are marked in bold. Clearly, the proposed hard EM algorithm can effectively estimate the model parameters with much better performance than the two-stage algorithm and the other benchmarks. As the profile function  $g^{(IC)}(\mathbf{x})$  gets more



**Figure 2.** Online monitoring of some IC and OC profiles via (a)  $T^2$  control chart and (b) GLR control chart. The change patterns in cases (i)-(iv) are the change of  $\mu_{\xi}$ ,  $\sigma_{\gamma}^2$ ,  $\omega$  and  $g(\mathbf{x})$  respectively.

complex, the accuracy of model optimization slightly decreases, especially in the optimization of  $\sigma_{\gamma}^2$  and  $\omega$ . The possible reason is that the specified DNN structure is not optimal for the other two cases. In comparison, the proposed DMEM outperforms other models in all cases with more accurate parameter estimation and lower MSE, demonstrating its advantage in modeling these data.

#### 4.2. Phase II monitoring

In this subsection, we test the detection performance regarding different profile anomalies, including the change of  $g(\cdot)$  and the change of the model parameters  $\mu_{\xi}$ ,  $\sigma_{\gamma}^2$ ,  $\omega$ . For statistical process control, the performance is typically evaluated using  $\alpha$ -error and  $\beta$ -error, or equivalently Average Run Length (ARL) for both IC and OC processes. However, in our case, both the ARL among profiles (each profile as one instance) and detection delay within each profile are critical for incremental profile monitoring. Therefore, in this article we specify  $ARL_0$  in terms of profiles in the IC process to a desirable level, i.e., 100, which corresponds to a type I error of 0.01, and use both  $ARL_1$  in terms of profiles and detection delay in terms of measurement points in the OC process as two evaluation metrics. In addition, the computational time is recorded to test whether the proposed methods are efficient enough for online applications.

In all cases we follow the conventional assumption that the true IC parameters are known in Phase II analysis and set  $g^{(IC)}(\mathbf{x}) = 0$  without loss of generality. For completeness, we also studied the performance using the estimated model parameters, which is provided in Appendix F due to space limitations. All results reported in this subsection are obtained by 1000 replicated simulations. To attain the desired  $ARL_0$  value of 100, the scaling coefficient  $c_{\alpha}$  for  $UCL_{T^2}(t)$  and  $UCL_{GLR}(t)$  are set to 0.00023. Note that for

both control charts, we find that the scaling coefficient is approximately the same in the simulation. For the change of model parameters  $\mu_{\xi}$ ,  $\sigma_{\gamma}^2$ ,  $\omega$ , we consider the following three cases: (i)  $\mu_{\xi}^{(OC)} = \delta\sigma_{\xi}$ , (ii)  $\sigma_{\gamma}^{2(OC)} = (1 + \delta)\sigma_{\gamma}^{2(IC)}$  and (iii)  $\omega^{(OC)} = (1 + \delta)\omega^{(IC)}$ . For the change of  $g(\mathbf{x})$ , we consider the following three cases: (iv)  $g^{(OC)}(\mathbf{x}) = (\delta\beta_1)' \mathbf{x}$ , (v)  $g^{(OC)}(\mathbf{x}) = \delta \sin(\beta_2' \mathbf{x})$  and (vi)  $g^{(OC)}(\mathbf{x}) = \mathbf{x}'(\delta\mathbf{B}_3)\mathbf{x}$ . Similar to the simulation of phase I optimization, the OC parameters  $\beta_1, \beta_2, \mathbf{B}_3$  are randomly generated in the range of  $(-1, 1)$  and the sparsity of  $\mathbf{B}_3$  is set to 0.8. The parameter  $\delta$  controls the shift magnitude in all cases.

Since there are no online control charts designed in the literature for incremental profile monitoring with a large number of measurement points in each profile, it is challenging to compare the proposed control charts with alternative methods. To demonstrate the superiority and effectiveness of our method, the widely used EWMA (exponentially weighted moving average) control chart and CUSUM (cumulative sum) control chart are chosen as the benchmarks, which also monitor the residuals  $\zeta_{i,1:t}$  in an online manner. The EWMA statistics are calculated by

$$EWMA(t) = \lambda_1 \frac{\zeta_{it}^2}{\sigma_{\xi}^{2(IC)} + \sigma_{\gamma}^{2(IC)}} + (1 - \lambda_1)EWMA(t-1),$$

where the parameter  $\lambda_1$  is set to 0.2. The two-side CUSUM statistics are calculated by

$$CUSUM_{upper}(t) = \max(0, CUSUM_{upper}(t-1) + \zeta_{it} - \lambda_2),$$

$$CUSUM_{lower}(t) = \min(0, CUSUM_{lower}(t-1) + \zeta_{it} + \lambda_2).$$

The parameter  $\lambda_2$  and the control limits of these two control charts are set to achieve the desired  $ARL_0 = 100$ .

**Table 3.** The ARL and detection delay for all methods.

Case	$\delta$	ARL				Detection delay			
		$T^2$	GLR	EWMA	CUSUM	$T^2$	GLR	EWMA	CUSUM
IC	0	100.00	100.00	100.00	100.00	99,054.1	<b>99,011.7</b>	99,221.1	99,084.9
(i)	−4	1.98	1.31	<b>1.10</b>	<b>1.10</b>	49,550.7	24,671.7	10,692.0	<b>91,98.8</b>
	−2	15.38	7.81	3.88	<b>3.82</b>	93,609.0	87,522.4	76,110.7	<b>73,987.1</b>
	−1	76.92	31.25	16.67	<b>16.39</b>	98,793.7	96,910.4	94,524.5	<b>94,004.0</b>
	1	90.91	37.04	<b>21.28</b>	22.73	98,992.8	97,347.7	95,745.6	<b>95,620.6</b>
	2	18.18	8.00	4.29	<b>4.17</b>	94,589.9	88,242.5	78,436.1	<b>76,172.4</b>
(ii)	4	1.94	1.35	1.12	<b>1.11</b>	48,528.8	26,677.6	11,800.9	<b>10,173.3</b>
	−0.02	333.33	<b>1.31</b>	100.00	100.00	99,700.0	<b>67,323.7</b>	99,275.5	99,085.0
	−0.01	250.00	<b>9.71</b>	100.00	100.00	99,600.3	<b>95,686.8</b>	99,221.1	99,085.0
	−0.005	250.00	<b>55.56</b>	100.00	100.00	99,600.3	<b>98,657.0</b>	99,221.1	99,084.9
	0.005	<b>25.00</b>	41.67	100.00	90.91	<b>97,389.9</b>	98,326.0	99,221.1	99,062.4
(iii)	0.01	<b>4.81</b>	9.71	100.00	90.91	<b>90,229.3</b>	95,475.4	99,221.1	99,062.4
	0.02	<b>1.14</b>	1.26	100.00	90.91	<b>51,719.9</b>	63,251.6	99,200.4	99,062.4
	−0.03	<b>1.14</b>	1.27	100.00	100.00	<b>51,642.8</b>	63,438.7	99,257.5	99,085.0
	−0.02	<b>2.16</b>	3.50	100.00	100.00	<b>79,916.8</b>	89,453.6	99,221.1	99,085.0
	−0.01	<b>13.51</b>	27.78	100.00	100.00	<b>95,815.1</b>	97,821.8	99,221.1	99,085.0
(iv)	0.01	250.00	<b>35.71</b>	100.00	90.91	99,600.3	<b>98,323.4</b>	99,221.1	99,062.4
	0.02	250.00	<b>3.89</b>	100.00	90.91	99,600.4	<b>90,398.6</b>	99,221.1	99,062.3
	0.03	333.33	<b>1.32</b>	100.00	90.91	99,700.0	<b>67,329.6</b>	99,199.4	99,062.3
	−0.15	<b>1.04</b>	1.06	66.67	52.63	<b>28,255.7</b>	34,293.7	98,659.3	98,251.9
	−0.10	<b>1.86</b>	2.39	76.92	62.50	<b>73,355.6</b>	81,891.7	99,074.1	98,497.8
(v)	−0.05	<b>35.71</b>	50.00	100.00	76.92	<b>97,975.9</b>	98,473.0	99,217.8	98,847.7
	0.05	<b>32.26</b>	55.56	100.00	83.33	<b>97,996.3</b>	98,552.7	99,221.8	98,870.1
	0.10	<b>1.89</b>	2.42	90.91	62.50	<b>73,193.5</b>	81,724.5	99,095.6	98,578.4
	0.15	<b>1.04</b>	1.06	66.67	55.56	<b>28,227.8</b>	34,778.0	98,791.2	98,376.9
	−0.20	<b>1.05</b>	1.10	71.43	66.67	<b>41,705.5</b>	51,600.8	98,955.3	98,631.3
(vi)	−0.15	<b>2.18</b>	3.27	83.33	76.92	<b>80,626.9</b>	89,199.7	99,138.2	98,733.1
	−0.10	<b>20.83</b>	38.46	90.91	76.92	<b>97,159.9</b>	98,257.7	99,174.6	98,773.6
	0.10	<b>17.86</b>	33.33	100.00	76.92	<b>96,893.9</b>	97,952.5	99,221.8	98,802.5
	0.15	<b>2.18</b>	3.38	100.00	76.92	<b>81,011.9</b>	89,044.4	99,104.5	98,756.2
	0.20	<b>1.05</b>	1.10	90.91	62.50	<b>41,893.3</b>	51,700.6	99,057.0	98,606.5
(vii)	−0.15	<b>1.09</b>	1.16	62.50	47.62	<b>40,650.2</b>	49,565.0	98,736.0	98,350.1
	−0.10	<b>3.45</b>	5.46	90.91	71.43	<b>86,238.6</b>	9,1963.8	99,061.5	98,795.8
	−0.05	<b>40.00</b>	62.50	100.00	90.91	<b>98,163.3</b>	98,619.3	99,253.7	98,940.6
	0.05	<b>47.62</b>	66.67	100.00	83.33	<b>98,374.1</b>	98,648.2	99,185.8	98,835.8
	0.10	<b>3.38</b>	5.46	90.91	71.43	<b>86,651.8</b>	92,344.7	99,094.2	98,728.5
	0.15	<b>1.10</b>	1.15	76.92	55.56	<b>40,708.3</b>	50,332.8	9,8907.8	98,469.3

Figure 2 illustrates the online monitoring process of IC and OC profiles by the proposed  $T^2$  control chart and GLR control chart on randomly generated profiles. For the change of model parameters  $\mu_\xi$ ,  $\sigma_\gamma^2$  and  $\omega$ , the proposed GLR control chart outperforms the  $T^2$  control chart. According to Corollary 1, the  $T^2$  control chart is not effective to detect the shift of  $\mu_\xi$ . In this case, an abnormal  $\mu_\xi$  leads to some deviations for the  $T^2$  statistics in the early period, and then the  $T^2$  statistic becomes stable as  $t$  increases. Whereas for the GLR control chart (Figure 2 i-b), the GLR statistics are shifted upwards significantly. In the other cases, the  $T^2$  control chart can only detect one-side change of  $\sigma_\gamma^2$  and  $\omega$ , specifically the increase of  $\sigma_\gamma^2$  and the decrease of  $\omega$ . The convergence values of the  $T^2$  statistics deviate from the IC convergence value of one. However, the GLR control chart is capable of detecting all kinds of changes, whereas the GLR statistics would diverge to infinity as the sample size goes to infinity. For the change of  $g(\mathbf{x})$ , both control charts can detect the anomalies. The  $T^2$  statistics are shifted upwards and the GLR statistics diverge to infinity. As the shift magnitude  $\delta$  increases, the detection delay for both control charts decreases.

The detailed detection results for these cases are shown in Table 3. The best values among these methods are marked in bold. From the table we can find that:

1. The EWMA control chart and CUSUM control chart perform well only when  $\mu_\xi$  changes. Note that the within-profile correlations are ignored in these two control charts. They focus on the mean value of the monitored residuals, therefore, they are not effective in monitoring other changes, especially the change of  $\sigma_\gamma^2$  and  $\omega$ .
2. The proposed  $T^2$  control chart performs best in the change of  $g(\mathbf{x})$  and the one-side change of  $\sigma_\gamma^2$  and  $\omega$ , while it performs worst in the remaining cases. The poor detection performance is due to the asymptotic properties shown in Corollary 1 and the one-side test issue.
3. The proposed GLR control chart is the most robust method, which shows good performance in all cases. The running time of the  $T^2$  control chart and the GLR control chart are around 0.84s and 1.88s per profile ( $1 \times 10^5$  points) respectively, both computationally efficient for online monitoring.

## 5. Real case study

In this section, we apply the proposed DMEM- $T^2$  control chart and DMEM-GLR control chart to a SLM-based AM process. As shown in Figure 3, one spatially-integrated sensor, which is an indium gallium arsenide (InGaAs)

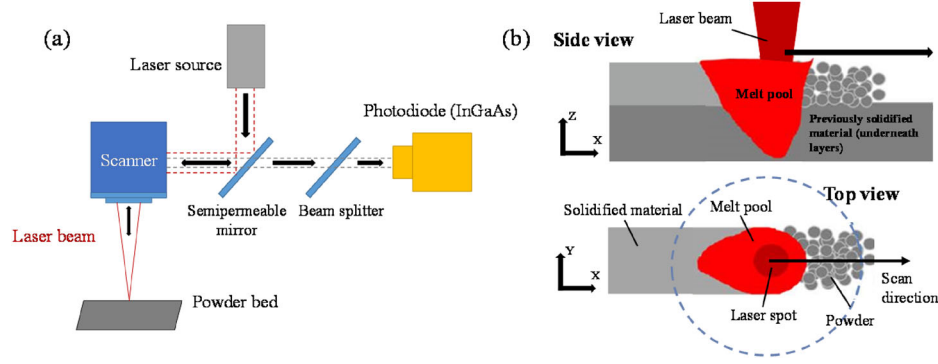


Figure 3. (a) The co-axial monitoring system. (b) Schematic side and top views of the melt pool along the laser scan direction.

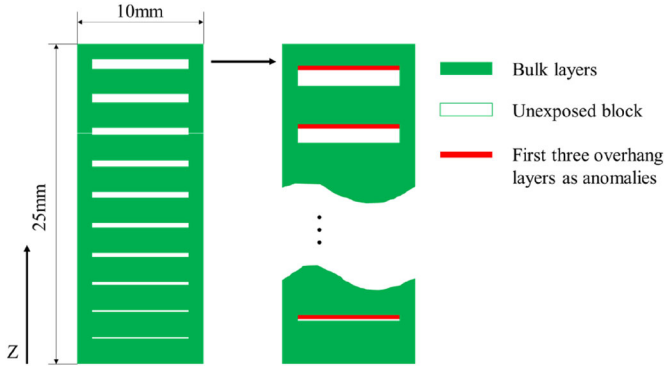


Figure 4. Schematic view of the additively manufactured specimens with unexposed blocks.

photodiode, is mounted co-axially to the laser path to measure the integral radiation within a field of view centered in the melt pool in the near/short infrared range.

In this study, we aim to detect the heat conduction anomalies as soon as possible during the manufacturing process. As shown in Figure 4, one parallelepiped specimen of size  $10 \times 10 \times 25$  mm is designed with some unexposed blocks to introduce heat conduction anomalies. In these blocks, no laser scan occurred for a number of consecutive layers. The first layer after an unexposed block has a large overhanging area with loose powder underneath. The heat exchange in this overhanging layer is altered by the fact that the loose powder has much less conductivity than the bulk material. Therefore, unexposed blocks tend to result in heat conduction anomalies with increasing severity as the number of unexposed layers increases. Figure 5 illustrates the photodiode signals of one IC layer and three OC layers which follow three, six and nine unexposed layers respectively. The photodiode signals turn higher with more unexposed layers. In addition, the heat conduction anomalies are possible to continue in a few of the following layers. In this case study, the first three overhang layers after each unexposed block are regarded as anomalies. Other bulk layers are supposed to be in control. A total of 10 unexposed blocks are designed with the number of unexposed layers increasing from 1 to 10 along the  $z$ -direction.

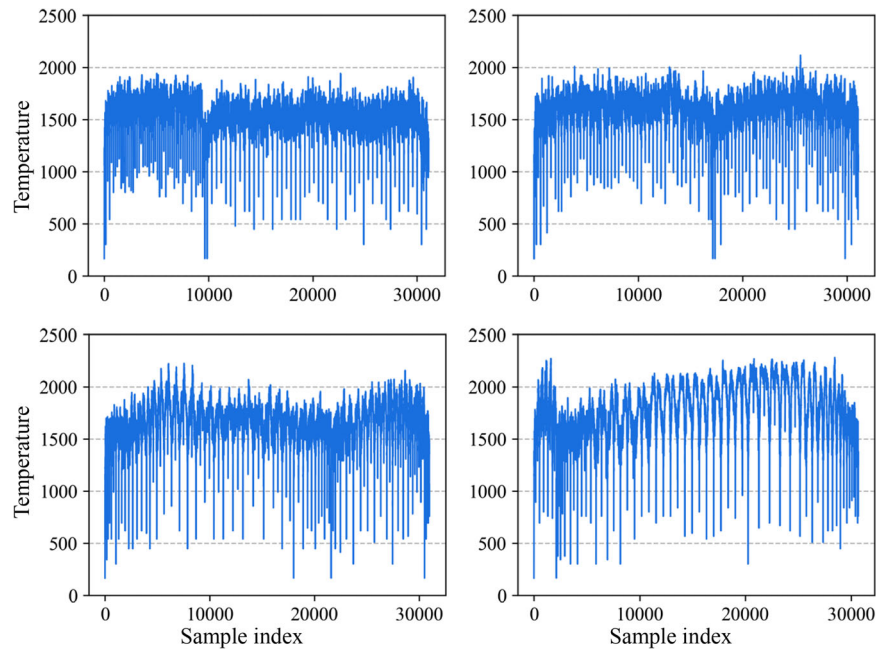
The experiment was repeated twice using a multi-laser L-PBF Trumpf system, so two specimens are manufactured. The data of 379 layers in each experiment are collected, including 294 bulk layers, 55 unexposed layers and 30

overhang layers. Each bulk layer and overhang layer contains about 30,000 measurement points. The process parameters are fixed with a scan speed of 1500 mm/s, a laser power of 4800 W, and a laser spot diameter of  $100 \mu\text{m}$ . The photodiode signal was initially acquired with a sampling rate of 100 kHz and then down-sampled in order to have one measurement point every  $30 \mu\text{m}$  along the laser scan path. The orientation of the laser scan and the laser scan path were changed every layer, so the coordinates of the measurements vary from layer to layer. The coordinates of the laser spot and the measured laser power are collected along with the photodiode signal. The dataset is part of an Open Data Science project between Trumpf GmbH and Politecnico di Milano and it is available at [www.ic.polimi.it/open-data-challenge](http://www.ic.polimi.it/open-data-challenge).

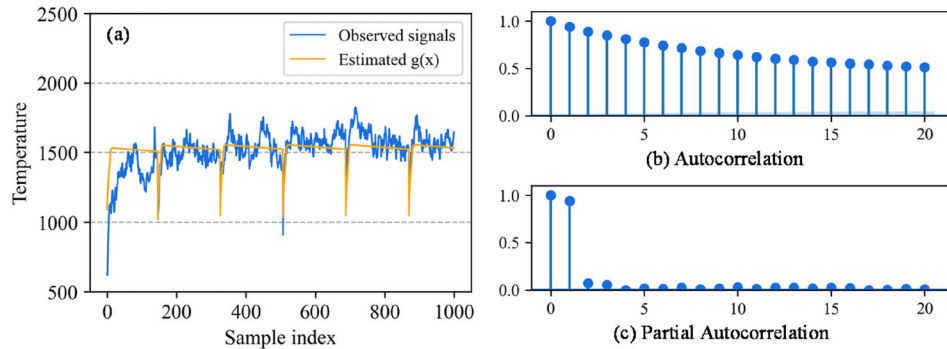
Based on the heat exchange process of the melt pool and the theory of thermodynamics, the photodiode signal can be regarded as a profile with some explanatory variables, including the position of the laser spot and the laser power. In order to better model the photodiode signal, five features are extracted from the observed data, namely, the distance between the laser spot and the scanning edge, the length of the previous scan trace, the length and index of the current scan trace and the distance between the current and previous laser spot. Finally, these five features, together with the measured laser power of the latest 10 measurement points, are used as the explanatory variables. All bulk layers are taken as the training data in Phase I optimization, and the overhang layers are used for performance evaluation and comparison of Phase II online monitoring.

In Phase I analysis, we find that the optimized DNN model is not accurate in fitting the first measurement point of each scan trace. Therefore, we treat it as an outlier and remove it to improve model fitting. Figure 6(a) shows the observed signals and fitted values in the first IC layer, where the observed signals fluctuates around the estimated  $\hat{g}(\mathbf{x})$ . Figure 6(b) and (c) shows the autocorrelation and partial autocorrelation of the residuals, from which we see the residuals indeed follow an autoregressive model AR(1). The normal Q-Q plots of the estimated  $\hat{\xi}_i$  of IC profiles in these two specimens are shown in Figure 7. The estimated  $\hat{\xi}$  approximately follows a Gaussian distribution. Therefore, the DMEM model fits the temperature data very well.

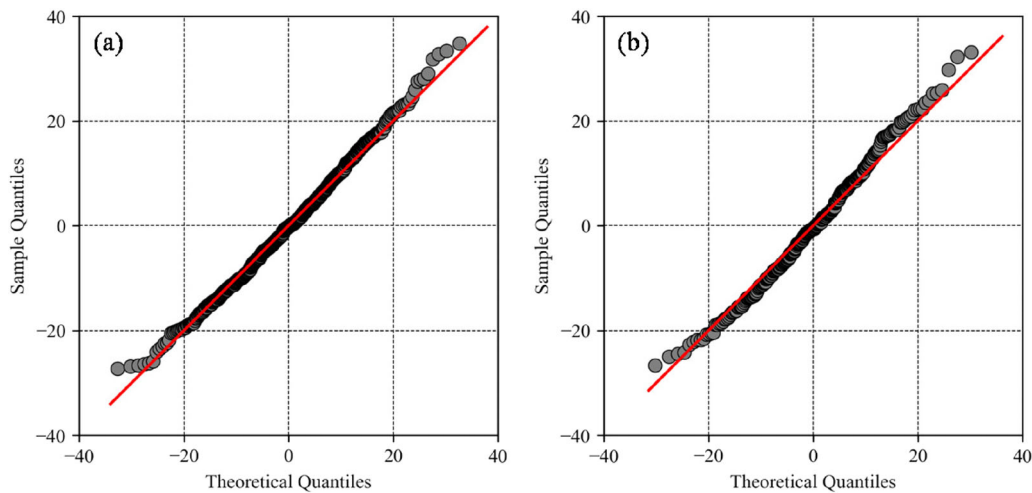




**Figure 5.** The photodiode signals of (a) one IC layer and three OC layers which follow (b) three, (c) six and (d) nine unexposed layers.



**Figure 6.** (a) The observed signals and the estimated  $\hat{g}(x)$  in the first IC layer. (b-c) The autocorrelation function and partial autocorrelation function of the residuals.



**Figure 7.** The normal Q-Q plots for the estimated  $\hat{\xi}_i$  of IC profiles in the (a) first and (b) second specimen.



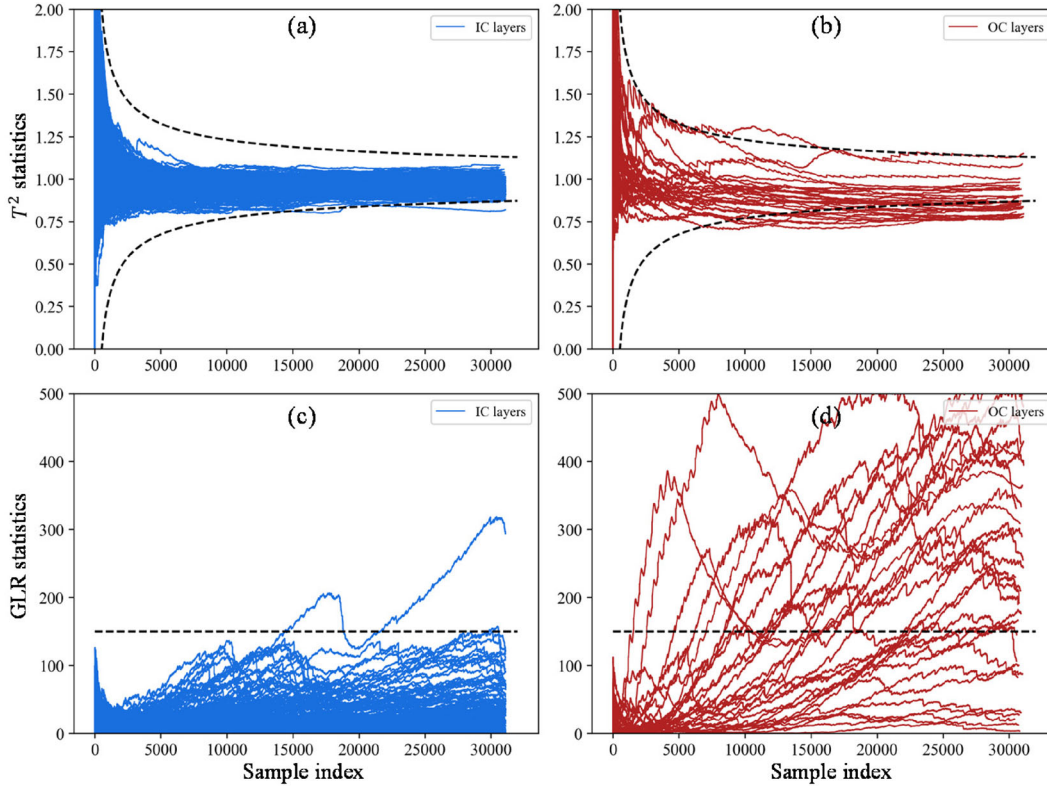


Figure 8. Phase II monitoring process of (a-b) DMEM- $T^2$  control chart and (c-d) DMEM-GLR control chart in the first manufacturing process.

Table 4. The detection results using specimen 1, specimen 2, and both specimens respectively as the data source.

		$\beta$ -error				Detection delay			
		$T^2$	GLR	EWMA	CUSUM	$T^2$	GLR	EWMA	CUSUM
Specimen 1	Layers	<b>0.267</b>	<b>0.267</b>	0.9	0.767	20,157.2	<b>18,344.5</b>	28,348.7	24,799.6
	Blocks	<b>0.1</b>	<b>0.1</b>	0.7	0.4	25,988.2	<b>21,394.2</b>	66,548.9	39,065.6
Specimen 2	Layers	0.333	<b>0.3</b>	0.933	0.633	19,367.7	<b>18,186.9</b>	30,767.3	23,980.4
	Blocks	<b>0.1</b>	<b>0.1</b>	0.8	0.2	26,976.5	<b>22,489.5</b>	83,045.4	28,870.0
Specimen 1&2	Layers	0.300	<b>0.283</b>	0.933	0.700	19,558.3	<b>17,837.7</b>	29,165.8	24,390.0
	Blocks	<b>0.1</b>	<b>0.1</b>	0.8	0.3	26,272.4	<b>21,117.6</b>	75,171.3	33,967.8

Based on the simulation study, we know that the one-side  $T^2$  control chart performs poorly in some cases with shifted model parameters for the random effect and residuals. To make it more robust, we also set a lower control limit  $LCL_{T^2}(t)$  for the  $T^2$  control chart, where

$$UCL_{T^2}(t) = \frac{\chi^2_{c_\alpha, \alpha/2}(t)}{t} \quad \text{and} \quad LCL_{T^2}(t) = \frac{\chi^2_{1-c_\alpha, \alpha/2}(t)}{t}.$$

The coefficient  $c_\alpha$  is determined to achieve a fixed  $\alpha$ -error at 0.01. Figure 8 illustrates the Phase II monitoring process via DMEM- $T^2$  control chart and DMEM-GLR control chart for the first specimen. The monitoring of the second specimen shows similar results. The  $\beta$ -error and detection delay are summarized in Table 4. Here we calculate the  $\beta$ -error in two ways, one for OC profiles and the other for OC blocks. The  $\beta$ -error of OC blocks is defined as the probability of failing to detect any anomaly in the three OC layers of one OC block. The best values among these methods are marked in bold. The proposed DMEM- $T^2$  control chart and DMEM-GLR control chart both perform well in detecting anomalous blocks. All the unexposed blocks except the first one can be detected

(9 out of 10 blocks). As there is only one unexposed layer in the first OC block, the anomaly is not obvious and thus both methods fail to detect it. The DMEM-GLR control chart is superior to all other methods in terms of the detection delay, indicating that it can more timely detect the anomalies. The  $\beta$ -error of OC layers is relatively lower than that of OC blocks. One reason is that the influence caused by abnormal heat exchange may not last for three layers, especially when the number of unexposed layers is small. The proposed methods only detect anomalies in one or two layers among the three OC layers in one OC block. As the number of unexposed layers increases, the detection by both proposed control charts becomes faster and more accurate. The computational time of DMEM- $T^2$  and DMEM-GLR method are around 2.11s and 2.51s per printing layer respectively, fast enough for online monitoring.

## 6. Conclusion and discussion

In this article, we proposed a deep mixed-effects modeling approach for real-time monitoring of the melt pool

temperature in AM processes. In the DMEM model, a DNN is proposed to capture the relationship between the temperature and process setting and other sensing data. A random-effect term is added to model the random variation of mean temperature among different printing layers, and a residual term is used to model the within-profile auto-correlation and measurement error. An efficient hard EM algorithm is developed to train the model in Phase I stage, which iteratively optimizes the DNN parameters and the ones for random-effect term and residual term. In Phase II monitoring, a  $T^2$  control chart and GLR-based control chart are developed to timely detect various process anomalies. The asymptotic properties of the  $T^2$  and GLR statistics with and without anomalies are further established. An efficient dynamic updating strategy is proposed to reduce the computational cost. The effectiveness of the proposed method is demonstrated thorough simulation studies and a real case study of an AM process. The results show that the DMEM-GLR control chart is the most robust method in handling various process changes in both simulation study and real case study.

There are several issues that are worthy of further investigation. First, the random terms in DMEM are modeled by a autoregressive model AR(1), which may not hold in other more complicated cases. A Gaussian process is an alternative approach and is more general for different situations by setting different kernel functions. Second, although very effective in our study, the proposed optimization algorithm may not be accurate when limited historical IC profiles are available. Finally, the proposed method focuses on the photodiode signals in monitoring the AM process. Other type of signals, such as thermal images may be more effective in AM process monitoring.

## Funding

This work was partially supported by National Natural Science Foundation of China grant NSFC-71932006, NSFC-72171003, NSFC-51875003.

## Notes on contributors

**Ruiyu Xu** received a BE degree at the Department of Energy and Resources Engineering, Peking University, Beijing, China, in 2019. He is currently pursuing a PhD degree at the Department of Industrial Engineering and Management, Peking University, Beijing. His research interests include time series analysis, deep learning, quality and reliability engineering and advanced manufacturing.

**Song Huang** received a BS degree at the Department of Mechanics and Engineering Science, Peking University, Beijing, China, in 2021. He is currently pursuing a PhD degree at the Department of Industrial Engineering and Management, Peking University, Beijing. His research interests include quality and reliability engineering, statistical learning, time series analysis and change point detection.

**Zheren Song** received a BE degree at the Department of Engineering Mechanics, Peking University, Beijing, China, in 2019. He is now a PhD candidate at the Department of Industrial Engineering and Management, Peking University, Beijing. His research topics involve deep generative learning, computer vision, additive manufacturing and quality engineering.

**Yuanyuan Gao** received a BS degree in aircraft design and engineering from Beihang University, Beijing, China, in 2018. She is currently

pursuing a PhD degree in industrial engineering and management with Peking University, Beijing. Her research interests are focused on data mining, advanced data analytics, quality, and reliability engineering. She is a recipient of the National Scholarship Award from Peking University and won the Championship in the INFORMS QSR Data Challenge. She is a member of IEEE, INFORMS, IISE.

**Jianguo Wu** received a BS degree in mechanical engineering from Tsinghua University in 2009, an MS degree in mechanical engineering from Purdue University in 2011, an MS degree in statistics in 2014 and a PhD degree in industrial and systems engineering in 2015 from the University of Wisconsin-Madison. Currently, he is an assistant professor in the Department of Industrial Engineering and Management at Peking University, Beijing, China. He was an assistant professor at the Department of Industrial, Manufacturing and Systems Engineering at UTEP, TX, USA from 2015 to 2017. His research interests are mainly in quality control and reliability engineering of intelligent manufacturing and complex systems through engineering-informed machine learning and advanced data analytics. He is a recipient of the STARS Award from the University of Texas Systems, Overseas Distinguished Young Scholars from China, and several Best Paper Award/Finalists from INFORMS/IIE Annual Meetings. He is an associate editor of the *Journal of Intelligent Manufacturing*, and a member of IEEE, INFORMS, IISE, and SME.

## ORCID

Ruiyu Xu  <http://orcid.org/0000-0001-5192-6266>

Jianguo Wu  <http://orcid.org/0000-0002-2885-8725>

## References

- Alberts, D., Schwarze, D. and Witt, G. (2017) In situ melt pool monitoring and the correlation to part density of Inconel® 718 for quality assurance in selective laser melting, in *Proceedings of the 2017 International Solid Freeform Fabrication Symposium*, The University of Texas at Austin, Austin, TX, pp. 1481–1494.
- Bisht, M., Ray, N., Verbist, F. and Coeck, S. (2018) Correlation of selective laser melting-melt pool events with the tensile properties of Ti-6Al-4V ELI processed by laser powder bed fusion. *Additive Manufacturing*, **22**, 302–306.
- Chen, Y., Birch, J.B. and Woodall, W.H. (2015) A phase I cluster-based method for analyzing nonparametric profiles. *Quality and Reliability Engineering International*, **31**, 1675–1689.
- Chua, Z.Y., Ahn, I.H. and Moon, S.K. (2017) Process monitoring and inspection systems in metal additive manufacturing: Status and applications. *International Journal of Precision Engineering and Manufacturing-Green Technology*, **4**, 235–245.
- Chuang, S.-C., Hung, Y.-C., Tsai, W.-C. and Yang, S.-F. (2013) A framework for nonparametric profile monitoring. *Computers & Industrial Engineering*, **64**, 482–491.
- Colosimo, B., Grossi, E., Caltanissetta, F. and Grasso, M. (2020) Penelope: A novel prototype for in situ defect removal in LPBF. *The Journal of The Minerals, Metals & Materials Society*, **72**, 1332–1339.
- Colosimo, B.M., Huang, Q., Dasgupta, T. and Tsung, F. (2018) Opportunities and challenges of quality engineering for additive manufacturing. *Journal of Quality Technology*, **50**, 233–252.
- Demir, A.G. and Previtali, B. (2017) Investigation of remelting and pre-heating in SLM of 18Ni300 maraging steel as corrective and preventive measures for porosity reduction. *The International Journal of Advanced Manufacturing Technology*, **93**, 2697–2709.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 1–22.
- Everton, S.K., Hirsch, M., Stravroulakis, P., Leach, R.K. and Clare, A.T. (2016) Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing. *Materials & Design*, **95**, 431–445.
- Forien, J.-B., Calt, N.P., DePond, P.J., Guss, G.M., Roehling, T.T. and Matthews, M.J. (2020) Detecting keyhole pore defects and monitoring process signatures during laser powder bed fusion: A correlation

- between in situ pyrometry and ex situ X-ray radiography. *Additive Manufacturing*, **35**, 101336.
- Grasso, M. and Colosimo, B. (2019) A statistical learning method for image-based monitoring of the plume signature in laser powder bed fusion. *Robotics and Computer-Integrated Manufacturing*, **57**, 103–115.
- Grasso, M. and Colosimo, B.M. (2017) Process defects and in situ monitoring methods in metal powder bed fusion: A review. *Measurement Science and Technology*, **28**, 044005.
- Grasso, M.L.G., Remani, A., Dickins, A., Colosimo, B.M. and Leach, R.K. (2021) In-situ measurement and monitoring methods for metal powder bed fusion—An updated review. *Measurement Science and Technology*, **32**, 112001.
- Guo, S., Guo, W.G. and Bain, L. (2020) Hierarchical spatial-temporal modeling and monitoring of melt pool evolution in laser-based additive manufacturing. *IIEE Transactions*, **52**, 977–997.
- Heeling, T. and Wegener, K. (2018) The effect of multi-beam strategies on selective laser melting of stainless steel 316L. *Additive Manufacturing*, **22**, 334–342.
- Hung, Y.-C., Tsai, W.-C., Yang, S.-F., Chuang, S.-C. and Tseng, Y.-K. (2012) Nonparametric profile monitoring in multi-dimensional data spaces. *Journal of Process Control*, **22**, 397–403.
- Jalalhamadi, B., Liu, J., Rios, J., Slotwinski, J., Peitsch, C., Goldberg, A. and Montalbano, T. (2019) In-process defect monitoring and correction in additive manufacturing of aluminum alloys, in *Proceedings of the Vertical Flight Society's 75th Annual Forum & Technology Display*, Vertical Flight Society, Philadelphia, PA, pp. 2279–2292.
- Jayasinghe, S., Paoletti, P., Sutcliffe, C., Dardis, J., Jones, N. and Green, P.L. (2022) Automatic quality assessments of laser powder bed fusion builds from photodiode sensor measurements. *Progress in Additive Manufacturing*, **7**, 143–160.
- Jensen, W.A., Birch, J.B. and Woodall, W.H. (2008) Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology*, **40**, 167–183.
- Kim, K., Mahmoud, M.A. and Woodall, W.H. (2003) On the monitoring of linear profiles. *Journal of Quality Technology*, **35**, 317–328.
- Ko, H., Moon, S.K. and Hwang, J. (2015) Design for additive manufacturing in customized products. *International Journal of Precision Engineering and Manufacturing*, **16**, 2369–2375.
- Kruth, J.-P., Mercelis, P., Van Vaerenbergh, J. and Craeghs, T. (2007) Feedback control of selective laser melting, in *Proceedings of the Proceedings of the 3rd International Conference on Advanced Research in Virtual and Rapid Prototyping*, Taylor & Francis Ltd, London, pp. 521–527.
- Liu, J., Liu, C., Bai, Y., Rao, P., Williams, C.B. and Kong, Z. (2019) Layer-wise spatial modeling of porosity in additive manufacturing. *IIEE Transactions*, **51**, 109–123.
- Mahmoudi, M., Ezzat, A.A. and Elwany, A. (2019) Layerwise anomaly detection in laser powder-bed fusion metal additive manufacturing. *Journal of Manufacturing Science and Engineering*, **141**(3), 031002.
- Mani, M., Feng, S., Lane, B., Donmez, A., Moylan, S. and Feserman, R. (2015) Measurement science needs for real-time control of additive manufacturing powder bed fusion processes, in *Additive Manufacturing Handbook*, CRC Press, Boca Raton, FL, pp. 629–652.
- Montgomery, D.C. (2007) *Introduction to Statistical Quality Control*, John Wiley & Sons, Hoboken, NJ.
- Nadipalli, V.K., Andersen, S.A., Nielsen, J.S. and Pedersen, D.B. (2019) Considerations for interpreting in-situ photodiode sensor data in pulsed mode laser powder bed fusion, in *Proceedings of the Joint Special Interest Group meeting between EUSPEN and ASPE Advancing Precision in Additive Manufacturing*, The European Society for Precision Engineering and Nanotechnology, Ecole Centrale de Nantes, Nantes, France, pp. 66–69.
- Neal, R.M. and Hinton, G.E. (1998) A view of the EM algorithm that justifies incremental, sparse, and other variants, in *Learning in Graphical Models*, Springer, Berlin/Heidelberg, Germany, pp. 355–368.
- Okaro, I.A., Jayasinghe, S., Sutcliffe, C., Black, K., Paoletti, P. and Green, P.L. (2019) Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. *Additive Manufacturing*, **27**, 42–53.
- Qiu, P. and Zou, C. (2010) Control chart for monitoring nonparametric profiles with arbitrary design. *Statistica Sinica*, **20**(4), 1655–1682.
- Qiu, P., Zou, C. and Wang, Z. (2010) Nonparametric profile monitoring by mixed effects modeling. *Technometrics*, **52**, 265–277.
- Rao, P.K., Liu, J.P., Roberson, D., Kong, Z.J. and Williams, C. (2015) Online real-time quality monitoring in additive manufacturing processes using heterogeneous sensors. *Journal of Manufacturing Science and Engineering*, **137**(6), 061007.
- Renken, V., von Freyberg, A., Schünemann, K., Pastors, F. and Fischer, A. (2019) In-process closed-loop control for stabilising the melt pool temperature in selective laser melting. *Progress in Additive Manufacturing*, **4**, 411–421.
- Ruggieri, A., Stranieri, F., Stella, F. and Scutari, M. (2020) Hard and soft EM in Bayesian network learning from incomplete data. *Algorithms*, **13**(12), 329–345.
- Sachlas, A., Bersimis, S. and Psarakis, S. (2019) Risk-adjusted control charts: Theory, methods, and applications in health. *Statistics in Biosciences*, **11**, 630–658.
- Song, L. and Mazumder, J. (2011) Feedback control of melt pool temperature during laser cladding process. *IEEE Transactions on Control Systems Technology*, **19**, 1349–1356.
- Tapia, G. and Elwany, A. (2014) A review on process monitoring and control in metal-based additive manufacturing. *Journal of Manufacturing Science and Engineering*, **136**, 060801.
- Tseng, P. (2001) Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, **109**, 475–494.
- Vasileska, E., Demir, A.G., Colosimo, B.M. and Previtali, B. (2020) Layer-wise control of selective laser melting by means of inline melt pool area measurements. *Journal of Laser Applications*, **32**, 022057.
- Wald, A. (1992) Sequential tests of statistical hypotheses, in *Breakthroughs in Statistics*, Springer, Berlin/Heidelberg, Germany, pp. 256–298.
- Waller, J.M., Parker, B.H., Hodges, K.L., Burke, E.R., Walker, J.L. and Generazio, E.R. (2014) *Nondestructive evaluation of additive manufacturing*, NASA Technical Memorandum, Washington, DC.
- Wang, R., Zhang, L. and Chen, N. (2018) Spatial correlated data monitoring in semiconductor manufacturing using Gaussian process model. *IEEE Transactions on Semiconductor Manufacturing*, **32**, 104–111.
- Wei, G.C. and Tanner, M.A. (1990) A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.
- Wen, H., Liu, L. and Yan, X. (2021) Regression-adjusted Poisson EWMA control chart. *Quality and Reliability Engineering International*, **37**, 1956–1964.
- Wen, Y., Wu, J., Zhou, Q. and Tseng, T.-L. (2018) Multiple-change-point modeling and exact Bayesian inference of degradation signal for prognostic improvement. *IEEE Transactions on Automation Science and Engineering*, **16**, 613–628.
- Wu, B., Ji, X., Zhou, J., Yang, H., Peng, D., Wang, Z., Wu, Y. and Yin, Y. (2021) In situ monitoring methods for selective laser melting additive manufacturing process based on images—A review. *China Foundry*, **18**, 265–285.
- Yang, H., Zhang, S., Lu, Y., Witherell, P. and Kumara, S. (2022) Spatiotemporal monitoring of melt-pool variations in metal-based additive manufacturing. *IEEE Robotics and Automation Letters*, **7**, 8249–8256.
- Zhang, J., Li, Z. and Wang, Z. (2009) Control chart based on likelihood ratio for monitoring linear profiles. *Computational Statistics & Data Analysis*, **53**, 1440–1448.
- Zou, C., Qiu, P. and Hawkins, D. (2009) Nonparametric control chart for monitoring profiles using change point formulation and adaptive smoothing. *Statistica Sinica*, **19**(3), 1337–1357.
- Zou, C., Tsung, F. and Wang, Z. (2007) Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. *Technometrics*, **49**, 395–408.