A novel sparse Bayesian learning and its application to fault diagnosis for multistation assembly systems

Jihoon Chung, Bo Shen, and Zhenyu (James) Kong

ABSTRACT
This article addresses the problem of fault diagnosis in multistation assembly systems. Fault diagnosis is to identify process faults that cause excessive dimensional variation of the product using dimensional measurements. For such problems, the challenge is solving an underdetermined system caused by a common phenomenon in practice; namely, the number of measurements is less than that of the process errors. To address this challenge, this article attempts to solve the following two problems: (i) how to utilize the temporal correlation in the time series data of each process error and (ii) how to apply prior knowledge regarding which process errors are more likely to be process faults. A novel sparse Bayesian learning method is proposed to achieve the above objectives. The method consists of three hierarchical layers. The first layer has parameterized prior distribution that exploits the temporal correlation of each process error. Furthermore, the second and third layers achieve the prior distribution representing the prior knowledge of process faults. Since posterior distributions of process faults are intractable, this article derives approximate posterior distributions via Variational Bayes inference. Numerical and simulation case studies using an actual autobody assembly process are performed to demonstrate the effectiveness of the proposed method.

1. Introduction
Multistation assemblies can be used to represent systems that perform assembly operations on multiple stations in order to assemble a final product. The quality of the final product relies on several factors, called Key Control Characteristics (KCCs) or process errors (Bastani et al., 2018). The positioning accuracy of fixture locators is a KCC in multistation assembly (Bastani et al., 2012). Fixture locators are used to clamp parts during the assembly process, so their deviation from nominal values may cause dimensional quality problems in the final product. Therefore, it is necessary to identify process errors that have mean shifts and/or large variance increases from their design specifications, namely, process faults (Bastani et al., 2016). Hence, fault diagnosis in multistation assemblies estimates the mean and variance of process errors, namely, the variations of fixture locators. This article focuses on process faults due to the mean shift.

Due to the high cost and physical constraints, the dimensional variation of KCCs, namely, fixture locators, cannot be directly monitored using sensors in multistation assemblies (Lee et al., 2020). Instead, Key Product Characteristics (KPCs), namely, the key measurements from the final product, can be used to estimate the KCCs and, consequently, identify process faults among process errors. A fault-quality linear model of multistation process represents the relationship between KPCs and process errors as follows (Huang et al., 2007; Huang et al., 2018):

\[ y = \Phi x + v, \]  

where \( y \in \mathbb{R}^{M \times 1} \) represents \( M \) dimensional measurements (i.e., KPCs), \( x \in \mathbb{R}^{N \times 1} \) denotes \( N \) process errors (i.e., KCCs), \( \Phi \in \mathbb{R}^{M \times N} \) is a fault pattern matrix obtained from all the process information in the multistage process (Ding et al., 2000), and \( v \in \mathbb{R}^{M \times 1} \) denotes the noise. Although the relationship between KPCs and KCCs is nonlinear in many manufacturing processes, a fault-quality linear model is used in general, as the nonlinear relationship could be approximated to a linear model utilizing Taylor series expansion; that is because the variations of KCCs are small, and the relationship between KCCs and KPCs is smooth and without sharp changes (Ding et al., 2002; Lee et al., 2020; Shi, 2023). Since process errors indicate the mean deviations of the fixture locators, process faults refer to nonzero elements in \( x \), namely, process errors with nonzero mean shifts (Bastani et al., 2018).

Following the fault-quality model in (1), some research on fault diagnosis for manufacturing systems has been investigated (Ding et al., 2002; Zhou et al., 2004). All these methods assume that the number of measurements (\( M \)) is greater than the number of process errors (\( N \)) (i.e., \( M > N \)).
However, this assumption may not always hold in actual manufacturing applications, as using an excessive number of sensors (measurements) will result in undue costs (Bastani et al., 2018). These approaches are unsuitable if this assumption is violated. This is because (1) becomes an underdetermined system that results in the non-existence of a unique solution. To overcome the challenge, the sparse solution assumption (Donoho, 2006) that \( x \) in (1) has a minimal number of nonzero elements is required. In the context of the fault diagnosis problem, it denotes the sparsity of process faults in the fault-quality linear model. This is reasonable, as it is likely to have a few process faults in practice (Bastani et al., 2012). Among the several sparse estimation methods, the Bayesian method called sparse Bayesian learning has received considerable attention recently due to its superior estimation performance that is guaranteed from the several theoretical properties (Tibshirani, 1996; Candès and Wakin, 2008; Zhang and Rao, 2011).

Several studies have used sparse Bayesian learning for fault diagnosis in manufacturing systems (Bastani et al., 2012; Li and Chen, 2016; Bastani et al., 2018; Lee et al., 2020). These studies successfully identified process faults by providing prior distribution of process errors (i.e., \( x \) in (1)) to promote the sparsity of process faults. Especially, the work in Bastani et al. (2018) applied Bayesian learning to diagnose mean shift faults, which is the most relevant work to our study. In Bastani et al. (2018), multiple KPCs samples were used to estimate the mean deviation of fixture locators. Specifically, given the average of multiple KPCs samples, which is \( y \) in (1), the mean deviation of the fixture locators (namely, \( x \) in (1)) can be estimated. However, this work does not consider the characteristics of multiple KPCs samples when they are collected sequentially. In practice, time series data of each process error from the multiple KPCs samples may have a strong temporal correlation in the multistation assembly process. For example, the locator position of the fixture system could drift due to wear (Zhou et al., 2004), causing dimensional quality issues in the samples. Specifically, Jin and Chen (2001); Chen et al. (2004) investigated the effect of pin wear over time and dimensional variation in the multistage assembly process. If the process faults are not mitigated immediately, the effect of fixture deviation is auto-correlated in terms of time due to the degradation of wear of production tooling over time (Shi, 2023). Similarly, the dimensional variability caused by machine-tool thermal distortions in the multistation processes is highly correlated between the samples assembled in a certain time period (Abellan-Nebot et al., 2012). Therefore, the temporal correlation of each process error causes product samples manufactured over a period of time to exhibit the same patterns of faults by a specific source of variation (Liu, 2010). The relationship between the sequentially collected multiple KPCs samples and their process errors can be formulated as the following Multiple Measurements Vectors (MMV) model (Cotter et al., 2005) that follows from (1):

\[
Y = \Phi X + V, \tag{2}
\]

where \( Y = [Y_1, \ldots, Y_L] \in \mathbb{R}^{M \times L} \) is a measurement matrix consisting of \( L \) KPCs samples, and \( Y_i \in \mathbb{R}^{M \times 1} \) is a vector that denotes the \( i \)th KPCs sample. \( X = [X_1, \ldots, X_L] \in \mathbb{R}^{N \times L} \) is a process error matrix, where \( X_j \in \mathbb{R}^{N \times 1} \) is a vector that represents process errors (KCCs) of the \( i \)th KPCs sample. \( V \in \mathbb{R}^{M \times L} \) is a noise matrix. Since \( L \) KPCs samples have the same process faults, all columns of \( X \) share the index of nonzero rows called support. It is called a common support assumption in the MMV model (Cotter et al., 2005). In addition, \( L \) elements in the \( j \)th row of \( X \), representing the time series data of the \( j \)th process error, are highly correlated as nonzero values if the \( j \)th process error is a process fault. However, the dynamic changes of process faults due to the complexity of the manufacturing process can easily violate the common support assumption in a large number of KPCs samples. Therefore, this article focuses on a small number of KPCs samples (\( L \)) to satisfy the common support assumption. Utilizing a small number of KPCs samples is also efficient in fault diagnosis of process faults in the multistation assembly process in terms of time and cost reduction.

Beyond the common support assumption and temporal correlation in the time series data of each process error, utilizing prior knowledge of process faults is an additional way to improve the identification of sparse process faults (Lee et al., 2020). Specifically, prior knowledge regarding which process errors are more likely to be process faults than others. This knowledge can be obtained from domain-specific knowledge from practitioners or collected based on the fault diagnosis at past time stamps. For example, based on their experience, manufacturing engineers in an assembly line usually know that some fixture locators may malfunction more frequently than others. In practice, prior knowledge provides only part of the actual process faults. In addition, the knowledge may contain some erroneous information as to process faults, which are actually not faults. Therefore, prior knowledge provides partial and even erroneous information about process faults. However, utilizing prior knowledge is expected to improve the identification of process faults if the correct and erroneous information can be properly distinguished.

In the sparse Bayesian learning literature, there exist studies that consider the temporal correlation of rows in matrix \( X \) in (2) based on the common support assumption (Luessi et al., 2013; Han and Song, 2018), and the work that utilizes the partial with some erroneous prior knowledge of support to improve the performance of sparse estimation (Fang et al., 2015; Guo et al., 2017), separately. However, these studies did not integrate these aspects to improve the sparse estimation.

To address this challenge, this article aims to develop a novel sparse Bayesian hierarchical learning method that simultaneously utilizes the temporal correlation in the time series data of each process error, as well as prior knowledge, which may contain erroneous information. The new method, namely, support knowledge aided with temporally correlated process error sparse Bayesian learning, is proposed to achieve this objective. The contributions of this work are summarized as follows:

1. From the methodological point of view, this article proposes a novel sparse Bayesian learning approach that
considers both temporal correlation in the time series data of each process error and prior knowledge of process faults to improve the sparse estimation. This method also derives an approximate posterior distribution of the sparse solution via Variational Bayes inference (Petersen et al., 2005) to address the intractable computational challenge.

2. From the application perspective, the proposed method is applied to fault diagnosis in multistation assembly systems. The method mitigates dimensional quality issues in the assembly operation by effectively identifying fixture locators that have excessive mean shifts. The effectiveness of the proposed method is validated in real-world simulation case studies that use an actual auto body assembly process.

The rest of this article is organized as follows. A brief review of related research work is provided in Section 2. The proposed methodology is presented in Section 3, followed by numerical case studies to validate its effectiveness in Section 4. Section 5 offers real-world case studies on fault diagnosis problems in multistation assembly process. Finally, conclusions and future work are discussed in Section 6.

2. Review of related work

The related existing studies of fault diagnosis in manufacturing systems are reviewed in Section 2.1. Then, the literature related to sparse Bayesian learning is provided in Section 2.2. Afterward, the research gaps in the existing work are identified in Section 2.3.

2.1. Fault diagnosis methodologies in manufacturing systems

There is a large body of research on fault diagnosis methodologies for manufacturing systems based on the fault-quality model in (1). Kong et al. (2008) developed a Principal Component Analysis (PCA)-based orthogonal diagonalization strategy to transform the measurement data. It enabled the estimation of the variance of process errors in a multistation assembly system. Zhou et al. (2004) applied a mixed linear model to represent the relationship between measurements and process faults in a multistage machining process. Then, it used a maximum likelihood estimator for both mean shift and variance change to detect the process change. Ceglarek and Shi (1996) presented a fault diagnosis method based on fault mapping procedures that combine PCA and pattern recognition. All these approaches assume the number of measurements is greater than the number of process errors (i.e., $M > N$ in (1)), which may not be consistent with industrial practice. However, if this assumption is violated, all of the approaches mentioned above are ineffective, due to the fault-quality linear model becoming an underdetermined system resulting in the non-existence of a unique solution.

To overcome an underdetermined system for fault diagnosis in the manufacturing system, sparse learning, which has been actively used in fault diagnosis and detection in manufacturing systems, can be considered. Han et al. (2018) developed a fault diagnosis method via dictionary learning and sparse representation-based classification for fault diagnosis in manufacturing processes. Zhang et al. (2022) proposed a new non-convex penalty called the generalized logarithm penalty, which enables sparsity and reduces noise disturbance for bearing fault diagnosis. Zhang et al. (2021) developed a dynamical multivariate functional data modeling via sparse subspace learning approach to detect process faults in a manufacturing processes. In addition to fault diagnosis and detection in a general manufacturing process, sparse learning has been widely utilized to address the issue of an underdetermined system in multistage assembly systems. Specifically, sparse Bayesian learning has been actively utilized to incorporate the sparsity of process faults as the prior distribution. Bastani et al. (2012) proposed a fault diagnosis approach that combined a state-space model and the relevant vector machine to figure out process faults using the sparse estimate of the variance change of process errors. Li and Chen (2016) developed a Bayesian variable selection-based method to identify both process faults and sensor faults in the assembly process. Bastani et al. (2018) proposed a spatially correlated sparse Bayesian learning technique to deal with the case when process errors have a spatial correlation. That work is based on the hypothesis that if one of the fixture locators deviates from its design specification, the neighboring locators are also expected to deviate. Lee et al. (2020) presented a Bayesian approach for identifying variation sources in a multitage manufacturing process using the sparse variance component prior. That work focuses on the identification of process faults that have variance increases.

2.2. Sparse Bayesian learning

Since Tipping (2001) and Zhang and Rao (2011) proposed Sparse Bayesian Learning (SBL), many researchers have significantly extended it. Wipf and Nagarajan (2009) first introduced SBL to sparse estimation for the single measurement vector model in (1). Then Wipf and Rao (2007) extended it to the MMV model (2), deriving the MSBL algorithm using the common support assumption. The advantage of SBL and MSBL is that global minima of both methods are always the sparsest solution compared with the minimization-based algorithms (Tibshirani, 1996), where the global minimum is usually not the sparsest solution (Zhang and Rao, 2011). In addition, SBL solutions have much fewer local minima than sparse learning with frequentist approaches such as the FOCUSS family (Cotter et al., 2005; Zhang and Rao, 2011). Based on the MMV model, several studies exploited the temporal correlation to improve the performance of sparse estimation. Zhang and Rao (2011) presented a block SBL framework where a positive definite matrix captures the correlation structure of each row of $X$ in (2). Luessi et al. (2013) established a hierarchical Bayesian framework to model the temporally smooth signals using a multinomial distribution as the prior distribution of each
row of $X$ in (2). Han and Song (2018) provided a Wishart distribution as the prior distribution to learn the temporal correlation from multiple measurement samples. In addition to considering the temporal correlation, work utilizing prior knowledge of the support has recently been actively studied under the SBL framework. To utilize prior knowledge in the Bayesian framework, Fang et al. (2015); Li et al. (2017); Yu et al. (2019) added one more hierarchical layer to conventional SBL to integrate the prior knowledge of support as the prior distribution.

2.3. Research gap analysis

The work summarized in Section 2.1 investigates process faults in multistation assembly systems for quality assurance. Recently, the sparsity of process faults has been used via SBL to deal with the low dimensional measurements in fault diagnosis problems, which is common in industrial practice. However, there is a lack of effort to identify process faults by considering temporal correlation in the time series data of each KCC and prior knowledge of support. Therefore, this article proposes a novel SBL method that simultaneously utilizes temporal correlation from multiple measurement samples. In addition to considering the temporal correlation, work utilizing prior knowledge of support as the prior distribution.

Research efforts in Section 2.2 introduce methods in SBL that consider the temporal correlation and prior knowledge of support. However, these two aspects were not integrated into the reported work. Therefore, this article proposes a novel SBL method that simultaneously utilizes temporal correlation in the time series data of each KCC and prior knowledge of process faults to improve the identification of sparse process faults in multistation systems.

3. Proposed research methodology

This section proposes a novel sparse Bayesian hierarchical method: support knowledge aided temporally correlated process error SBL (SA-TSBL). The proposed SA-TSBL is described in Section 3.1, followed by Bayesian inference of the proposed method in Section 3.2.

3.1. Proposed methodology

The proposed methodology is a sparse Bayesian hierarchical model using multiple KPCs samples that have time series data of each process fault. The method considers the correlation in the time series data of each process error, and utilizes prior knowledge of process faults to improve the sparse estimation. To exploit the temporal correlation in the time series data of each process error, the proposed method transforms the MMV model in (2) to the following block single measurement vector model (Zhang and Rao, 2011), where $\text{Vec}(A)$ is a vectorization of the matrix $A$, $I_L$ is an Identity matrix with the size $L \times L$, and $A \otimes B$ is Kronecker product of the matrices $A$ and $B$.

$$y = Dx + v,$$  \hspace{1cm} (3)

where $y = \text{Vec} (Y^T) \in \mathbb{R}^{ML \times 1}$, $D = \Phi \otimes I_L$, $x = \text{Vec} (X^T) \in \mathbb{R}^{NL \times 1}$. Assume noise vector $v$ follows a Gaussian distribution with zero mean and variance $\lambda$. In fact (3) can be rewritten as

$$y = [\phi_1 \otimes I_L, ..., \phi_N \otimes I_L] [x_1^T, ..., x_N^T]^T + v,$$  \hspace{1cm} (4)

where $\phi_i$ is the $i$th column of $\Phi$, and $x_i$ consists of the $i$th process errors of $L$ KPCs samples (i.e., $x_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, ..., \tilde{x}_{iL})$, where $\tilde{x}_{ij}$ denotes the element in the $i$th row and the $j$th column of matrix $X$ in (2)). In other words, $x_i \in \mathbb{R}^{L \times 1}$ is the $i$th block of $x$ in (3), as illustrated in Figure 1. Therefore, $K$ nonzero rows in $X$ in (2) are represented as $K$ nonzero blocks in $x$ in the proposed method. It implies that there exist $K$ process errors that have shifted from the design specification, namely, process faults. Since $x_i$ consists of the time series data of $i$th process error from $L$ KPCs samples, the proposed method exploits the correlation between the $L$ elements of $x_i$. Since noise vector $v$ follows Gaussian distribution with zero mean and variance $\lambda$, the Gaussian likelihood is provided for the block model in (3) as follows:

$$p(y|x; \lambda) \sim N(Dx, \lambda I_{ML}).$$

The proposed method consists of the following three layers. The prior distribution in the first layer is provided to exploit the correlation of the time series data of each KCC. The second and third layers consist of prior distributions representing the prior knowledge of process faults among process errors. A graphical representation of the proposed method is shown in Figure 2.

In the first layer of the hierarchical model in Figure 2, the Gaussian distribution is provided as the prior distribution for process error ($x$) since it is the most common distribution considered for process errors in the literature and practice (Bastani et al., 2012; Lee et al., 2020). The prior distribution is given by

$$p(x|z; B) \sim N(0, \Sigma_0),$$

where $\Sigma_0$ is

$$\Sigma_0 = \begin{bmatrix} z_1^{-1}B_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & z_N^{-1}B_N^{-1} \end{bmatrix},$$

$z = \{z_1, ..., z_N\}$, and $B = \{B_1, ..., B_N\}$. Since $z_i$ consists of the time series data of the $i$th process error from $L$ KPCs samples, $z_i^{-1}$ controls the sparsity of the $i$th process error ($x_i$). For example, when $z_i^{-1}$ converges to zero, the associated block $x_i$ will be driven to zero (Tipping, 2001). $B_i$ is a positive definite matrix that captures the correlation of the time series data of the $i$th process error. The proposed method assumes independence between process errors (e.g., different fixture locators), which is a common assumption in the literature (Bastani et al., 2016; Li and Chen, 2012).
It results in the block diagonal matrix $\Sigma_0$.

The second layer in Figure 2 specifies Gamma prior distributions over $a$. The distributions have an individual rate parameter $b_i$ for each parameter $x_i$, namely,

$$p(x|b) = \prod_{i=1}^{N} \text{Gamma}(x_i|a, b_i) = \prod_{i=1}^{N} \frac{\Gamma(a)}{\Gamma(a-b_i)} x_i^{-a} e^{-b_i x_i}, \quad (5)$$

where $b = \{b_1, ..., b_N\}$. (5) effectively integrates the prior information of process faults into the SBL framework. Basically, $a$ and $b$ are set to be very small values (e.g., $10^{-4}$) to provide a large variance of a prior distribution over $x_i$ in SBL (Tipping, 2001). It encourages large values of $x_i$ and promotes the $i$th process errors as non-process faults (i.e., $x_i = 0$ in (4)) (Fang et al., 2015). Specifically, the marginal distribution of the time series data of the $i$th process error, namely, $p(x_i) = \int p(x_i|x, B)p(x_i|a, b_i)dx_i$, follows a multivariate student-t distribution, in which the probability is concentrated at zero (Tipping, 2001). Assuming that set $P$ consists of the indexes of process errors that are likely to be process faults based on prior knowledge. Then, the corresponding rate parameters $\{b_i, i \in P\}$ are set as a relatively large value (e.g., one) to provide a small variance of the prior distribution over $x_i$ (Fang et al., 2015). The prior distribution allows the small value of $x_i$ and encourages the $i$th process errors to be process faults (i.e., $x_i \neq 0$ in (4)). However, assigning a fixed value to $b_i$ in (5) creates limitations in dealing with the situation when the set $P$ has some erroneous information about process faults.

To address this issue, the third layer in Figure 2 assigns a prior distribution over $\{b_i, i \in P\}$ (Li et al., 2017). A Gamma distribution is assigned, due to the support of the variable. In addition, the distribution enables users to derive the closed form of the approximate posterior distribution of $\{b_i, i \in P\}$, which will be described in the following section. The prior distribution over $\{b_i, i \in P\}$ is provided as follows:

$$\text{Gamma}(b_i|p, q) = \frac{1}{\Gamma(p)^{q(b_i+1)}} b_i^{-p} e^{-q b_i}, \quad i \in P. \quad (6)$$

$p$ and $q$ in (6) are specified to characterize the prior distribution. These two values in the proposed method can be adjusted by the users to satisfy the following two conditions as follows:

- First, the values are set to provide a large mean value of the prior distribution of $\{b_i, i \in P\}$.
- Second, the values are determined to have a large variance in the prior distribution of $\{b_i, i \in P\}$.

The first condition encourages the $i$th process errors as process faults based on prior knowledge. Compared with a small fixed value of $b_i = 10^{-4}$ that was used to promote a large $x_i$ in previous research (Tipping, 2001; Fang et al., 2015), the large mean value of the prior distribution of $\{b_i, i \in P\}$ creates a small $x_i$. The small $x_i$ provides a prior distribution such that the $i$th process errors are likely to be process faults of the mean shift.

The second condition is designed to deal with the case when erroneous prior knowledge is provided. Assume incorrect prior knowledge that the $i$th process error may be a process fault is provided. If the variance of the prior distribution of $\{b_i, i \in P\}$ is small, the $i$th process error is likely to be misdiagnosed as a process faults, due to the large mean value of the prior distribution of $\{b_i, i \in P\}$ from the first condition. However, if the variance of the prior distribution of $\{b_i, i \in P\}$ is large, $\{b_i, i \in P\}$ is not significantly affected by the prior distribution. Instead, $\{b_i, i \in P\}$ is highly affected by the data itself, where the $i$th process error is not the process fault. This allows $\{b_i, i \in P\}$ to be learned as a small value from data and provides the sparsity prior to the $i$th process error (Fang et al., 2015).

For $\{b_i, i \in P\}$ in the proposed method, the parameters are set to a fixed small value of $10^{-4}$ as in the previous study (Tipping, 2001).

### 3.2. Bayesian inference of the proposed methodology

The proposed method in Section 3.1 has several hidden variables that need to be estimated, these include, the process errors ($x$), the variable controlling the sparsity of process errors ($z$), and the variable used to provide prior knowledge of process faults ($b_i, \forall i \in P$). In addition, there are hyper-parameters related to temporal correlations in the time
series data of the \( i \)th KCCs \((B_i)\) and noise \((\lambda)\), which also require estimation. To avoid too many parameters to being estimated, causing a challenging task in sparse estimation, the proposed method sets \( B_i = B \) (Zhang and Rao, 2011).

However, the posterior distribution of hidden variables in the proposed method (7) does not have a closed form, due to the complexity of the proposed hierarchical model in Section 3.1. Specifically, the denominator in (7) cannot be calculated in closed form. Let \( b_i \) \((i \in P)\) as \( \tilde{b} \) for convenience.

\[
P(x, z, \tilde{b}|y) = \frac{P(y, x, z, \tilde{b})}{\sum_{x, z, \tilde{b}} P(y, x, z, \tilde{b})} \tag{7}
\]

To overcome this challenge, this article derives approximate posterior distributions of hidden variables via Variational Bayes Inference (VBI). Specifically, Variational Bayes Expectation Maximization (VBEM) (Petersen et al., 2005) is utilized to estimate hidden variables and hyperparameters in the proposed method to identify mean shift process faults. VBEM consists of (i) E-step: Variational Bayesian expectation step to estimate hidden variables \( x, z, \tilde{b} \) by approximating the posterior distribution of hidden variables; and (ii) M-step: Variational Bayesian maximization step to update hyperparameters \( B \) and \( \lambda \) by maximizing the expected value of the logarithm of the complete likelihood.

Let \( \theta \) be a vector of all hidden variables in the proposed method \((i.e., \theta = (x, z, \tilde{b}))\). VBI approximates the posterior distribution of \( \theta \), denoted as \( q(\theta) \), by minimizing the Kullback–Leibler (KL) divergence between \( q(\theta) \) and the true posterior distribution, namely, \( p(\theta|y) \) \((i.e., D_{KL}(q(\theta)||p(\theta|y)))\). \( q(\theta) \) is factorized as

\[
q(\theta) = q(x)q(z)q(\tilde{b})
\]

by the mean-field approximation (Cohn et al., 2010), that all the variables are independent. The approximate posterior distribution \( q(\theta) \), where \( \theta_i \) is the \( i \)th element in the set \( \theta \) is derived as follows by minimizing the \( D_{KL}(q(\theta)||p(\theta|y)) \) under the mean-field approximation.

\[
\ln q(\theta) = \mathbb{E} \ln p(\theta|y) + \text{const}, \tag{8}
\]

where \( \mathbb{E}_{\theta|y} \) denotes the expectation taken with the set \( \theta \) without \( \theta_0 \), can be obtained through normalization. The detailed derivation of (8) is provided in the Supplemental Online Materials. Equation (8) is used in the following E-step of VBEM to approximate the posterior distributions of hidden variables.

**E-step of VBEM:** The posterior distributions of hidden variables that are related to process errors \((x)\), sparsity of process errors \((z)\), and prior knowledge of process faults \((\tilde{b})\) are approximated by (8), respectively, as follows:

\[
\ln q(x) = \mathbb{E} \ln p(y|x, \tilde{x}, \tilde{b})q(x|\tilde{b}) + \text{const}
\]

\[
\ln q(z) = \mathbb{E} \ln p(y|x, z, \tilde{b})q(z|x)q(\tilde{b}) + \text{const}
\]

\[
\ln q(\tilde{b}) = \mathbb{E} \ln p(y|x, \tilde{x}, \tilde{b})q(\tilde{b}) + \text{const}
\]

Based on the statistical inference, the posterior distributions of hidden variables can be derived as

\[
q(x) = N(x|\mu_x, \Sigma_x), \tag{12}
\]

\[
q(z) = \prod_{i=1}^{N} \text{Gamma}(\alpha_i|\tilde{a}_i, \tilde{b}_i), \tag{13}
\]

\[
q(\tilde{b}) = \prod_{i \in P} \text{Gamma}(b_i|p + a, \mathbb{E}[x_i] + q). \tag{14}
\]

Based on the mean-field approximation, \( q(\theta) \) is the multiplication of Gaussian and Gamma distributions from (12), (13), and (14). Let \( \text{diag}(x_1, ..., x_m) \) be a diagonal matrix with principal diagonal elements being \( x_1, ..., x_m \) and \( \text{Tr}(A) \) be a trace of matrix \( A \). Then, the expectations and moments of distributions in (12), (13), and (14) are

\[
\mu_\lambda = \frac{1}{\lambda} \Sigma_x D^T y, \tag{15}
\]

\[
\mathbb{E}[x_i] = \frac{a + \frac{b}{2}}{\text{Tr}[\Sigma_x + \Sigma(x_{x[i]})] + \mathbb{E}[b_i]}, (\forall i), \tag{16}
\]

\[
\mathbb{E}[b_i] = \frac{p + a}{q + \mathbb{E}[a]}, (i \in P), \tag{17}
\]

\[
\Sigma_x = \left( \frac{1}{\lambda} D^T D + \mathbb{E}[AB] \right)^{-1}, \tag{18}
\]

where \( \mathbb{E}[AB] = \text{diag} \mathbb{E}[x_1 B, ..., \mathbb{E}[x_N] B] \). The estimator of the time series data of the \( i \)th process error \((\mathbb{E}[x_i])\) can be obtained from \( \mu_\lambda \) in (15) as follows:

\[
\mathbb{E}[x_i] = \mu_\lambda ((i - 1)L + 1 : iL), i = 1, ..., N. \tag{19}
\]

Detailed derivations of (12), (13), and (14) are provided in the Supplemental Online Materials.

**M-step of VBEM:** Temporal correlations in the time series data of each KCC \((B)\) and noise \((\lambda)\) are estimated in this step. Let \( \hat{\theta} = \{B, \lambda\} \). Posterior distributions of \( x, \tilde{b} \) obtained in (12), (13), and (14) are denoted as \( q(x; \hat{\theta}^{OLD}), q(z; \hat{\theta}^{OLD}), \) and \( q(\tilde{b}; \hat{\theta}^{OLD}) \) respectively. \( \hat{\theta} \) can be updated by maximizing the complete likelihood as follows:

\[
\hat{\theta}^{NEW} = \arg\max_{\theta} \mathbb{E} \ln p(y|x, \tilde{x}, \tilde{b}; \theta)q(x; \theta^{OLD})q(\tilde{b}; \theta^{OLD}),
\]

\[
= \arg\max_{\theta} \mathbb{E} \ln p(y|x; \theta)p(x|\tilde{x}; \theta)q(x; \theta^{OLD})q(\tilde{b}; \theta^{OLD}).
\]

Let

\[
Q(\theta) = \mathbb{E} \ln p(y|x; \theta)p(x|\tilde{x}; \theta),
\]

which results in

\[
Q(B, \lambda) = \mathbb{E} \ln p(y|x; \lambda)q(x; \theta^{OLD}) + \mathbb{E} \ln p(x|\tilde{x}; B)q(\tilde{b}; \theta^{OLD}),
\]

\[
= \mathbb{E} \ln p(y|x; \lambda)q(x; \theta^{OLD}) + \mathbb{E} \ln p(x|\tilde{x}; B)q(\tilde{b}; \theta^{OLD}). \tag{20}
\]
B and \(\lambda\) are estimated as (21) and (22), respectively, by maximizing the (20).

\[
B = \left[\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[x_i](\Sigma_{x_i} + \mathbb{E}[x_i]\mathbb{E}[x_i]^T)^{-1}\right].
\tag{21}
\]

\[
\lambda = \frac{\|y - D\mu_x\|_2^2 + \lambda NL - \text{Tr}(\Sigma_x E_q(x, \theta^{\text{true}})(\Sigma^{-1}_0))}{\text{ML}},
\tag{22}
\]

where \(\|x\|_p\) denotes \(\ell_p\) norm of the vector \(x\). Detailed derivations of (21) and (22) are described in the Supplemental Online Materials.

Algorithm 1 shows the procedure of the proposed SA-TSBL method. Given the multiple KPCs samples (\(y\)) and fault pattern matrix (\(\Phi\)), the proposed method estimates the following variables and parameters in E and M steps, respectively:

- E-step: Process errors (\(\mu_x\)), the variable related to the sparsity of process errors (\(x\)), and the variable that is used to provide prior knowledge of process faults (\(b\)).
- M-step: Temporal correlations in the time series data of each KCC (\(B\)) and noise (\(\lambda\)).

These steps iterate until the estimator of process errors (\(\mu_x\)) is rarely updated, namely, \(\|\mu_x^{-1} - \mu_x\|_\infty < \gamma\), where \(\|\cdot\|_\infty\) indicates the infinity norm and \(\gamma\) is a user-defined threshold (e.g., \(\gamma = 10^{-6}\)) or it reaches the maximum number of iterations (T). Then, the mean deviation of process errors, which is the output of Algorithm 1, are calculated by the following procedure:

- Step 1: The matrix \(\mu_x\) is defined in which the \(i\)th row represents (19) derived from \(\mu_x\).
- Step 2: The average of each row of matrix \(\mu_x\), that is, a vector \(\hat{\mu}_x\), is provided as the output of Algorithm 1, indicating the mean deviation of process errors.

Therefore, nonzero values in a vector \(\hat{\mu}_x\) are process faults of the mean shifts.

\[\text{Algorithm 1: Proposed SA-TSBL method}\]

**Input:** Multiple KPCs samples (\(y\)), Fault pattern matrix (\(\Phi\)).

**Set:** \(a = b_i(i \in P) = 10^{-4}\). \(p\) and \(q\) are set based on two conditions in Section 3.1.

**Initialize:** \(B = I_L, \lambda = 1, b_i(i \in P) = 1, t = 1\).

**While** \(\|\mu_x^{-1} - \mu_x\|_\infty \geq \gamma\) or \(t \leq T\) do

**E-step of VBEM:**
- Update \(\mu_x\) using (15)
- Update \(x\) using (16)
- Update \(b\) using (17)

**M-step of VBEM:**
- Update \(B\) using (21)
- Update \(\lambda\) using (22)

\(t = t + 1\)

End

**Output:** Mean deviations of process errors \(\hat{\mu}_x\).

4. Numerical case studies

This section provides three scenarios to compare the performance between the proposed method and benchmark methods:

- Section 4.1 shows the performance evaluation by varying the temporal correlation. This study is to validate the effectiveness of strong temporal correlation in sparse estimation.
- Section 4.2 provides the numerical study to investigate the impact of the number of KPCs samples (i.e., measurement samples) on the performance of process faults identification.
- Section 4.3 illustrates the sparse estimation performance by varying the ratio between the number of measurements and process errors (i.e., the severity of the underdetermined systems).

All the numerical case studies consist of 100 independent trials. \(p\) and \(q\) in (6) are determined as 1 and 0.1, respectively, in the case studies to satisfy the two conditions provided in Section 3.1. The code of the proposed SA-TSBL algorithm is implemented in Matlab 2021a. The CPU of the computer used in this article is an Intel® Xeon® E-2286M 2.40 GHz Processor.

The benchmark methods selected in this study are as follows, which are widely used in SBL:

- MSBL proposed in Wipf and Rao (2007) is a basic SBL method for the MMV model that assumes independence in the time series data of each KCC.
- T-MSBL proposed in Zhang and Rao (2011) is a typical SBL method for the MMV model that considers temporal correlation in the time series data of each KCC.
- SCBL proposed in Bastani et al. (2018) is the SBL method that considers prior knowledge of support in the MMV model. It assumes independence in the time series data of each KCC.
- SA-MSBL proposed in Yu et al. (2019) is the SBL method that considers prior knowledge of support in the MMV model. It assumes independence in the time series data of each KCC.
- SA-SBL proposed in Fang et al. (2015) and Li et al. (2017) is the SBL method that considers prior knowledge of process faults in the SMV model.
- Lee et al. (2020) proposed the SBL method that considers prior knowledge of the support in the MMV model. It assumes independence in the time series data of each KCC.

**Data Generation:** The data generation process is summarized in Figure 3. The solution matrix \(X_{true} \in \mathbb{R}^{N \times L}\) is randomly generated with \(K\) nonzero rows. Indexes of the nonzero rows are randomly chosen in each trial. The
nonzero rows in $X_{true}$ are generated as an AR(1) process that initiates from the standard Gaussian distribution, since AR(1) processes are sufficient to represent the temporal structure of the small number of measurement samples ($L$) (Zhang and Rao, 2011). The AR coefficient, defined as $\beta$, represents the temporal correlation. A dictionary matrix $\Phi \in \mathbb{R}^{M \times N}$ is constructed with columns drawn from the surface of a unit hyper-sphere uniformly (Donoho, 2006). Finally, the measurements matrix is built by $Y = \Phi X_{true} + V$ in the final step of Figure 3, where $V$ is a Gaussian noise matrix with zero-mean (Zhang and Rao, 2011). The variance of the noise matrix is chosen to meet the determined value of the Signal-to-Noise Ratio (SNR). SNR is defined by $\text{SNR}(\text{dB}) = 20 \log_{10}(\frac{\|\Phi X_{true}\|_F}{\|V\|_F})$ (Zhang and Rao, 2011).

**Performance Evaluation:** Since the objective of this article is to diagnose the mean deviation of process errors, the target of the proposed method is to accurately estimate $X_{true}$, which is the row-wise mean of $X_{true}$.

Two performance measures are used in this article. One is the failure rate defined in Li et al. (2017). It measures the accuracy of detecting process faults, which are the nonzero rows in $X_{true}$. Assume that the number of process faults $K$ is given. Then, the row indexes of the $K$ largest $\ell_2$-norms from $\tilde{\mu}_i$ in Algorithm 1 are identified. If the indexes are different from the indexes of nonzero rows in $X_{true}$, it is considered to be a failed trial. The failure rate is the percentage of failed trials in the total trials. The other performance measure is a Normalized Mean Squared Error (NMSE) that is defined by $\frac{\|\tilde{X} - X_{true}\|_F^2}{\|X_{true}\|_F^2}$, where $\|A\|_F$ is a Frobenius norm of the matrix $A$. Averages of failure rate and NMSE from 100 trials are used as performance measures. In addition, the standard deviations and quantile statistics of the two measures are provided in the Supplemental Online Materials.

### 4.1. Performance evaluation in various temporal correlations

This case study shows the performance of all methods in various temporal correlations when the SNR is 100dB. The size of the dictionary matrix $\Phi$ is $5 \times 40$. The number of nonzero rows of $X_{true}$ ($K$), and measurement samples ($L$) are four and four, respectively. Temporal correlation ($\beta$) varies among 0.0, 0.3, 0.6, 0.9, and 0.95. The proposed method shows the best performance in all temporal correlations, as shown in Table 1.

Specifically, the failure rate and NMSE of the proposed method and T-MSBL tend to decrease as $\beta$ increases, since these two methods capture the temporal correlation. In contrast, both measures of MSBL, SA-MSBL, and Lee et al. (2020) show the opposite trend as the three methods assume independence among samples.

### Remark 1. (Performance Evaluations of SA-TSBL, SA-MSBL, SA-SBL, and Lee et al. (2020))

The SA-MSBL, SA-SBL, Lee et al. (2020), and the proposed SA-TSBL have several cases based on the set of prior knowledge (i.e., $P$ in (6)) even in the same problem. Both correct and erroneous information of support exists in the set $P$ as prior knowledge. To differentiate between the correct and erroneous information of support, two subsets, namely, $P_C$ and $P_E$ are defined. $P_C$ consists of the indexes of nonzero rows obtained from prior knowledge among true nonzero rows in $X_{true}$. In contrast, $P_E$ consists of the indexes of nonzero rows in prior knowledge but are actually zero in $X_{true}$. The cardinalities of $P_C$ and $P_E$ in this article are assumed to satisfy two conditions. The first condition is the cardinality of set $P_C$ and $P_E$ should be less than or equal to 75% and 50% of the number of nonzero rows in $X_{true}$ (i.e., $K$), respectively. For example, if $K$ is four, the cardinality of $P_C$ and $P_E$ are less than or equal to four and three, respectively. The first condition illustrates the cardinality of prior knowledge is similar to $K$. The condition enables the performance evaluation of the proposed method in comprehensive situations. Since the cardinality of $P_C$ is less than or equal to four, the prior knowledge is partial (i.e., $P_C$ misses some rows that are actually nonzero in $X_{true}$) in all cases. In addition, the cases with a cardinality of $P_E$ greater than zero show the situations when the erroneous prior knowledge (i.e., $P_E$ contains the rows that are actually zero in $X_{true}$) exist. The second condition is the cardinality of $P_C$ is greater than or equal to that of $P_E$. The second condition prevents erroneous prior knowledge from being dominant prior knowledge. Therefore, performances of these three methods are evaluated as the average of all cases with different cardinalities of $P_C$ and $P_E$, where each case consists of 100 trials.
maximum value sequentially from the bottom. The minimum value denotes the smallest value within the $1.5/C_2$ Inter Quartile Range (IQR) below the first quartile. Similarly, the maximum value is defined as the largest value that is within the $1.5/C_2$ IQR above the third quartile. A blue square in Figure 4 shows the performance of the proposed method without any prior knowledge (i.e., $j_{PC} = j_{PE} = 0$), and the black dot represents the average of all cases used as a performance evaluation measure of the proposed method. The trend of black dots shows the performance improvements of the proposed method as temporal correlation (b) increases. In addition, boxplots in Figure 4 represent most cases with partial and some erroneous prior knowledge achieve better performance in both measures than those without prior knowledge. The results show the effectiveness of prior knowledge in the proposed method, which successfully distinguishes the correct and incorrect prior knowledge in various temporal correlations.

### 4.2. Performance evaluation in the various numbers of measurement samples

The objective of this case study is to compare the performance of all methods using the various numbers of measurement samples (L). The size of the dictionary matrix $\Phi$ is $7 \times 55$, and the number of nonzero rows in $X_{\text{true}} (K)$ is four. SNR and temporal correlation (b) are set as $35$ dB, and $0.95$, respectively. The number of measurement samples varies from two to four. Table 2 shows that the proposed method achieves the best performance in all cases. In addition, the results in Table 2 illustrate that most of the MMV models achieve performance improvements as the number of measurement samples increases, this is due to the common support assumption (Cotter et al., 2005). Figure 5 shows the performance evaluation of the proposed method with various sizes of $PC$ and $PE$. Especially, the performance of T-MSBL and the proposed method without prior knowledge (blue square in Figure 5) is similar. However, the proposed method achieves a better performance than T-MSBL by utilizing partial and some erroneous prior knowledge. Specifically, the performance of the proposed method improves from 3% to 32% of those without any prior knowledge (an average of relative improvement from square to dots in Figure 5) in the various numbers of measurement samples.

### 4.3. Performance evaluation in various ratios between the number of measurements and process errors

Results in Table 3 illustrate the performance of all methods by varying the ratio between the number of measurements and process errors (i.e., underdetermined ratio). In this study, the number of measurements (M) is fixed at 10, and the underdetermined ratio (N/M) is selected from three, five, seven and nine with $25$ dB for SNR, respectively. The number of nonzero rows in $X_{\text{true}} (K)$ and measurement samples (L) are set as four and three, respectively. The temporal correlation (b) is $0.99$. As the underdetermined ratio increases, it becomes more challenging to identify the sparse process faults. However, capturing temporal correlation and utilizing the partial and some erroneous prior knowledge enables the proposed method to achieve

#### Table 1. Performance comparison with various temporal correlations (b).

| b     | Failure Rate | NMSE |
|-------|--------------|------|
| 0.0   | 0.38         | 0.76 |
| T-MSBL| 0.34         | 0.70 |
| MSBL  | 0.64         | 0.96 |
| SA-MSBL| 0.96        | 0.96 |
| SA-SBL| 0.97         | 0.97 |
| Lee et al. (2020) | 0.93 | 0.99 |
| SCBL  | 0.94         | 0.95 |
| SA-TSBL (Proposed) | 0.23 | 0.23 |

#### Table 2. Performance comparison with various number of measurement samples (L).

| L   | Failure Rate | NMSE |
|-----|--------------|------|
| 2   | 0.59         | 0.76 |
| T-MSBL | 0.17         | 0.19 |
| 3   | 0.83         | 1.10 |
| MSBL | 0.77         | 1.04 |
| 4   | 0.88         | 0.68 |
| SA-MSBL | 0.90        | 0.67 |
| 4   | 0.89         | 0.70 |
| SA-SBL | 0.90        | 0.67 |
| 4   | 0.71         | 0.63 |
| Lee et al. (2020) | 0.50 | 0.52 |
| 4   | 0.97         | 1.25 |
| SCBL | 0.92         | 1.29 |
| 4   | 0.97         | 1.17 |
| SA-TSBL (Proposed) | 0.45 | 0.52 |

Figure 4. The boxplot for (a) failure rate and (b) NMSE in all cases of SA-TSBL in various temporal correlations.

Figure 5. The boxplot for (a) failure rate and (b) NMSE in all cases of SA-TSBL in various temporal correlations.

4.2. Performance evaluation in the various numbers of measurement samples

4.3. Performance evaluation in various ratios between the number of measurements and process errors
the best performance even in a high underdetermined ratio. Figure 6 shows prior knowledge of support is still valuable in various underdetermined ratios. This study shows that the proposed algorithm can be applied to applications such as neuroimaging where highly underdetermined systems exist.

5. Real-world simulation case studies

An assembly model from a real auto body assembly process is utilized as a real-world case study. As shown in Figure 7, the assembled product is a floor pan of a car. It consists of four parts namely, the left floor pan, left bracket, right floor pan, and right bracket. They are assembled in three stations as illustrated in Figure 8. During the assembly process, the parts are held by fixtures and part-mating features, which are the KCCs in this process (Bastani et al., 2016). KPCs are measured at four points, namely, M1, M2, M3, and M4, respectively, as shown in Figure 7. Each part has its own measurement place such as M1 on part 1, M2 on part 2, M3 on part 3, and M4 on part 4. These measurements can be measured in each station when the corresponding part has been assembled in the previous stations. For example, M4 on part 4 cannot be measured in station 1 because part 4 has not yet been assembled at station 1 (Bastani et al., 2016). In addition, KPCs are measured in three directions (X, Y, and Z) at each point. In this assembly process, a total of 33 process errors exist, which are fixture locator dimensional errors (Bastani et al., 2018). The dimension of the fault pattern matrix is 12 × 33, and is established based on the literature (Huang et al., 2007; Kong et al., 2009; Bastani...
et al., 2016). Since the variations of KCCs can be propagated to deviations of other KCCs in the subsequent stations, the transfer and accumulation of deviations in KCCs between multiple stations are considered in the formulation of the fault pattern matrix \( U \) (Ding et al., 2000; Bastani et al., 2016). The matrix is provided in the Supplemental Online Materials. Since the number of measurements (12) is less than the number of process errors (33), an underdetermined system is created in the fault quality linear model. It sparse estimation is requires to identify process faults.

To generate the multiple KPCs samples from the auto body assembly process, the time series data of each process fault is generated from an AR(1) process initiated from the standard Gaussian distribution as in Section 4. The generated temporal correlated process faults \( (X_{true}) \) and fault pattern matrix provide the multiple KPCs samples \( (Y) \), as in Figure 3. In addition, performance evaluation measures, benchmark methods, and the values of \( p \) and \( q \) in (6) used in Section 4 are still utilized in Section 5. Prior knowledge of process faults is also provided in the same way as Section 4. Sections 5.1 and 5.2 show the performance evaluation by varying the number of KPCs samples, and temporal correlation \( \beta \), respectively.

### 5.1. Performance evaluation in the various number of KPCs samples

This case study aims to demonstrate the effectiveness of sparse estimation of process faults by varying the number of KPCs samples \( (L) \) when a strong correlation exists \( (\beta = 0.99) \). The number of KPCs samples varies from two to four, and SNR is 50 dB. Three process faults \( (K) \) are randomly determined among 33 process errors. The proposed method shows the best performance in all various numbers of KPCs samples, as shown in Table 4. All methods except SA-SBL generally tend to improve performance in both measures as the number of KPCs samples increases. However, the performances of all methods do not improve significantly as the number of KPCs samples increases, compared with Section 4.2. The reason for this behaviour could be the fault pattern matrix in this multistation assembly process, which is very structured compared with the random design matrix \( \Phi \) in Section 4. The random matrix achieves an accurate sparse estimation based on some theoretical properties of the random matrix, such as low mutual coherence that measures the highest correlation between columns of \( \Phi \) (Candès and Wakin, 2008). In contrast, if the mutual coherence is high, it causes inaccurate sparse estimation. As shown in the Supplemental Online Materials, the columns of fault pattern matrix \( \Phi \) in the assembly process are highly correlated, and mutual coherence is one, causing the challenging sparse estimation task compared with Section 4.2. Therefore, the results in this section are not significantly improved compared with Section 4.2, which uses a random matrix with low mutual coherence, even if the number of KPCs samples is increased. Figure 9 shows that utilizing prior knowledge of process faults is still effective in the proposed method for failure rate and NMSE, even in the structured design matrix \( \Phi \). The property lets the proposed method obtain a more accurate sparse estimation than T-MSBL, MSBL, and SCBL that cannot incorporate the prior knowledge of process faults.

### 5.2. Performance evaluation in various temporal correlations

This case study presents the performance of all methods in various temporal correlations \( \beta \), when there exist five KPCs samples \( (L) \). Temporal correlation varies among 0.0, 0.3, 0.6,
The study has three process faults ($K$) with noise level 80 dB. Table 5 shows the proposed method achieves the best performance, and the performance of other methods exhibit similar trends to the case in Section 4.1. The performance of the proposed method and T-MSBL improves in general as $b$ increases, and MSBL and SA-MSBL show the opposite trends, since they assume the independence in the time series data of each process error. Lee et al. (2020) shows a poor performance, which might be caused by the high mutual coherence of the fault pattern matrix. Figure 10 illustrates the prior knowledge of process faults is still effective in various temporal correlations even with the structured matrix $\Phi$.

### 6. Conclusions

This article proposes a novel sparse hierarchical Bayesian method, SA-TSBL, to effectively identify the sparse process faults in multistation assembly systems. The method identifies process faults by considering the temporal correlation of each KCC and utilizing partial, with some erroneous, prior knowledge of process faults. Since posterior distributions of process errors in the proposed method are computationally intractable, this article derives approximate posterior distributions of process errors via VBI. The effectiveness of the proposed method is validated by both numerical cases and real-world simulation application that uses an actual auto body assembly system. Based on these studies, it is evident that the direct use of multiple KPCs samples in the proposed SA-TSBL is more effective to process mean shift identification than in a previous study using the average of multiple KPCs samples in the SMV model ((1)). This is because the proposed method can fully utilize multiple KPCs samples without any information loss. In addition, the results in case studies highlight that the proposed method achieves a high performance in process faults estimation when the time series data of each KCC have a strong temporal correlation. Furthermore, utilizing the prior knowledge of process faults improves the performance of the proposed method in the case studies, even if the knowledge has some erroneous information. This is possible through the Bayesian framework in the proposed method that distinguishes between the correct and incorrect prior knowledge. Practically, the proposed method improves at least 27% and 28% of the failure rate and NMSE of Bastani et al. (2018); Lee et al. (2020), the state-of-the-art sparse Bayesian
diagnosis method for the process faults in the multistage assembly process in real-world simulation case studies.

In this work, all fixture locators are assumed to be independent, which is a common assumption in the literature. However, the fixture locators in the multistation assembly process have some spatial correlation with each other if the locators are in the composite tolerance mode (Kong et al., 2006). Therefore, considering the spatial correlation among the process errors is a promising direction for future work. In addition, instead of utilizing the Gamma distribution in (5) enforcing the sparsity of process error in the proposed method, other distributions such as Laplacian distribution can also be considered as prior distribution since the probability of these distributions is concentrated at zero.

Funding
This project was funded by the following grant from the Department of Defense: N00014-19-1-2728.

Notes on contributors
Jihoon Chung received his BS degree in industrial engineering from Hanyang University, Seoul, Korea, in 2015. He obtained his MS degree in industrial and systems engineering at the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2017. He is currently pursuing a PhD degree in industrial and systems engineering from Virginia Tech, Blacksburg, VA, USA. His research interests include statistical learning and data analytics in smart manufacturing.

Bo Shen is an assistant professor in the Department of Mechanical and Industrial Engineering at the New Jersey Institute of Technology. He received his PhD in industrial and systems engineering at Virginia Tech, Blacksburg, VA, in August 2022. He also received his BS degree in statistics from the University of Science and Technology of China, Hefei, China, in July 2017. His research interests include optimization and machine learning, and data analytics in smart manufacturing.

Zhenyu (James) Kong received his BS and MS degrees in mechanical engineering from Harbin Institute of Technology, China, in 1993 and 1995, respectively, and his PhD degree from the Department of Industrial and System Engineering, University of Wisconsin Madison, Madison, WI, USA, in 2004. He is currently a professor with the Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA. His research interests include sensing and analytics for smart manufacturing, and modeling, synthesis, and diagnosis for large and complex manufacturing systems. He is a fellow of the Institute of Industrial and Systems Engineers and the American Society of Mechanical Engineers. He was recognized as one of the 20 Most Influential Academics in Smart Manufacturing by the Society of Manufacturing Engineering in 2021.

ORCID
Zhenyu (James) Kong http://orcid.org/0000-0002-8827-502X

Data availability statement
The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

References
Abellan-Nebot, J.V., Liu, J., Subirón, F.R. and Shi, J. (2012) State space modeling of variation propagation in multistation machining processes considering machining-induced variations. Journal of Manufacturing Science and Engineering, 134(2), 021002.

Bastani, K., Barazandeh, B. and Kong, Z.J. (2018) Fault diagnosis in multistation assembly systems using spatially correlated Bayesian learning algorithm. Journal of Manufacturing Science and Engineering, 140(3), 031003.

Bastani, K., Kong, Z., Huang, W., Huo, X. and Zhou, Y. (2012) Fault diagnosis using an enhanced relevance vector machine (rvm) for partially diagnosable multistation assembly processes. IEEE Transactions on Automation Science and Engineering, 10(1), 124–136.

Bastani, K., Kong, Z., Huang, W. and Zhou, Y. (2016) Compressive sensing-based optimal sensor placement and fault diagnosis for multi-station assembly processes. IIE Transactions, 48(5), 462–474.

Candes, E.J. and Wakin, M.B. (2008) An introduction to compressive sampling. IEEE Signal Processing Magazine, 25(2), 21–30.

Ceglar, D. and Shi, J. (1996) Fixture failure diagnosis for autobody assembly using pattern recognition. Journal of Manufacturing Science and Engineering, 118(1), 55–66.

Chen, Y., Jin, J. and Shi, J. (2004) Integration of dimensional quality and locator reliability in design and evaluation of multi-station body-in-white assembly processes. IIE transactions, 36(9), 827–839.

Cohn, I., El-Hay, T., Friedman, N. and Kupferman, R. (2010) Mean field variational approximation for continuous-time Bayesian networks. The Journal of Machine Learning Research, 11, 2745–2783.

Cotter, S.F., Rao, B.D., Engan, K. and Kreutz-Delgado, K. (2005) Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Transactions on Signal Processing, 53(7), 2477–2488.

Ding, Y., Ceglar, D. and Shi, J. (2002) Fault diagnosis of multistage manufacturing processes by using state space approach. Journal of Manufacturing Science and Engineering, 124(2), 313–322.

Ding, Y., Ceglar, D. and Shi, J. (2000) Modeling and diagnosis of multistage manufacturing processes: Part i state space model, in Proceedings of the 2000 Japan/USA Symposium on Flexible Automation, ASME, Ann Arbor, MI, pp. 23–26.

Donoho, D.L. (2006) For most large underdetermined systems of linear equations the minimal ℓ1 norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 59(6), 797–829.

Fang, J., Shen, Y., Li, F., Li, H. and Chen, Z. (2015) Support knowledge-aided sparse Bayesian learning for compressed sensing, in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE Press, Piscataway, NJ, pp. 3786–3790.

Guo, Y., Sun, B., Li, N. and Fang, D. (2017) Variational Bayesian inference-based counting and localization for off-grid targets with faulty prior information in wireless sensor networks. IEEE Transactions on Communications, 66(3), 1273–1283.

Han, N. and Song, Z. (2018) Bayesian multiple measurement vector problem with spatial structured sparsity patterns. Digital Signal Processing, 75, 184–201.

Han, T., Jiang, D., Sun, Y., Wang, N. and Yang, Y. (2018) Intelligent fault diagnosis method for rotating machinery via dictionary learning and sparse representation-based classification. Measurement, 118, 181–193.

Huang, W., Lin, J., Bezdrcny, M., Kong, Z. and Ceglarek, D. (2018) Stream-of-variation modelling: Part i: A generic three-dimensional variation model for rigid-body assembly in single station assembly processes. Journal of Manufacturing Science and Engineering, 140(3), 821–831.

Huang, W., Lin, J., Kong, Z. and Ceglarek, D. (2007) Stream-of-variation (sova) modeling: Part ii: a generic 3d variation model for rigid body assembly in multistation assembly processes. Journal of Manufacturing Science and Engineering, 129(4), 832–842.

Jin, J. and Chen, Y. (2001) Quality and reliability information integration for design evaluation of fixture system reliability. Quality and Reliability Engineering International, 17(5), 355–372.

Kong, Z., Ceglarek, D. and Huang, W. (2008) Multiple fault diagnosis method in multistation assembly processes using orthogonal
diagonalization analysis. *Journal of Manufacturing Science and Engineering*, **130**(1), 011014.

Kong, Z., Huang, W. and Oztekin, A. (2009) Variation propagation analysis for multistation assembly process with consideration of gd&t factors. *Journal of Manufacturing Science and Engineering*, **131**(3), 051010.

Kong, Z., Kumar, R., Gogineni, S., Zhou, Y. and Ceglarek, D. (2006) Mode-based tolerance analysis in multi-station assembly using stream of variation model. *Trans. NAMRI/SME*, **34**, 469–476.

Lee, J., Son, J., Zhou, S. and Chen, Y. (2020) Variation source identification in manufacturing processes using bayesian approach with sparse variance components prior. *IEEE Transactions on Automation Science and Engineering*, **17**(3), 1469–1485.

Li, H., Jiang, Y., Fang, J. and Rangaswamy, M. (2017) Adaptive subspace signal detection with uncertain partial prior knowledge. *IEEE Transactions on Signal Processing*, **65**(16), 4394–4405.

Li, S. and Chen, Y. (2016) A Bayesian variable selection method for joint diagnosis of manufacturing process and sensor faults. *IIE Transactions*, **48**(4), 313–323.

Liu, J. (2010) Variation reduction for multistage manufacturing processes: A comparison survey of statistical-process-control vs stream-of-variation methodologies. *Quality and Reliability Engineering International*, **26**(7), 645–661.

Luessi, M., Babacan, S.D., Molina, R. and Katsaggelos, A.K. (2013) Bayesian simultaneous sparse approximation with smooth signals. *IEEE Transactions on Signal Processing*, **61**(22), 5716–5729.

Petersen, K.B., Winther, O. and Hansen, L.K. (2005) On the slow convergence of EM and VBEM in low-noise linear models. *Neural Computation*, **17**(9), 1921–1926.

Shi, J. (2023) In-process quality improvement: Concepts, methodologies, and applications. *IIE Transactions*, **55**(1), 2–21.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**(1), 267–288.

Wipf, D.P. and Rao, B.D. (2007) An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Transactions on Signal Processing*, **55**(7), 3704–3716.

Yu, D., Guo, Y., Li, N. and Wang, M. (2019) Sa-m-sbl: An algorithm for CSI-based device-free localization with faulty prior information. *IEEE Access*, **7**, 61831–61839.

Zhang, C., Yan, H., Lee, S. and Shi, J. (2021) Dynamic multivariate functional data modeling via sparse subspace learning. *Technometrics*, **63**(3), 370–383.

Zhang, Z., Huang, W., Liao, Y., Song, Z., Shi, J., Jiang, X., Shen, C. and Zhu, Z. (2022) Bearing fault diagnosis via generalized logarithm sparse regularization. *Mechanical Systems and Signal Processing*, **167**, 108576.

Zhang, Z. and Rao, B.D. (2011) Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE Journal of Selected Topics in Signal Processing*, **5**(3), 912–926.

Zhou, S., Chen, Y. and Shi, J. (2004) Statistical estimation and testing for variation root-cause identification of multistage manufacturing processes. *IEEE Transactions on Automation Science and Engineering*, **1**(1), 73–83.