A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics

Chen-Fu Chien, *Member, IEEE*, and Shih-Chung Chuang

***Abstract*—Root cause detecting and rapid yield ramping for advanced technology nodes are crucial to maintain competitive advantages for semiconductor manufacturing. Since the data structure is increasingly complicated in a fully automated wafer fabrication facility, it is difficult to diagnose the whole production system for fault detection. A number of approaches have been proposed for fault diagnosis and root cause detection. However, many constraints in real settings restrict the usage of conventional approaches, due to the big data with complicated data structure. In particular, a batch may not be considered as a run in the present sub-batch processing system for wafer fabrication, in which the processing paths of the wafers in a batch could be different. Motivated by realistic needs, this paper aims to develop a root cause detection framework for the sub-batch processing system. Briefly, the proposed framework consists of three phases: data preparation, data dimension reduction, and the sub-batch processing model construction and evaluation. The proposed approach has been validated by a sequence of simulations and an empirical study conducted in a leading semiconductor manufacturing company in Taiwan. The results have shown practical viability of the proposed approach. Indeed, the developed approach is incorporated in the engineering data analysis system in this case company.**

***Index Terms*—Sub-batch processing system, big data analytics, root cause detection, data mining, longitudinal data analysis.**

I. Introduction

SEMICONDUCTOR manufacturing is one of the most complicated manufacturing systems that contain hundreds of process steps, in which the data is generated when wafers move from step to step in a wafer fabrication facility (fab). Economies of speed are critical as semiconductor products in consumer electronics era compete with increasingly shortening time-to-market [1]. The average sale price of the IC with short product life is quickly reduced [2], [3]. Thus, rapid root cause detection for yield ramp up for advanced technology migration is critical for semiconductor companies to maintain competitive advantages [3], [4]. Conventionally, the engineers diagnose the system and specific module to detect the root cause based on experiences and domain knowledge. Big data with a huge number of process features, in-line metrologies, product test results, and product data will be recorded automatically when the semiconductor products are fabricated in the advanced manufacturing system [5]. Although the automatically collected big data can be analyzed to extract useful manufacturing intelligence, the data structure becomes increasingly complicated for advanced nano technology nodes. Thus, the detection of root causes is exponentially difficult that cannot rely only on domain knowledge and conventional approaches.

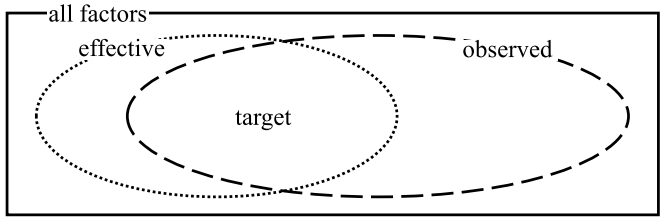

Fig. 1. Categories of the factors.

Suppose that all factors in the advanced technology nodes for wafer fabrication can be expressed by $X_p, p = 1, \ldots, A$, and Y is the interested response variable such as the yield. The factors can be categorized as illustrated in Fig. 1. The big rectangle expresses the set of all A factors; the dash ellipse expresses the set of observed factors, the dot ellipse expresses the set of the effective factors that can affect Y; the intersection of dash and dot ellipses expresses the set of the target factors. In general, it is hard to collect all effective factors since the semiconductor manufacturing is very lengthy and complicated. The objective of the analysis is to detect the effective factors in the observed factors. In this study, we use $F_p, p = 1, \ldots, A$ denote all the factors, in which $F_p, p = 1, \ldots, P$ are the observed factors and $F_p, p = P+1, \ldots, A$ are the unobserved factors.

Manuscript received October 6, 2013; revised September 1, 2014; accepted September 6, 2014. Date of publication September 9, 2014; date of current version November 10, 2014. This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant NSC100-2628-E-007-017-MY3 and Grant NSC103-2622-E-007-002, in part by the Toward World-Class Universities Project of NTHU under Grant 102N2075E1, in part by Taiwan Semiconductor Manufacturing Company under Grant 100A0259JC, and in part by Hsinchu Science Park under Grant 102A26.

The authors are with the NTHU-TSMC Center for Manufacturing Excellence, Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cfchien@mx.nthu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSM.2014.2356555

Data mining and big data analytics approaches have been developed to extract potentially useful information and manufacturing intelligence from massive data in various problems in different fields such as bioinformatics [6], marketing trend and demand forecast [1], sales forecast [7], human resource management [8], financial engineering [9], health and medical decision making [10], fault location [11], and manufacturing [12], [13]. In particular, data mining and manufacturing intelligence approaches have been implemented for semiconductor manufacturing including demand forecast [1], human resource management [8], yield enhancement [5], [14], wafer bin map analysis [15], [16], cycle time reduction [12], [13], and fault location [19]. In addition, various statistical analysis approaches were developed to model the relation between the responses and the factors for the traditional manufacturing system. However, little research has been done to address the issues involved in the sub-batch processing systems that are common for advanced technology nodes for wafer fabrication.

In advanced technology nodes, the wafers are processed by batch or sub-batch to speed up production. Indeed, the difference of quality characteristics of the wafers within a batch is averagely smaller than that between different batches. However, few studies have been done to develop a suitable model for describing this phenomenon. Most conventional approaches to address such problems are developed in some trial-and-error ways that are lack of theoretical basis. It is the kernel for statistical analyses to develop a suitable model.

To fill the gap, this study aims to develop an effective approach to detect the root causes in the batch and sub-batch processing system for dealing with semiconductor manufacturing big data in real settings. In particular, the effects of the sub-batch processing system are firstly decomposed, and thus a suitable model called sub-batch processing model (SBPM) is developed to fit the characteristics of the sub-batch processing system. Motivated by realistic needs, a framework based on the developed SBPM was proposed for root cause detection for wafer fabrication, in which random forest (RF) [20], [21] was employed for handling dimensionality and collinearity. The proposed approach was validated by a number of simulations and then an empirical study conducted in a leading semiconductor manufacturing company in Taiwan. The results have shown practical viability of the proposed approach.

The remainder of this paper is organized as follows. Section II reviews the traditional approaches for root cause detection in the batch processing system. Section III introduces the model for the sub-batch processing system. Section IV proposes a framework to detect the root causes for the sub-batch processing system. Section V validates the proposed approach by a sequence of simulations and an empirical study from a leading semiconductor manufacturing company in Taiwan. Section VI concludes with discussions of contributions and future research directions.

II. Traditional Approaches for Batch Processing Data

The batch processing system and data collection can be generally illustrated as Fig. 2. There are multiple stages for various functions, and several potential factors observed in each stage that can be denoted by $F_1, \ldots, F_P$. Suppose that there are K batches of products. Use N_k to denote the number of products contained in the kth batch, and let $N = \sum_{k=1}^{K} N_k$. The data of process record and the response are denoted by $\mathbf{X}_{kn} = (X_{1kn}, \ldots, X_{Pkn})$ and Y_{kn}, $k = 1, \ldots, K$, $n = 1, \ldots, N_k$, where $X_{pkn}, p = 1, \ldots, P$ and Y_{kn} denote the measured value of $F_p, p = 1, \ldots, P$ and Y of the nth product of batch k, respectively. Without loss of the generality, the factors can be nominal or numerical. A nominal factor is used to denote the processed tool, chamber, and recipe for wafer fabrication. A numerical factor is used to record the variables such as temperature, pressure, critical dimension and so on. Conventional approaches to derive the relationship between $\mathbf{X}_{kn}$ and Y_{kn} are briefly reviewed below.

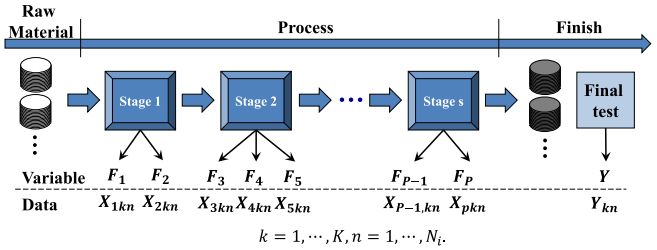

Fig. 2. Batch processing system and data collection.

A. Analysis by Ignoring Batch Dependency

Approach by ignoring batch dependency is commonly used to analyze the data from the batch processing system owing to the lack of suitable model. Since the number of potential factors P is large, it is hard to construct a regression model with all the factors, though not all factors are observed. Researchers have proposed a number of feature selection approaches such as stepwise regression [22], least angle regression (LARS) regression [23] and least absolute shrinkage and selection operator (Lasso) regression [24]. In addition, some methodologies for single factor analysis such as one-way analysis of variance (ANOVA) and Kruskal-Wallis (K-W) test [25] can be used to detect the effect of some specific factor. Let $\mathbf{Q} = \{q_1, \ldots, q_Q\} \subset \{1, \ldots, P\}$ be the set of indices of the modeled factors, and denote $\mathbf{X}_{kn}^{\mathbf{Q}} = (X_{q_1kn}, \ldots, X_{q_Qkn})$. Then, the model

$$Y_{kn} = g\left(\mathbf{X}_{kn}^{\mathbf{Q}}\right) + \varepsilon_{kn}, \quad k = 1, \ldots, K, \quad n = 1, \ldots, N_k \quad (1)$$

can be used to evaluate the effects of the factors, where g is a function to fit the relation of Y_{kn} and $\mathbf{X}_{kn}^{\mathbf{Q}}$. However, the assumption of the independency in error term e_{kn} is not appropriate in the present problem, in which ε_{kn}s within the same batch are dependent that will lead to a poor model fitting and failure to identify root causes.

B. Analysis by Batch Level Data

Approach by analyzing batch level data evaluates the factor effect based on batch level data. Let $\overline{Y}_k$ be the mean of $Y_{kn}, n = 1, \ldots, N_k$ and $\overline{\mathbf{X}}_k^{\mathbf{Q}} = (\overline{X}_{q_1k}, \ldots, \overline{X}_{q_Qk})$, while $\overline{X}_{pk}$

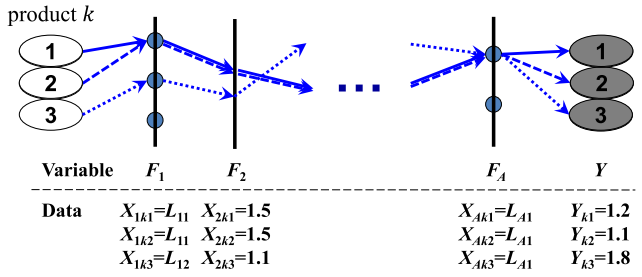

Fig. 3. Stream of process. The solid lines and dashed lines are actually overlapped; the little difference of them is in order to facilitate read.

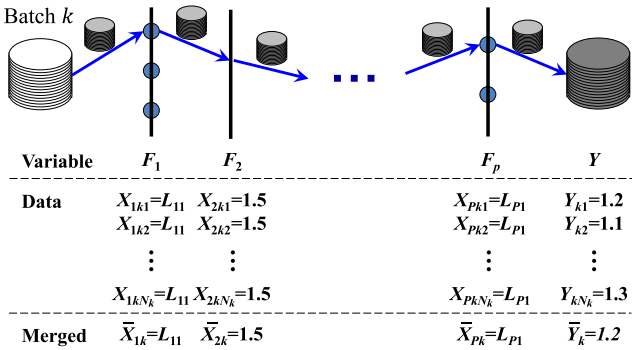

Fig. 4. Batch processing system. The process records of the products in batch 1 are all the same. Thus, it can be denoted by $\bar{\mathbf{X}}_k=(\bar{X}_{1k},\ldots,\bar{X}_{Pk})$.

denotes the mean of $X_{pkn}, n = 1,\ldots,N_k$. Then, the model becomes

$$\bar{Y}_k = g\left(\bar{\mathbf{X}}_k^{\mathbf{Q}}\right) + \varepsilon_k,\ \ k = 1,\ldots,K. \tag{2}$$

This approach merges the observations of one batch into a single value to solve the problem of error dependency, in which $\varepsilon_k,\ k = 1,\ldots,K$ are assumed to be mutually independent. However, the use of this approach will lose the within-batch information. Also, it is hard to be implemented when $X_{pkn}, n = 1,\ldots,N_k$ cannot be considered via the average. For example, if F_p is a nominal factor and $X_{pkn}, n = 1,\ldots,N_k$ are not of the same level, it will be difficult to merge the observations. Although this approach can roughly solve the problem by fitting the mode of the batches, it may cause the bias in the result subject to the data quality.

C. Analysis by Taking Batch as a Factor

Approach by taking batch as a factor treats the batch information as a dummy variable to deal with the data from the batch processing system to eliminate the common effect within one batch to solve the problem of error dependency. Comparing to the analysis by batch level data, this approach will lose the between-batch information since the dummy variable will cause a serious alias to some nominal factors. Thus, this approach is not appropriate to deal with the problem in the batch processing system.

III. Fundamental

A. Model for the Sub-Batch Processing System

1) Model for Traditional Batch Processing System: As the batch processing system shown in Fig. 2, suppose that all factors can be denoted by $F_p, p = 1,\ldots,A$, in which $F_p, p = 1,\ldots,P$ are observed and the others are not. Let $\mathbf{X}_{kn}^A = (X_{1kn},\ldots,X_{Akn})$ denote the vector of the measures of the factors of product n of batch k. Note that $X_{pkn}, p = P+1,\ldots,A$ cannot be observed, though they exist. Without focusing on any factors as (1), the relation between the factors and the response can be written as

$$Y_{kn} = g\left(\mathbf{X}_{kn}^A\right) + \varepsilon_{kn},\ \ k = 1,\ldots,K,\ \ n = 1,\ldots,N_k, \tag{3}$$

where g is a function to describe the variation of Y_{kn} by $\mathbf{X}_{kn}^A$, and $\varepsilon_{kn}, k = 1,\ldots,K,\ \ n = 1,\ldots,N_k$ denotes the pure noise assumed to be independent and identically distributed (*i.i.d.*) random variables. In (3), $g(\mathbf{X}_{kn}^A)$ is called stream effect that can be decomposed as follows:

A stream of product process indicates a sequence of setting of $F_1,\ldots,F_A$ as illustrated as Fig. 3, in which the stream of product 1 is the same as product 2 and is different from product 3. Thus, the effect to the response caused by the stream is so-called stream effect as $g(\mathbf{X}_{kn}^A)$ in (3). In order to simplify the statistical analysis, g is considered as a linear function, and then (3) can be rewritten as

$$Y_{kn} = \beta_0 + \sum_{p=1}^{A}\beta_p X_{pkn} + \varepsilon_{kn}, k = 1,\ldots,K,\ \ n = 1,\ldots,N_k, \tag{4}$$

where β_0 is intercept and $\beta_p, p = 1,\ldots,A$ are the associated coefficients.

The model of (4) can be evaluated by statistical analysis if the number of factors A is small and all the factors are observed ($A = P$). However, this approach will become infeasible to deal with the present problem in the advanced technology nodes. Thus, only few interested and observed factors are evaluated and the rest are treated as noises. Recall $\mathbf{Q}$ defined in the Section II to be the set of indices of the interested factors. Then, (4) can be further rewritten as

$$Y_{kn} = \beta_0 + \sum_{p\in\mathbf{Q}}\beta_p X_{pkn} + \xi_{kn} + \varepsilon_{kn},\quad k = 1,\ldots,K,\ \ n = 1,\ldots,N_k, \tag{5}$$

where $\mathbf{Q}^c = \{1,\ldots,A\}\backslash\mathbf{Q}$ and $\xi_{kn} = \sum_{p\in\mathbf{Q}^c}\beta_p X_{pkn}$. $\sum_{p\in\mathbf{Q}}\beta_p X_{pkn}$ is called the factor effect and ξ_{kn} is called the remained stream effect. $X_{pkn}, p \in \mathbf{Q}$ denote the random variables for describing the involved uncertainties, and hence ξ_{kn}s are treated as the random variables.

A conventional batch processing system can be illustrated as Fig. 4. The process record of the products in the same batch are the same, and thus the data for analysis can be rearranged as $\bar{\mathbf{X}}_k^A = (\bar{X}_{1k},\ldots,\bar{X}_{Ak})$ and $\bar{Y}_k,\ \ k = 1,\ldots,K$, where $\bar{X}_{pk}$ denotes the process record of F_p of the products of batch k and $\bar{Y}_k$ denotes the average response of the products in batch k. Thus, Equation (5) can be rewritten as

$$\bar{Y}_k = \beta_0 + \sum_{p\in\mathbf{Q}}\beta_p\bar{X}_{pk} + \xi_k + \bar{\varepsilon}_k, k = 1,\ldots,K, \tag{6}$$

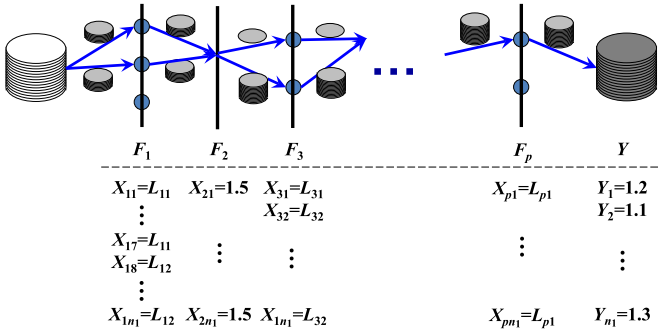

Fig. 5. Sub-batch process of system.

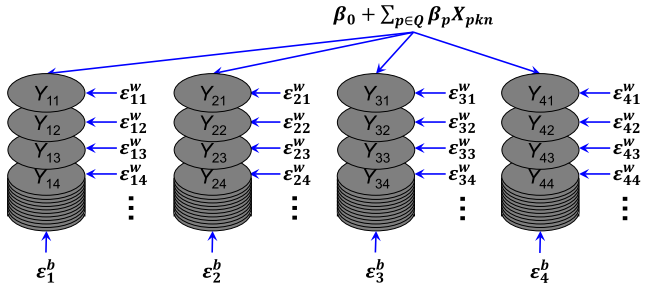

Fig. 6. Decomposition of effect in the sub-batch processing system.

where $\xi_k = \xi_{k1} = \cdots = \xi_{kN_k}$ and $\overline{\varepsilon}_k = N_k^{-1} \sum_{n=1}^{N_k} \varepsilon_{kn}$. In this model, ξ_ks and $\overline{\varepsilon}_k$s can be treated as independent random variables, respectively. Let $\zeta_k = \xi_k + \overline{\varepsilon}_k$, and then ζ_ks can be treated as independent random variables. Therefore, Equation (6) can be further rewritten as

$$\overline{Y}_k = \beta_0 + \sum_{p \in \mathbf{Q}} \beta_p \overline{X}_{pk} + \zeta_k, k = 1, \ldots, K. \tag{7}$$

The effect of the factors can be evaluated by $\overline{\mathbf{X}}_k^{\mathbf{Q}}$ and $\overline{Y}_k$ as the approach introduced in Section II-B. Note that ζ_ks cannot be assumed to be identically distributed if the numbers of the products in batches are not uniform. However, this approach cannot deal with the sub-batch data in present problem.

2) Generalized Model for the Sub-Batch Processing System: As Fig. 5 illustrates the sub-batch processing system in semiconductor manufacturing, the product batch may be divided into serveral parts to be processed and then merged in the later stage. Note that the form of partition may not be fixed and thus the streams of the products in the same batch will be different. For analyzing the present problem, the model assumption should be reconsidered as follows.

Recall the model shown in (5). For any batch k, ξ_{kn}, $n = 1, \ldots, N_k$ are different and mutually dependent since the processes of the products in one batch are the same in some stages and different in some stages. Therefore, $\varepsilon_{kn}^{'} = \xi_{kn} + \varepsilon_{kn}$, $k = 1, \ldots, K$, $n = 1, \ldots, N_k$ cannot be treated as independent errors and then executing the traditional statistical inference shown in Section II.

Although ξ_{kn}s are not mutually independent, some assumptions of independency and correlation can be claimed. Since the processes of the products of one batch are independent of another batch, it can be rationally assumed that the between-batch remained stream effects are independent; that is, $\xi_{k_1 n_1}$ is independent of $\xi_{k_2 n_2}$ if $k_1 \neq k_2$. For any two products in the same batch, the joint distribution of their remained stream effects can be considered as a two-dimensional distribution that is independent of the index of the batch and the indices of the products. Hence, it can be rationally assumed that the correlation coefficients of the remained stream effects within the batch are fixed; that is, the correlation coefficient of ξ_{kn_1} and ξ_{kn_2} is equal to a fixed constant c for any k and $n_1 \neq n_2$. Also, the ξ_{kn}s can be assumed identically distributed since there is no essential difference between the processes of all products. Let $\zeta_{kn} = \xi_{kn} + \varepsilon_{kn}$. Then, $\zeta_{k_1 n_1}$ is independent of $\zeta_{k_2 n_2}$ if $k_1 \neq k_2$, and the correlation coefficient of ζ_{kn_1} and ζ_{kn_2} is equal to a fixed constant ρ^2 for any k and $n_1 \neq n_2$. Thus, the ζ_{kn}s can be assumed identically distributed, where $\rho^2 = c \times Var(\xi_{11}) / (Var(\xi_{11}) + Var(\varepsilon_{11}))$. Equation (5) can be further rewritten as

$$Y_{kn} = \beta_0 + \sum_{p \in \mathbf{Q}} \beta_p X_{pkn} + \zeta_{kn}, \quad k = 1, \ldots, K, \ n = 1, \ldots, N_k, \tag{8}$$

where $\zeta_{11}, \ldots, \zeta_{1N_1}, \zeta_{21}, \ldots, \zeta_{2N_2}, \ldots, \zeta_{KN_K}$ have the identically marginal distribution with mean 0 and variance σ_ζ^2. Let $\zeta_k = (\zeta_{k1}, \ldots, \zeta_{kN_k}), k = 1, \ldots, K$. Then, $\zeta_k, k = 1, \ldots, K$ are mutually independent and have correlation matrix $\mathbf{V}_{N_k}$, $k = 1, \ldots, K$, respectively, where $\mathbf{V}_{N_k}$ is a matrix of $N_k \times N_k$ with diagonal elements 1 and other elements ρ. Thus, ζ_{kn}s can be treated as a sequence of dependent noises.

Furthermore, ζ_{kn} can be decomposed as $\zeta_{kn} = \varepsilon_k^b + \varepsilon_{kn}^w$, where ε_k^b denotes the center of the remained stream effect of the products of batch k that can be treated as between-batch noise, and ε_{kn}^w denotes the sum of ε_{kn}. The rest of the remained stream effect can be treated as within-batch noise. Furthermore, $\varepsilon_k^b, k = 1, \ldots, K$ and $\varepsilon_{kn}^w, k = 1, \ldots, K$, $n = 1, \ldots, N_k$ are assumed to be mutually independent. Hence, Equation (7) can be rewritten as

$$Y_{kn} = \beta_0 + \sum_{p \in \mathbf{Q}} \beta_p X_{pkn} + \varepsilon_k^b + \varepsilon_{kn}^w, \quad k = 1, \ldots, K, n = 1, \ldots, N_k, \tag{9}$$

where the variance of ε_k^b and ε_{kn}^w are $\sigma_b^2 = \rho^2 \sigma_\zeta^2$ and $\sigma_w^2 = \sigma_\zeta^2 - \sigma_b^2$, respectively. The model as (9) is SBPM, in which the related effect decomposition can be illustrated as Fig. 6.

B. Inferences of SBPM

1) Maximum Likelihood Estimators: The maximum likelihood estimator (MLE) of SBPM can be derived as follows, while more related studies can be found in longitudinal data analysis [26]–[29]. Let $\mathbf{X}_{kn}^{\mathbf{Q}} = (1, X_{q_1 kn}, \ldots, X_{q_Q kn})$ and Y_{kn} be the factors and the response variable of the nth product in the kth batch, where the first element of $\mathbf{X}_{kn}^{\mathbf{Q}}$ 1 is used to describe the intercept. Let $\mathbf{X}_k^{\mathbf{Q}} = (\mathbf{X}_{k1}^{\mathbf{Q}'}, \ldots, \mathbf{X}_{kN_k}^{\mathbf{Q}'})'$, $\mathbf{X}^{\mathbf{Q}} = (\mathbf{X}_1^{\mathbf{Q}'}, \ldots, \mathbf{X}_K^{\mathbf{Q}'})'$, $\mathbf{Y}_k = (Y_{k1}, \ldots, Y_{kN_k})'$ and $\mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_K')'$, and use the lowercases $\mathbf{x}_{kn}^{\mathbf{Q}}$, y_{kn}, $\mathbf{x}_k^{\mathbf{Q}}$, $\mathbf{x}^{\mathbf{Q}}$,

$\mathbf{y}_k$ and $\mathbf{y}$ to denote their observations, in which the first column of $\mathbf{X}_{kn}^{\mathbf{Q}}$ "1" is used to represent the intercept of model. Recall the model (9) and assume that ε_k^b and ε_{kn}^w are from normal distributions of mean 0 with variance σ_b^2 and σ_w^2 respectively. Let $\boldsymbol{\beta} = (\beta_0, \beta_{q_1}, \ldots, \beta_{q_Q})'$. Under the assumption of normal distribution of the noise, the likelihood function of $(\boldsymbol{\beta},\ \sigma_\zeta^2, \rho^2)$ can be written as

$$L\left(\boldsymbol{\beta}, \sigma_\zeta^2, \rho^2 \mid \mathbf{x}^{\mathbf{Q}}, \mathbf{y}\right) = \prod_{k=1}^{K} \frac{1}{(2\pi)^{\frac{N_k}{2}} \left|\Sigma_{N_k}\right|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}\left(\mathbf{y}_k - \mathbf{x}_k^{\mathbf{Q}}\boldsymbol{\beta}\right)' \Sigma_{N_k}^{-1}\left(\mathbf{y}_k - \mathbf{x}_k^{\mathbf{Q}}\boldsymbol{\beta}\right)\right\}, \quad (10)$$

where $\Sigma_{N_k}=\sigma_\zeta^2\mathbf{V}_{N_k}$. Equation (10) can be directly obtained from (8). Then, the related log-likelihood function of $(\boldsymbol{\beta}, \sigma_\zeta^2, \rho^2)$ becomes

$$\begin{aligned} l\left(\boldsymbol{\beta}, \sigma_\zeta^2, \rho^2 \mid \mathbf{y}, \mathbf{x}^{\mathbf{Q}}\right) &= \ln L\left(\boldsymbol{\beta}, \sigma_{all}^2, \rho^2 \mid \mathbf{x}^{\mathbf{Q}}, \mathbf{y},\right) \\ &= -\frac{1}{2}\left[\sum_{k=1}^{K} N_k \ln 2\pi + \sum_{k=1}^{K} N_k \ln\sigma_\zeta^2 + \sum_{k=1}^{K} ln\left|\mathbf{V}_{N_k}\right| + \sigma_\zeta^{-2}\sum_{k=1}^{K}\left(\mathbf{y}_k - \mathbf{x}^{\mathbf{Q}}\boldsymbol{\beta}\right)' \mathbf{V}_{N_k}^{-1}\left(\mathbf{y}_k - \mathbf{x}^{\mathbf{Q}}\boldsymbol{\beta}\right)\right] \end{aligned} \quad (11)$$

where ln denotes the natural logarithm function. Hence,

$$\begin{aligned} \frac{\partial}{\partial\boldsymbol{\beta}} l\left(\boldsymbol{\beta}, \sigma_\zeta^2, \rho^2 \mid \mathbf{x}^{\mathbf{Q}}, \mathbf{y}\right) &= -\frac{\sigma_\zeta^{-2}}{2}\frac{\partial}{\partial\boldsymbol{\beta}}\left[\sum_{k=1}^{K} 2\boldsymbol{\beta}' \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\mathbf{y}_k + \boldsymbol{\beta}' \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\mathbf{x}_k^{\mathbf{Q}}\boldsymbol{\beta}\ \right] \\ &= \sigma_\zeta^{-2}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\left(\mathbf{y}_k - \mathbf{x}_k^{\mathbf{Q}}\boldsymbol{\beta}\right). \end{aligned}$$

$$\left(\because \mathbf{V}_{N_k}^{-1} \text{ is symmetric, } \left(\mathbf{y}_k' \mathbf{V}_{N_k}^{-1}\mathbf{x}^{\mathbf{Q}}\boldsymbol{\beta}\right)' = \boldsymbol{\beta}' \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\mathbf{y}_k\right) \quad (12)$$

Let (12) equal to zero, and thus

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\mathbf{x}_k^{\mathbf{Q}}\right)^{-1}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\mathbf{y}_k, \quad (13)$$

The associated Hessian matrix can be shown as follows:

$$\begin{aligned} H(\boldsymbol{\beta}) &= \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} l\left(\boldsymbol{\beta}, \sigma_\zeta^2, \rho^2 \mid \mathbf{x}^{\mathbf{Q}}, \mathbf{y}\right) \\ &= -\sigma_\zeta^{-2}\frac{\partial}{\partial\boldsymbol{\beta}'}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\mathbf{x}_k^{\mathbf{Q}}\boldsymbol{\beta} \\ &= -\sigma_\zeta^{-2}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'}\mathbf{V}_{N_k}^{-1}\mathbf{x}_i^{\mathbf{Q}} < 0. \end{aligned} \quad (14)$$

The last inequality is validated, since $\mathbf{V}_{N_k}$ is positive-definite matrix. Thus, the MLE of $\boldsymbol{\beta}$ follows and depends on only ρ^2. Next, we take the partial derivative of l with respect to σ_ζ^2 for obtaining its MLE as follows:

$$\begin{aligned} \frac{\partial}{\partial\sigma_\zeta^2} l\left(\boldsymbol{\beta}, \sigma_\zeta^2, \rho^2 \mid \mathbf{x}^{\mathbf{Q}}, \mathbf{y}\right) &= -\frac{\sigma_\zeta^{-2}}{2}\sum_{k=1}^{K} N_k + \frac{\sigma_\zeta^{-4}}{2}\sum_{k=1}^{K}\left(\mathbf{y}_k - \mathbf{x}_k^{\mathbf{Q}}\boldsymbol{\beta}\right)' \mathbf{V}_{N_k}^{-1}\left(\mathbf{y}_k - \mathbf{x}_k^{\mathbf{Q}}\ \boldsymbol{\beta}\right). \end{aligned} \quad (15)$$

Let it equal to zero, and then we have

$$\widehat{\sigma}_\zeta^2 = \left(\sum_{k=1}^{K} N_k\right)^{-1}\sum_{k=1}^{K}\left(\mathbf{y}_k - \mathbf{x}_k^{\mathbf{Q}}\widehat{\boldsymbol{\beta}}\right)' \mathbf{V}_{N_k}^{-1}\left(\mathbf{y}_k - \mathbf{x}_k^{\mathbf{Q}}\widehat{\boldsymbol{\beta}}\right) \quad (16)$$

Also, the associated Hessian matrix is shown as follows:

$$\frac{\partial^2}{\partial\left(\sigma_\zeta^2\right)^2} l\left(\boldsymbol{\beta}, \sigma_{all}^2, \rho^2 \mid \mathbf{x}^{\mathbf{Q}}, \mathbf{y}\right) \overset{\sigma_\zeta^2=\widehat{\sigma}_\zeta^2}{\Longrightarrow} -\frac{1}{2}\sigma_\zeta^{-4}\sum_{k=1}^{K} N_k < 0. \quad (17)$$

Therefore, the MLE of σ_ζ^2 becomes (16) that depends on $\boldsymbol{\beta}$ and ρ^2. Thus, both MLEs of $\boldsymbol{\beta}$ and σ_ζ^2 can be written as the functions of ρ^2. That is, if the MLE of ρ^2 can be found, then those of $\boldsymbol{\beta}$ and σ_ζ^2 follow.

To derive the MLE of ρ^2, the log-likelihood function of (11) can be derived by replacing $\boldsymbol{\beta}$ and σ_ζ^2 by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}_\zeta^2$, and then the log-likelihood function of ρ^2 can be obtained as

$$\begin{aligned} l\left(\rho^2 \mid \mathbf{y}, \mathbf{x}^{\mathbf{Q}}\right) &= l\left(\widehat{\boldsymbol{\beta}}\left(\rho^2\right), \widehat{\sigma}_\zeta^2\left(\rho^2\right), \rho^2 \mid \mathbf{x}^{\mathbf{Q}}, \mathbf{y}\right) \\ &= -\frac{1}{2}\left[\sum_{k=1}^{K} N_k \ln 2\pi + \sum_{k=1}^{K} N_k \ln\widehat{\sigma}_\zeta^2\left(\rho^2\right) + \sum_{k=1}^{K} \ln\left|\mathbf{V}_{N_k}\right| + \sum_{k=1}^{K} N_k\right] \end{aligned} \quad (18)$$

Hence, the desired solution of MLE of ρ^2 can be expressed as follows:

$$\widehat{\rho}^2 = \underset{\rho^2\in[0,1)}{\operatorname{argmin}}\left\{\sum_{k=1}^{K} N_k \ln\widehat{\sigma}_\zeta\left(\rho^2\right) + \sum_{k=1}^{K} \ln\left|\mathbf{V}_{N_k}\right|\right\}. \quad (19)$$

Indeed, the closed form of $\widehat{\rho}^2$ does not exist. However, since searching the solution of $\widehat{\rho}^2$ is a one-dimensional optimization problem with bounded domain, it can be obtained by some numeric approach [30], [31]. Furthermore, for more complex variance-covariance setting of Σ_{N_k}, generalized estimating equations (GEE) can be employed [27], [29], [32].

2) Test of Factor Effects With Large Sample: Recall the MLE of $\boldsymbol{\beta}$ in (13). Firstly, we derive the distribution of $\widehat{\boldsymbol{\beta}}$.

Let $\mathbf{S}^{-1} = \sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'} \Sigma_{N_k}^{-1} \mathbf{x}_k^{\mathbf{Q}}$. Then,

$$E\left[\widehat{\boldsymbol{\beta}}\right] = E\left[\left(\sum_{k=1}^{K} \mathbf{x}_k^{'} \mathbf{V}_{N_k}^{-1} \mathbf{x}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'} \mathbf{V}_{N_k}^{-1} (\mathbf{x}_k^{\mathbf{Q}} \boldsymbol{\beta} + \zeta_k)\right]$$
$$= E\left[\mathbf{S}\mathbf{S}^{-1}\boldsymbol{\beta} + \mathbf{S}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'} \mathbf{V}_{N_k}^{-1} \zeta_k\right]$$
$$= E\left[\boldsymbol{\beta}\right] + \mathbf{S}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'} \mathbf{V}_{N_k}^{-1} E[\zeta_k] = \boldsymbol{\beta}, \tag{20}$$

and

$$Var\left(\widehat{\boldsymbol{\beta}}\right) = E\left[\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right]$$
$$= E\left[\mathbf{S}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'} \Sigma_{N_k}^{-1} \zeta_k \zeta_k' \Sigma_{N_k}^{-1} \mathbf{x}_k^{\mathbf{Q}}\, \mathbf{S}\right]$$
$$= \mathbf{S}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'} \Sigma_{N_k}^{-1} E\zeta_k \zeta_k' \Sigma_{N_k}^{-1} \mathbf{x}_k^{\mathbf{Q}} \mathbf{S}$$
$$= \mathbf{S}\sum_{k=1}^{K} \mathbf{x}_k^{\mathbf{Q}'} \Sigma_{N_k}^{-1} \Sigma_{N_k} \Sigma_{N_k}^{-1} \mathbf{x}_k^{\mathbf{Q}} \mathbf{S} = \mathbf{S}, \tag{21}$$

where $\zeta_k = \left(\zeta_{k1}, \ldots, \zeta_{kN_k}\right)'$. Thus, $\widehat{\boldsymbol{\beta}} \sim N_{Q+1}(\boldsymbol{\beta}, \mathbf{S})$, where N_{Q+1} denotes the $(Q+1)$-variate normal distribution. When the sample size is large, the distribution of $\widehat{\boldsymbol{\beta}}$ is approximated to $N_{Q+1}\left(\boldsymbol{\beta}, \widehat{\mathbf{S}}\right)$, where $\widehat{\mathbf{S}}$ can be derived by plugging $\widehat{\sigma}_\zeta^2$ and $\widehat{\rho}^2$ into the $\mathbf{S}$. Let $\mathbf{U}$ and $\mathbf{u}$ are arbitrary matrix of $r \times (Q+1)$ and $r \times 1$, respectively. Then, the hypothesis test is: H_0:$\mathbf{U}\boldsymbol{\beta} = \mathbf{u}$ and the alternative test H_1:$\mathbf{U}\boldsymbol{\beta} \neq \mathbf{u}$. Because $\mathbf{U}\widehat{\boldsymbol{\beta}} - \mathbf{u} \sim N_r\left(\mathbf{0}_r, \mathbf{U}\mathbf{S}\mathbf{U}'\right)$ under H_0, the distribution of

$$T = \left(\mathbf{U}\widehat{\boldsymbol{\beta}} - \mathbf{u}\right)' \left(\mathbf{U}\widehat{\mathbf{S}}\mathbf{U}'\right)^{-1} \left(\mathbf{U}\widehat{\boldsymbol{\beta}} - \mathbf{u}\right) \tag{22}$$

is approximated to χ_r^2 distribution. Thus, the testing rule can be performed by

$$\Phi(\mathbf{X}, \mathbf{Y}) = \begin{cases} 1, & T > \chi_{r,\alpha}^2, \\ 0, & \text{otherwise}, \end{cases} \tag{23}$$

where $\chi_{r,\alpha}^2$ denotes the $1-\alpha$ quantile of χ^2distribution with r degree of freedom.

C. Factor Evaluation Tool

As the number of the factors is too large to conduct the analysis in the available time, dimension reduction can be firstly employed. A number of methodologies for dimension reduction such as principal component analysis [33] and exploratory factor analysis [34] have been developed. However, many of them are not suitable for root cause detection in semiconductor manufacturing, since the transformation, combination, or extracted common factors of the factors may not have physical meaning in real settings. Thus, for dimension reduction of semiconductor manufacturing data, it is better to have a factor evaluation tool and then extract important factors to conduct further analysis.

Random forest (RF) is appropriate to evaluate the factors since it can make a good trade-off between the explanation ability and the singularity of the factors based on the measurement of increasing mean square of predicted error. In particular, RF is an ensemble classifier that can handle the multivariate problem with dimensionality and collinearity by aggregating a number of decisions or predictions of weak learners [20], [21]. The classification and regression tree (CART) can be employed as the weaker learner. Furthermore, it can simultaneously deal with different nominal and numerical factors and provide the importance for each factor. Thus, this study employed RF to filter the unimportant factors due to the collinearity, via the following steps:

Step 1. Pick about one-third of observations randomly from the original dataset for validation, while the remainder is treated as the training dataset.

Step 2. Sample v bootstrapping sub-dataset from the training dataset and construct a decision tree for each sub-dataset respectively. Then we have v decision trees denoted by $D_i, i = 1, \ldots, v$.

Step 3. Construct the model by aggregating $D_i, i = 1, \ldots, v$.

Step 4. Validate the model by the reserved data in Step 1. If the accuracy is not stable enough, enlarge v and re-execute step 2, 3 and 4.

Note that the decision trees constructed in Step 2 do not need pruning, and only partial factors are randomly selected to construct the tree at each node. In general, the number of the factors to grow the tree is suggested as $\log_2 P + 1$ or $\sqrt{P}$, where P is the number of considered factors [21].

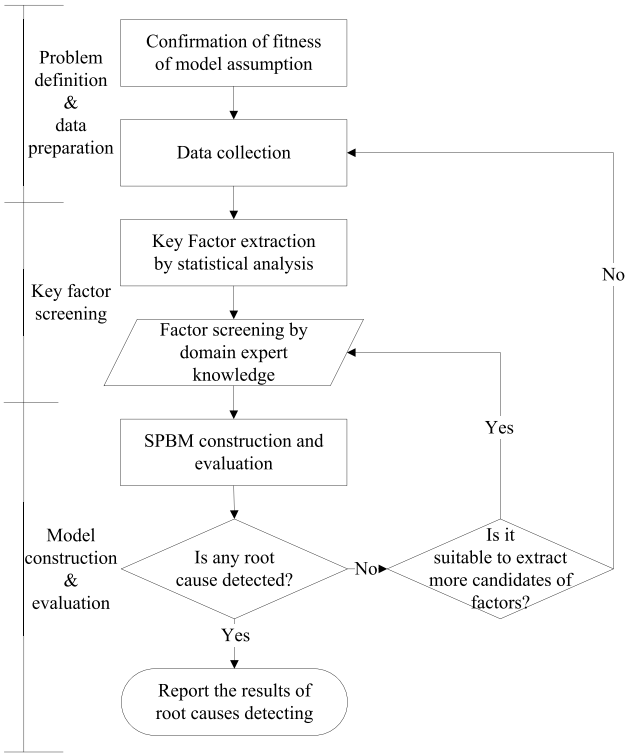

Fig. 7. SBPM-based root cause detection framework.

IV. Framework for Root Cause Detection

Fig. 7 illustrates the proposed SBPM-based framework for root cause detection including three steps: (a) problem

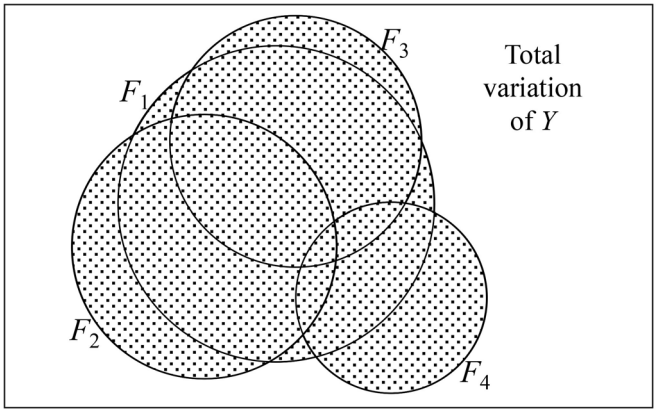

Fig. 8. Example for illustrating the importance of factors.

definition and data preparation, (b) key factor screening, and (c) model construction and evaluation.

A. Problem Definition and Data Preparation

Firstly, the proposed framework starts with problem definition and data preparation to check the characteristics of the processing system. The proposed SBPM is effective to address the processing system characterized by Fig. 5. Note that conventional batch processing system and non-batch processing system can be treated as special cases of the sub-batch processing system, and thus the proposed SBPM-based framework can be employed to deal with the systems.

Big data including process history, in-line metrology and various off-line testing data are automatically collected during wafer fabrication. Before further analysis, data preparation is employed.

B. Key Factor Screening

Secondly, key factor screening is conducted to select a number of factors that can cover major information of the response variable, Y. Since the process steps are highly interdependent, it is hard to divide the information of Y into the parts associated to specific factors. Indeed, the importance of factors to Y should be measured based on two aspects: (a) its explanation ability to Y and (b) its singularity among all factors. For example, suppose that there are 4 factors denoted by $F_1, \ldots, F_4$. The big rectangle in Fig. 8 denotes the total information of Y, and 4 circles express the parts of information of Y that can be explained by $F_1, \ldots, F_4$, in which the volumes of circles corresponds to the explanation abilities of factors to Y. The overlaps can be treated as the collinearity of factors. Thus, the explanation abilities of factors are graded according to the volumes of the circles, and the singularity of factors are graded according to the volumes of non-overlapping area of the circles. The singularity of factor can be treated as an indicator to measure the extra information provision when other factors are considered. Hence, in terms of explanation ability, the factors should be ranked as $F_1 > F_2 > F_3 > F_4$, and in terms of singularity, the factors should be ranked as $F_4 > F_2 > F_3 > F_1$. However, the trade-off between the explanation ability and singularity is a difficult decision in the sub-batch processing systems.

Thus, the proposed framework integrated RF and one-dimensional SBPM for factor screening. Suppose that there are P collected factors in the database. P indicators of factor importance, denoted by $V_{Rp}, p = 1, \ldots, P$, can be derived from RF algorithm. Note that V_{Rp} is a larger-the-better measure. The usage of one-dimensional SBPM is described as follows. For the factor F_p, the corresponding one-dimensional SBPM is

$$Y_{kn} = \beta_0 + \beta_p x_{pkn} + \varepsilon_k^b + \varepsilon_{kn}^w, \quad k = 1, \ldots, K, n = 1, \ldots, N_k. \tag{24}$$

The p-value of β_p can be calculated by $1-F(T)$, where T is shown as (22) with $\mathbf{U} = (0, 1)$ and $\mathbf{u} = 0$, and F denotes the cumulative distribution function of χ_1^2 distribution. Hence P p-values, denoted by $V_{Bp}, p = 1, \ldots, P$, can be calculated. Note that V_{Bp} is a smaller-the-better measure.

For the trade-off between the indicators from RF and one-dimensional SBPM, the weighted rank-sum function

$$\lambda_p = \omega \eta_R + (1 - \omega)\, \eta_B, p = 1, \ldots, P \tag{25}$$

is used, where $\eta_{Rp} = \sum_{i=1}^{P} I\left(V_{Rp} \geq V_{Ri}\right)$ is the rank of V_{Rp} among $V_{Rp}, p = 1, \ldots, P$, $\eta_{Bp} = \sum_{i=1}^{P} I\left(V_{Bp} \leq V_{Bi}\right)$ is the rank of V_{Bp} among $V_{Bp}, p = 1, \ldots, P$, I is the indicator function defined as

$$I(E) = \begin{cases} 1, & \text{if } E \text{ is true,} \\ 0, & \text{otherwise,} \end{cases} \tag{26}$$

and $\omega \in (0, 1)$ is a customized parameter to set the weights between SBPM and RF. The larger λ_p means the F_p is more important. A threshold is set to determine the factors to be selected as the candidates for model construction.

C. Model Construction and Evaluation

Thirdly, the stepwise regression algorithm [31], [35], [36] with Akaika information criterion (AIC) [37] is employed to construct the proposed algorithm for SBPM with variable selection.

Let $\mathbf{P}'$ be the set of indices of the factors selected from factor screening and use $|\mathbf{P}'|$ to denote the number of elements contained in $\mathbf{P}'$. Suppose that $\mathcal{P}$ is an arbitrary subset of $\mathbf{P}'$, and $F^{\mathcal{P}}$ denotes the set of the factors whose indices are contained in $\mathcal{P}$. Then, its associated data for analysis is $\mathbf{x}^{\mathcal{P}}$ that is a matrix of $N \times (|\mathcal{P}|+1)$, as the definition in Section III. Let $\boldsymbol{\beta}(\mathcal{P})$, $\sigma_\zeta^2(\mathcal{P})$, $\rho^2(\mathcal{P})$ denote the associated parameters and $l^{\mathcal{P}}(\boldsymbol{\beta}(\mathcal{P}),\ \sigma_\zeta^2(\mathcal{P}),\ \rho^2(\mathcal{P})|\ \mathbf{x}^{\mathcal{P}}, \mathbf{y})$ denote the associated log-likelihood function of the SBPM constructed by $F^{\mathcal{P}}$. Use $\widehat{\boldsymbol{\beta}}(\mathcal{P})$, $\widehat{\sigma}_\zeta^2(\mathcal{P})$, $\widehat{\rho}^2(\mathcal{P})$ to denote the MLEs of $\boldsymbol{\beta}(\mathcal{P})$, $\sigma_\zeta^2(\mathcal{P})$, $\rho^2(\mathcal{P})$, and $\hat{l}^{\mathcal{P}} = l(\widehat{\boldsymbol{\beta}}(\mathcal{P}),\ \widehat{\sigma}^2(\mathcal{P}),\ \widehat{\rho}^2(\mathcal{P})|\ \mathbf{y}, \mathbf{x}(\mathcal{P}))$ to denote the optimized $l^{\mathcal{P}}(\boldsymbol{\beta}(\mathcal{P}),\ \sigma_\zeta^2(\mathcal{P}),\ \rho^2(\mathcal{P})|\ \mathbf{x}^{\mathcal{P}}, \mathbf{y})$. Hence, the AIC of SBPM constructed by $F^{\mathcal{P}}$ is defined as

$$\mathrm{AIC}_{\mathrm{SPBM}}^{\mathcal{P}} = -2\hat{l}^{\mathcal{P}} + 2\left(1 + \sum_{p \in \mathcal{P}} d_p\right), \tag{27}$$

where d_p is the increment of degree of freedom when F_p is added in to the model. For the linear model, $d_p = 1$ if F_p is

a numerical factor, and $d_p =$ the number of factor levels $- 1$ if F_p is a nominal factor.

The proposed algorithm for SBPM with variable selection based on $\mathbf{P}'$ consists of the following steps:

Step 1. Set $t = 0$ and $\mathcal{P}^0 = \phi$, where ϕ denotes the empty set.

Step 2. Calculate M=AIC$^{\mathcal{P}^0}_{\text{SPBM}}$.

Step 3. $t = t + 1$.

Step 4. Find $p^* = \arg\min_{p \in \mathbf{P}' \setminus \mathcal{P}^{t-1}} \text{AIC}^{\{\mathcal{P}^{t-1}, p\}}_{\text{SPBM}}$. Set $\mathcal{P}^t = \{\mathcal{P}^{t-1}, p^*\}$.

Step 5. If $M >$AIC$^{\mathcal{P}^t}_{\text{SPBM}}$ or $t = |\mathbf{P}'|$, set $M =$ AIC$^{\mathcal{P}^t}_{\text{SPBM}}$ and go to Step 3.

Step 6. Set $\mathcal{P}^* = \mathcal{P}^{t-1}$, $\widehat{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}}(\mathcal{P}^{t-1})$, $\widehat{\sigma}^{2*} = \widehat{\sigma}^2(\mathcal{P}^{t-1})$, $\widehat{\rho}^{2*} = \widehat{\rho}^{\,2}(\mathcal{P}^{t-1})$ and the final model follows.

Finally, the fitted model is as follows:

$$\hat{Y}_{kn} = \mathbf{X}^{\mathcal{P}^*}_{kn} \widehat{\boldsymbol{\beta}}^*. \tag{28}$$

Indeed, the above steps are essentially based on the stepwise forward selection for efficiency. If the efficiency is not important, the proposed algorithm can be based on stepwise backward elimination or stepwise bi-direction construction.

For the model evaluation, if the fitted model is not satisfied, the factors without physical meaning can be removed, and then perform the above algorithm from Step 3 with reset t and $\mathcal{P}^t$ to get an adjusted model. Also, the residuals should be examined to check the goodness of fit of the constructed model [38]. Indeed, the SBPM contains two types of residuals: (a) between-batch residuals $e^b_k = \sum_{k=1}^{L} e^{\zeta}_{kn}/N_k$, $k = 1, \ldots, K$ and (b) within-batch residuals $e^w_{kn} = e^{\zeta}_{kn} - e^b_k$, $k = 1, \ldots, K, n = 1, \ldots, N_k$, where $e^{\zeta}_{kn} = Y_{kn} - \hat{Y}_{kn}$.

D. Time Complexity

The time complexity of the proposed framework to deal with semiconductor manufacturing big data involved in sub-batch systems can be estimated via the involved two computing stages: (a) key factor extraction by statistical analysis and (b) model construction and evaluation. Suppose that N observations and P observed factors are considered. In (a), assume that the number of trees to be grown in RF is M_T. Since the time complexity of CART with the number of factor $\sqrt{P}$ is $O(\sqrt{P}N \log N)$, the computing time complexity of RF is $O(M_T\sqrt{P}N \log N)$. The time complexity to fit a one-dimensional linear regression model by least square approach is $O(N)$. For a SBPM, a sequence of iterations will be performed to derive the optimal likelihood function. For each iteration, the time complexity is similarly to fit a linear regression model that is equivalent to solve $\boldsymbol{\beta}$ and σ^2_w under a fixed ρ. Let the number of iterations be κ, the time complexity of SBPM part in (a) will be $O(\kappa PN)$. Note that κ is not a fixed number, yet $\kappa = 30$ is enough to derive a desired precision in most of the realistic cases. The total time complexity for calculating $\lambda_p, p = 1, \ldots, P$ and sorting them is at most $O(P \log P)$ via a quick sorting algorithm. The rest of computing in (a) is trivial and can be ignored. Thus, the time complexity of (a) can be expressed by $O(M_T\sqrt{P}N \log N + \kappa PN + P \log P)$. If M_T and κ are both fixed, it can be replaced by $O(NP \log N \log P)$. In (b), suppose that the number of candidate factors selected in (a) is P'. For the stepwise forward selection algorithm, at most $P'(P' + 1)/2$ SBPM will be constructed, and the time complexity of a SBPM is limited by $O(\kappa P'^2 N)$ that is the time complexity of constructing a P'-variate SBPM. Thus, the time complexity of (b) can be expressed by $O(P'^4 N)$ if κ is fixed. Hence, the time complexity of the proposed framework is up to $O(P^4 N log N)$, that can be reduced to $O(P^2 N\ logN)$ if $P' \approx O(\sqrt{P})$.

V. Validation

A. Simulation Study

1) Simulation Setting: Let Y denote the response variable, i.e., yield. Some terminologies are defined below to facilitate the simulation setting:

- Effective factor: the factor that effectively affects Y.
- Ineffective factor: the factor that is not an effective factor.
- Between-batch noise: the noise used to describe the center of the remained stream effect in the sub-batch processing system, which is denoted by ε^b_k in (9).
- Within-batch noise: the noise used to describe the sum of the independent effect and the pure noise of products, which is denoted by ε^w_{kn} in (9).
- SBPM-based analysis: The analysis based on SBPM.
- Product-based analysis: The analysis based on the model shown as (1), which ignores the dependency of the products within batch (see Section II-A).
- Batch-based analysis: The analysis based on the model shown as (2) and batch level data (see Section II-B).
- Factor-assumed analysis: The analysis based on the model shown as (1) with a dummy factor to handle the batch issue (see Section II-C).

In this study, both nominal and numerical factors are considered in three types of simulations to facilitate the comparison: (a) numerical factor case, (b) nominal factor case and (c) mixed factor case. In particular, 1000 factors containing 100 interested factors are considered, in which the numbers of numerical factors and nominal factors are half-and-half in the mixed scenarios. In the 100 interested factors, 10 factors are randomly selected and set as the effect factors of effect 1. In the remaining 900 factors, 100 factors are randomly selected and set as the effective factors of effect τ, where τ of range 0.1 to 0.3 is a parameter to control the simulation scenario. Pure noises are *i.i.d.* generated from a normal distribution with mean 0 and variance σ^2_w, where σ^2_w of range 0.2 to 5 is a parameter to control the simulation scenario. Uses $F_p, p = 1, \ldots, 100$ denote the 100 interested factors $F_p, p = 101, \ldots, 1000$ denote the other 900 factors, and use X_{pkn} and ε^w_{kn} denote the value of factor F_p and pure noise of product n of batch k. Thus, the underlying model can be written as

$$Y_{kn} = \beta_0 + \sum_{p=1}^{100} \beta_p X_{pkn} + \sum_{p=101}^{1000} \beta_p X_{pkn} + \varepsilon^w_{kn}, \tag{29}$$

$k = 1, \ldots, K, n = 1, \ldots, N_k$, where $\beta_p, p = 1, \ldots, 1000$ are the corresponding effects and β_p will be set as 0 if F_p is an ineffective factor. K is a parameter to control the simulation

scenario, and N_k is randomly chosen from 10 to 25. In order to simulate the real situation well, the observations of X_{pkn}s are generated as follows. In terms of numerical factors, the observations of X_{pkn}s are randomly and independently generated from a standard normal distribution. In terms of nominal factors, the observations of X_{pkn}s are used to describe the used tools, chambers, recipes and etc., and in general, we use "level" to express it. For example, if a product in some stage is processed by Tool A, it has level A in this factor. If it is processed by Tool B, this factor has level B. Two levels are set for each of the nominal factors. To fit the situation of sub-batch process system shown in Section III-A2, two types of nominal factors are considered: (a) factor of a batch processing stage and (b) factor of a sub-batch process stage. For (a), the observations of $X_{pkn},\ n = 1, \ldots, N_k$ are set the same to imitate the characteristics of the batch processing system (that is, the products in the same batch take the same level in the particular factor). In this case, X_{pk1} is randomly generated from a Bernoulli distribution with mean θ_p randomly chosen from 2/5 to 3/5. $X_{pk1} = 0$ or 1 denoted the products of batch k with the level-$p0$ or level-$p1$ in F_p. For (b), the observations of $X_{pkn},\ n = 1, \ldots, N_k$ are not necessary to be the same to imitate the characteristics of the sub-batch processing system. In this case, $X_{pkn},\ n = 1, \ldots, N_k$ are randomly and independently generated from a Bernoulli distribution with mean θ_p randomly chosen from 2/5 to 3/5. The response Y can be generated according to (29). Since the setting of β_0 would not affect the analysis performance, it is set as 0.

2) Validation of Single Factor Analysis: For comparing the results of single factor analyses of the SBPM-based analysis, the product-based analysis, the batch-based analysis, and the factor-assumed analysis, their p-values for each of the factors are calculated, respectively. The factors with p-values less than the significant level α, set as 0.05 in this study, will be regarded as effective factors. Two measures, (a) Type I error R_α and (b) screening accuracy R_a, were used to evaluate the performance:

$$R_\alpha = \frac{p_s - p_c}{p_t - p_e}, \tag{30}$$

$$\mathrm{R_a} = \frac{2\mathrm{p_c} + \mathrm{p_t} - \mathrm{p_e} - \mathrm{p_s}}{\mathrm{p_t}}, \tag{31}$$

where $\mathrm{p_t}$, $\mathrm{p_e}$, $\mathrm{p_s}$ and $\mathrm{p_c}$ are the numbers of total factors, effective factors, selected factors and catching factors as illustrated in Fig. 9. In general, R_α in (30) is a nominal-the-better measure that targets the significant level α; $\mathrm{R_a}$ in (31) is a larger-the-better measure that corresponds to the consistency between the dot and dash circles.

The simulations are repeated and the average of R_αs and $\mathrm{R_a}$s of 100 repeats are shown as Figs. 10–15. In terms of Type I error, Fig. 10, the case of numeric factors, shows the derived Type I errors of four approaches are all closed to 0.05. Since the settings of numeric factors in the simulations are independent of batch/sub-batch, the relation of the products between batches is the same as that within a batch. Thus, all four approaches can lead to the same Type I errors. In Fig. 11, the case of nominal factors, only

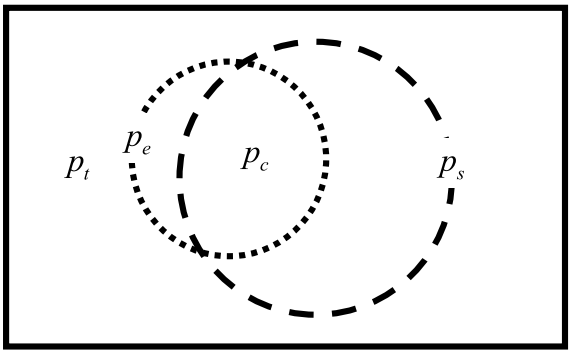

Fig. 9. Illustration for the numbers of the factors in various sets. The solid rectangle expresses the set of total factors with size p_t; the dot circle expresses the set of effective factors with size p_e; the dash circle expresses the set of selected factors with size p_s; and the intersection of dot and dash circles expresses the set of catching factors with size p_c.

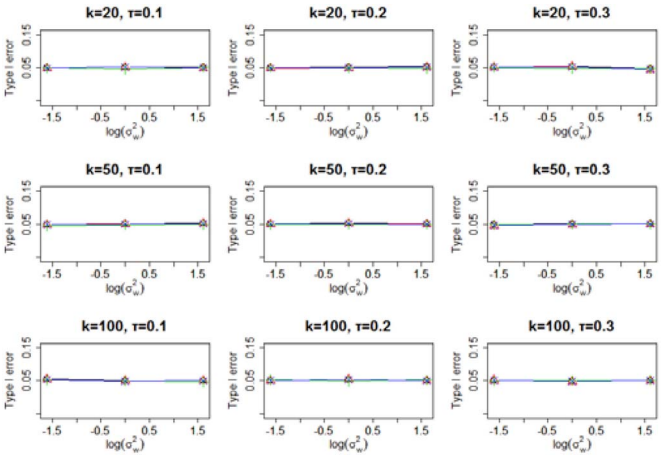

Fig. 10. Type I error of numerical input case. Four notations "◯," "Δ," "+," and "×" correspond to results of SBPM, product-based, batch-based, and factor-assumed analyses, respectively.

SBPM-based and batch-based analyses can lead to the controlled Type I errors. The results show that product-based and factor-assumed analyses cannot control the Type I error well in the case of nominal factors; in other words, the model risk cannot be controlled well. Note that the difference between the set significant level 0.05 and Type I errors of product-based and factor-assumed analyses are decreasing as $\log(\sigma_{\mathrm{w}}^2)$. The phenomenon can be reasonable explained as follows: Indeed, the deviations of Type I errors of the batch-based and factor-assumed analyses are mainly derived from the assumption that $\zeta_{\mathrm{kn}}, \mathrm{k} = 1, \ldots, \mathrm{K}, \mathrm{n} = 1, \ldots, \mathrm{N_k}$ in (8) are mutually independent. $\zeta_{\mathrm{kn}} = \xi_{\mathrm{kn}} + \varepsilon_{\mathrm{kn}}$ contains the remained stream effect and pure noise, where the dependencies of ζ_{kn}s are contributed by the remained stream effects. A bigger σ_{w}^2 will lead to relatively lower dependencies of ζ_{kn}s, and thus a lower Type I error. The same problem appears in the case of mixed factors as shown in Fig. 12, in which only SBPM-based and batch-based analyses can lead to the correct Type I errors. Therefore, product-based and factor-assumed analyses cannot evaluate the risk well in the cases that contain nominal factors and thus are inappropriate in practice.

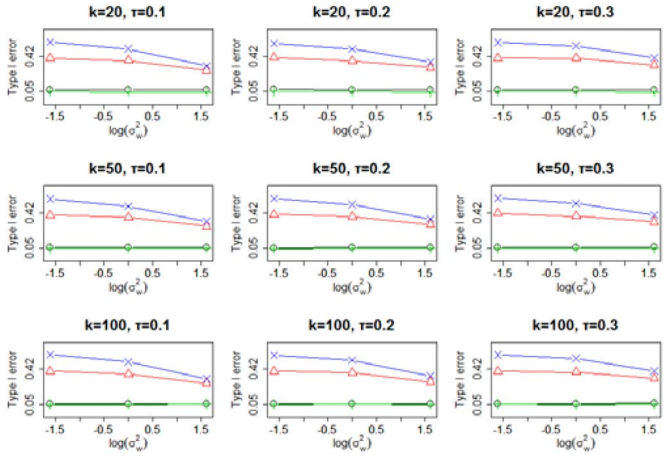

Fig. 11. Type I error of nominal input case. The symbol descriptions can be referred in Fig. 10.

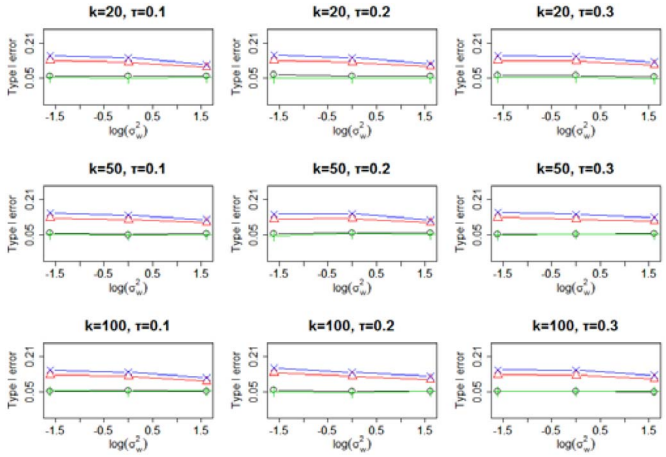

Fig. 12. Type I error of mixed input case. The notations can be referred in Fig. 10.

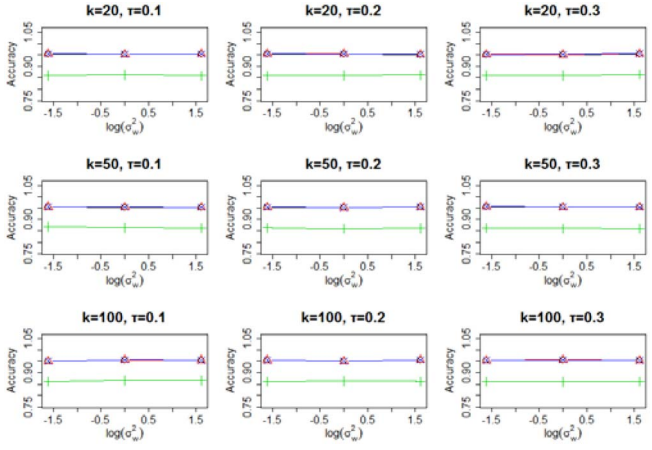

Fig. 13. Accuracy of numerical input case. Four notations "◯," "Δ," "+," and "×" correspond to results of SBPM, product-based, batch-based, and factor-assumed analyses, respectively.

In terms of accuracy, Fig. 13 denoting the case of numeric factor shows the accuracy of batch-based analysis is lower than the others. Comparing with Fig. 11, the Type I errors of

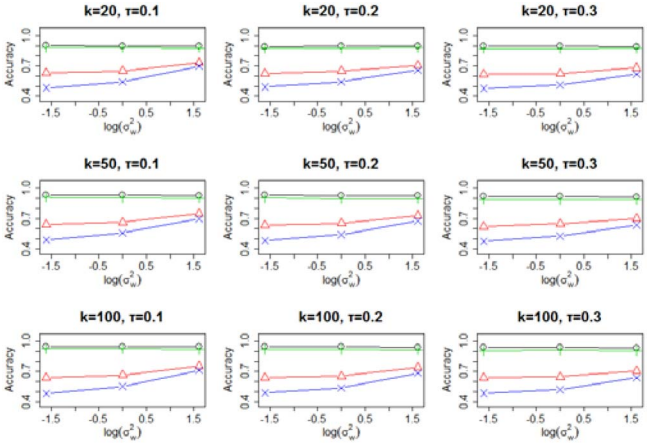

Fig. 14. Accuracy of nominal input case. The symbol descriptions can be referred in Fig. 13.

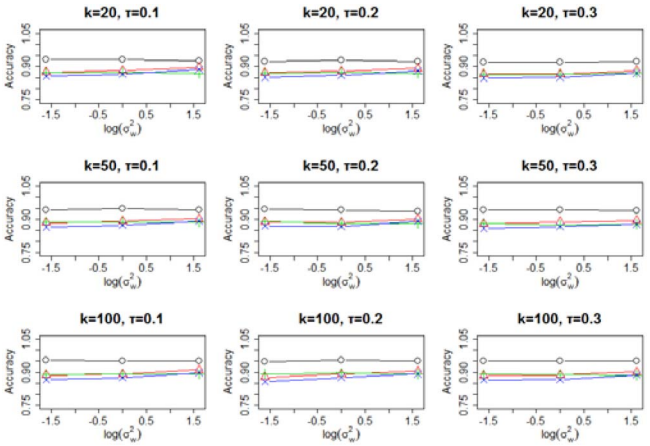

Fig. 15. Accuracy of mixed input case. The notations can be referred in Fig. 13.

four approaches are similar, yet the Type II error defined as $(p_e - p_c)/p_e$ of batch-based analysis is higher than the others. That is, the power to detect the effective factors of batch-based analysis is lower than the others, since the batch-based analysis does not consider the information within the batches. In Fig. 14, the case of nominal factors, the accuracies of product-based and factor-assumed analyses are much lower than that of SBPM-based and batch-based analyses since the Type I errors of product-based and factor-assumed analyses are much higher than SBPM-based and batch-based analyses. Furthermore, the accuracy of batch-based analysis is lower than that of SBPM-based analysis since the Type II error of batch-based analysis is higher than that of SBPM-based analysis. Indeed, the accuracies of product-based and factor-assumed analyses are increasing as $\log(\sigma_w^2)$. Similarly, since a higher σ_w^2 will lead to the lower correlations between ζ_{kn}s, and therefore the lower Type I error and accuracy follows. The accuracy of SBPM-based analysis is outstanding among the other approaches for the case of mixed factors in Fig. 15, since the proposed SBPM-based framework can deal with both the numeric and nominal factors effectively in the sub-batch processing system.

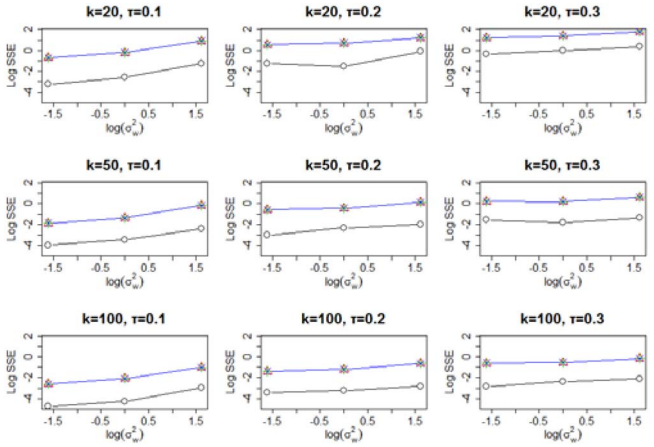

Fig. 16. Log-SSE of numerical input case. There are four approaches compared. Notations "◯," "Δ," "+," and "×" correspond to the results of the SBPM-based root cause detection approach, stepwise regression, Lasso, and LARS.

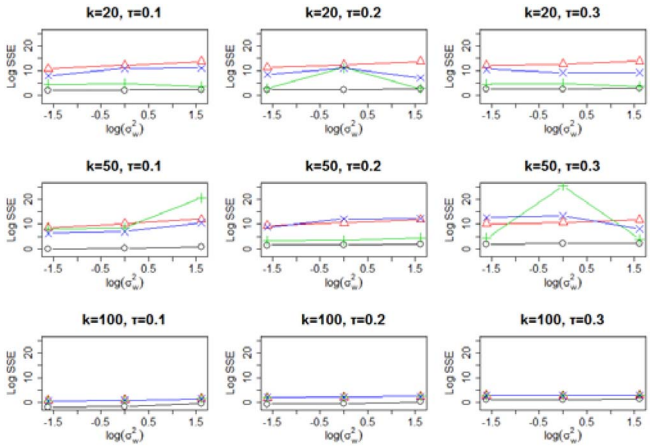

Fig. 17. Log-SSE of nominal input case. The symbol descriptions can be referred in Fig. 16.

3) Validation of SBPM-Based Root Cause Detection Framework: The proposed SBPM-based root cause detection framework is validated by comparing it with several model selection approaches including stepwise regression [23]. To highlight the effect of effective factors, the evaluation measure is defined as

$$R_{SSE} = \sum_{p=1}^{100} \widehat{\beta}_p - \beta_p, \tag{32}$$

where $\beta_p = 1$, if F_p is an effective factor; $\beta_p = 0$, if F_p is an ineffective factor. R_{SSE} is a smaller-the-better measure.

Similarly, the simulations are conducted and the average values of R_{SSE}s of 100 repeats are shown as Figs. 16–18. The performance of the SBPM -based root cause detection frame leads to a better performance than the others in all simulation scenarios. In general, the R_{SSE} is increasing as τ and σ_w^2 and decreasing as k. The performance of stepwise regression, Lasso and LARS [23] are similar in the case of numerical factors (see Fig. 16), but they are different in the case of nominal factors (see Fig. 17). The performance of Lasso is unstable in the case of nominal factors.

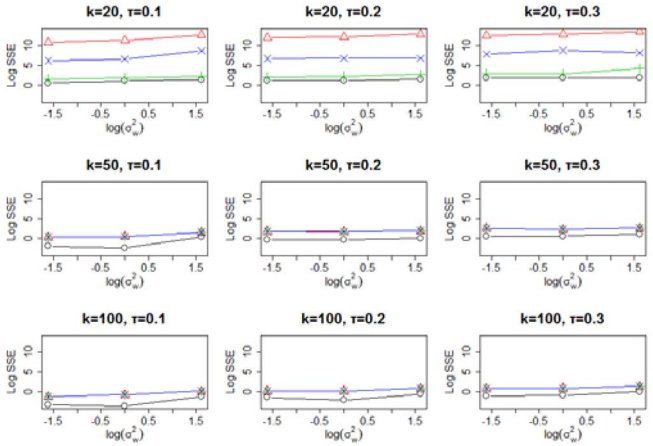

Fig. 18. Log-SSE of mixed input case. Symbol descriptions can be referred in Fig. 16.

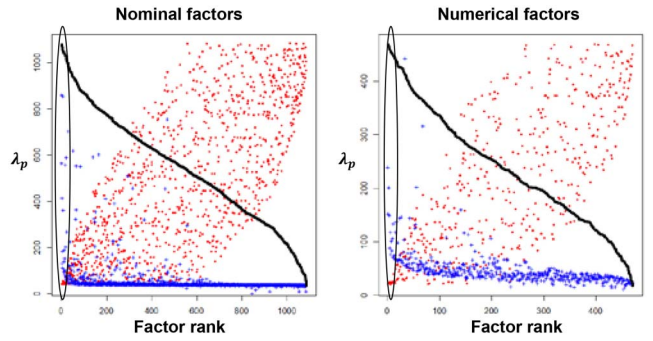

Fig. 19. Illustration of key factor screening of empirical study. The relation between factor rank and λ_p is described as the bold solid line. The symbol "·" and "+" show the relative V_{Bp} and V_{Rp} of the associated factor of the factor rank.

B. Empirical Study

To validate the proposed framework, an empirical study was conducted in a leading wafer fab consists of sub-batch processing systems in Taiwan for root cause detection.

1) Problem Definition and Data Preparation: The response Y was the wafer CP yield, and the considered factors were the parameters of processing record and wafer acceptance test (WAT). The collected data included 447 wafers of 18 lots with 1 CP-yield and 3919 factors. The factors consisted of 938 tool-related factors, 469 recipe-related factors, 469 module-related factors, 469 vendor-related factors and 1574 WAT-related factors, and the first four were nominal factors and the last one was numerical factors.

Firstly, since the raw data contained a lot of missing elements, data preparation was performed including imputation of missing elements [39], [40]. There were 1641 factors remained, with 739 tool-related, 203 recipe-related, 102 module-related, 41 vendor-related and 471 WAT-related factors.

2) Key Factor Screening: A number of key factors were selected from the remaining 1641 factors. Since data properties are different, the key factor screening was performed for both nominal and numerical factors. In particular, the measure λ_p in (25) with $\omega = 0.5$ was used to rank and select the key factors as illustrated in Fig. 19. The left and right plots are

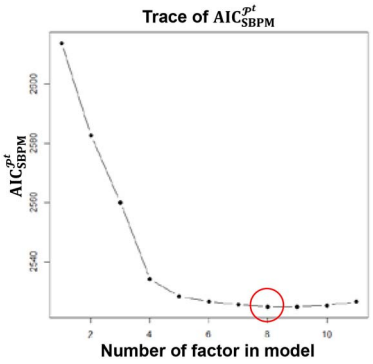

Fig. 20. Trace of AIC stepwise regression with forward selection. The red circle indicates where the minimal AIC is.

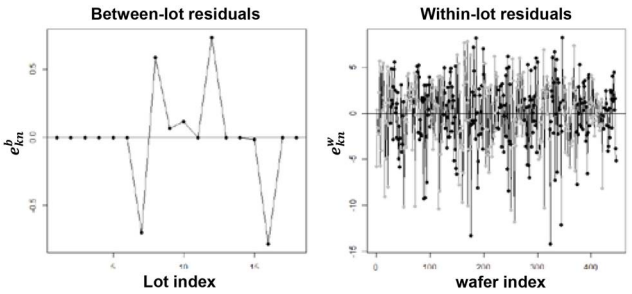

Fig. 21. Residuals. Left: between-lot residuals. Right: within-lot residuals. The colors of points are used to distinguish from lots.

associated to the nominal and numerical factors, respectively. For example, the 1170 nominal factors are ranked according to the λ_p. The relation of factor ranks and λ_p can be represented as the bold solid line. The symbol "·" and "+" are the relative V_{Bp} and V_{Rp} of the factor associated to the factor rank. The factors of high λ_p and $V_{Bp} \leq 0.05$ were extracted as the key factors of 31 nominal factors and 14 numerical factors and highlighted by ellipses.

3) Model Construction and Evaluation: Fig. 20 shows the trace of $\text{AIC}^{\mathcal{P}^t}_{\text{SBPM}}$ when the factors were added into the model, in which the $\text{AIC}^{\mathcal{P}^t}_{\text{SBPM}}$ was minimized as the number of factors was 8. Hence, the associated 8-factors model was fitted as follows:

$$\begin{aligned}\hat{Y}_{kn} = {} & \beta_0 + \beta_1 S667\text{chrcp}_{kn} + \beta_2 S284\text{chid}_{kn} \\ & + \beta_3 S1230\text{chid}_{kn} + \beta_4 S1106\text{chrcp}_{kn} \\ & + \beta_5 WAT872_{kn} + \beta_6 WAT1067_{kn} \\ & + \beta_7 WAT99_{kn} + \beta_8 WAT1529_{kn},\end{aligned}$$

$k = 1, \ldots, 18, n = 1, \ldots, N_k$, where the factors with prefixes *S-* and *WAT-* denoted the nominal factor and the numerical factor respectively. Note that "chrcp" and "chid" meant the used chamber recipe and chamber index. The result showed that CP-yield mainly depended on the tool and WAT parameters, while other factors such as tool module and vendor were relatively ineffective to the CP-yield.

The between-lot residuals and within-lot residuals were illustrated in Fig. 21. Since the number of lots $k = 18$ was less than the number of nominal predictors, a part of between-lot residuals were equal to zero. However, it did not affect the

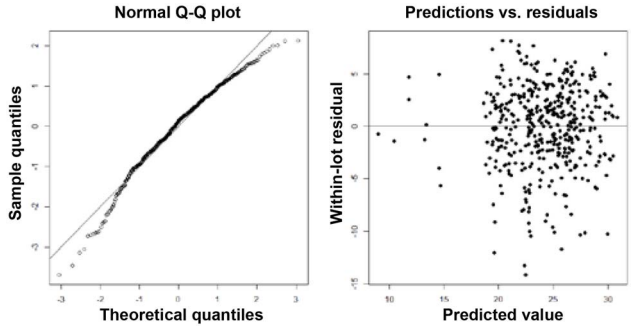

Fig. 22. Within-lot residuals diagnosis. Left: quintile–quentile plot of residuals of within-lot residuals. Right: expected values of response to residuals of within-lot residuals.

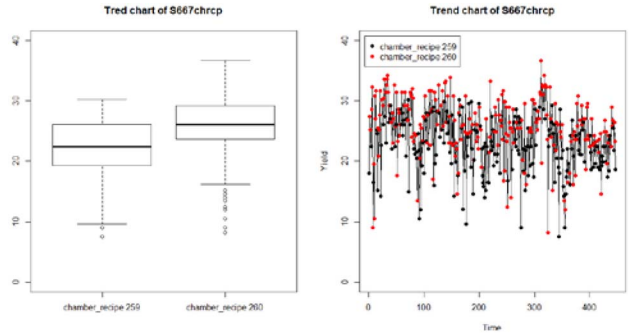

Fig. 23. Box plot and trend chart of S667chrcp.

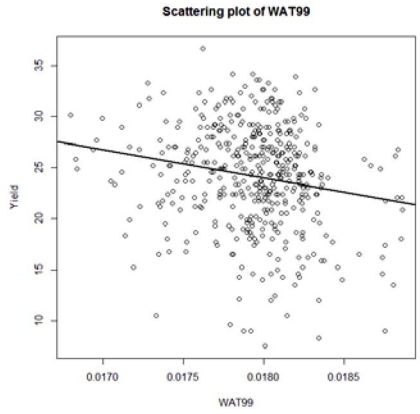

Fig. 24. Scattering plot of WAT99.

effect estimation too much, that could be avoided by collecting more data. Furthermore, Fig. 22 shows within-lot residual analysis. The quantile-quantile plot in Fig. 22 showed a little negative skew that could be fixed by taking a transformation. The right side of Fig. 22 showed the plot of $\hat{Y}$ to the within-batch residuals showing no obvious pattern.

Two effective factors were detected in this study. Fig. 23 showed that the CP-yield of wafers processed by chamber recipe 260 in the 667th stage was averagely greater than that by chamber recipe 259 over all the time, showing that the chamber recipe 259 in 667th stage was a root cause of yield-loss. Fig. 24 showed that the CP-yield was decreasing in WAT99. The above comparison and empirical study have validated the proposed approach.

VI. Concluding Remarks

Although sub-batch processing systems are common in advanced wafer fab, little research has been done to address the needs in real settings. This study proposed an effective framework to detect the root causes for sub-batch processing system for semiconductor manufacturing. The proposed framework contains the key factor screening step to enhance the efficiency and integrates the RF and SBPM to deal with the collinearity involved in semiconductor manufacturing. It can robustly deal with the massive data collected from the sub-batch processing systems with relatively high catching accuracy and small errors.

This study has dealt with the issue of sub-batch processing systems involved in semiconductor manufacturing big data analytics. The proposed approach was validated by simulations and an empirical study conducted in a leading semiconductor manufacturing company in Taiwan. More studies should be done to address other big data issues for advanced intelligent semiconductor manufacturing [41]. Indeed, the proposed approach has been embedded in engineering data analysis system in the case company for on-line implementation. Furthermore, the proposed framework is adaptive to various types of data and underlying models for semiconductor manufacturing big data analytics. Also, future study can be done to develop robust hypothesis testing when the sample size is small in the early stage of ramping advanced technology nodes.

References

[1] C.-F. Chien, Y.-J. Chen, and J.-T. Peng, "Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and product life cycle," *Int. J. Product. Econ.*, vol. 128, no. 2, pp. 496–509, 2010.

[2] R. C. Leachman, S. Ding, and C.-F. Chien, "Economic efficiency analysis of wafer fabrication," *IEEE Trans. Autom. Sci. Eng.*, vol. 4, no. 4, pp. 501–512, Oct. 2007.

[3] C.-F. Chien, S. Dauzere-Peres, H. Ehm, J. Fowler, Z. Jiang, S. Krishnaswamy, T. Lee, M. Lars, and R. Uzsoy, "Modeling and analysis of semiconductor manufacturing in a shrinking world: Challenges and successes," *Eur. J. Ind. Eng.*, vol. 5, no. 3, pp. 254–271, 2011.

[4] S. P. Cunningham, C. J. Spanos, and K. Voros, "Semiconductor yield improvement: Results and best practices," *IEEE Trans. Semicond. Manuf.*, vol. 8, no. 2, pp. 103–109, May 1995.

[5] C.-F. Chien, W. C. Wang, and J. C. Cheng, "Data mining for yield enhancement in semiconductor manufacturing and an empirical study," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 192–198, 2007.

[6] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[7] S. Thomassey, "Sales forecasts in clothing industry: The key success factor of the supply chain management," *Int. J. Product. Econ.*, vol. 128, no. 2, pp. 470–483, 2010.

[8] C.-F. Chien and L. Chen, "Using rough set theory to recruit and retain high-potential talents for semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 528–541, Nov. 2007.

[9] A. Sharma and P. K. Panigrahi, "A review of financial accounting fraud detection based on data mining techniques," *Int. J. Comput. Appl.*, vol. 39, no. 1, pp. 37–47, 2012.

[10] S. Evans, S. Leman, C. Deters, R. Fusaro, and H. Lynch, "Automated detection of hereditary syndromes using data mining," *Comput. Biomed. Res.*, vol. 30, pp. 337–348, Oct. 1997.

[11] C.-F. Chien, S. Chen, and Y. Lin, "Using Bayesian network for fault location on distribution feeder," *IEEE Trans. Power Del.*, vol. 17, no. 13, pp. 785–793, Jul. 2002.

[12] A. Kusiak, "Rough set theory: A data mining tool for semiconductor manufacturing," *IEEE Trans. Electron. Packag. Manuf.*, vol. 24, no. 1, pp. 44–50, Jan. 2001.

[13] J. A. Harding, A. Kusiak, M. Shahbaz, and T. Srinivas, "Data mining in manufacturing: A review," *J. Manuf. Sci. Eng.*, vol. 128, pp. 969–976, Nov. 2006.

[14] D. Braha and A. Shmilovici, "Data mining for improving a cleaning process in the semiconductor industry," *IEEE Trans. Semicond. Manuf.*, vol. 15, no. 1, pp. 91–101, Feb. 2002.

[15] S. Hsu and C.-F. Chien, "Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing," *Int. J. Product. Econ.*, vol. 107, no. 1, pp. 88–103, 2007.

[16] C.-W. Liu and C.-F. Chien, "An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing," *Eng. Appl. Artif. Intell.*, vol. 26, nos. 5–6, pp. 1479–1486, 2013.

[17] C. Kuo, C.-F. Chien, and J. Chen, "Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time," *IEEE Trans. Autom. Sci. Eng.*, vol. 8, no. 1, pp. 103–111, Jan. 2011.

[18] C.-F. Chien, C.-Y. Hsu, and C.-W. Hsiao, "Manufacturing intelligence to forecast and reduce semiconductor cycle time," *J. Intell. Manuf.*, vol. 23, no. 6, pp. 2281–2294, 2012.

[19] Q. P. He and J. Wang, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.

[20] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[21] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognit.*, vol. 44, no. 2, pp. 330–349, 2011.

[22] S. Weisberg, *Applied Linear Regression*. New York, NY, USA: Wiley, 1980.

[23] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Brit.*, vol. 58, no. 3, pp. 267–288, 1996.

[25] W. H. Kruskal and W. A. Wallis, "Use of ranks on one-criterion variance analysis," *J. Amer. Statist. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952.

[26] H. D. Patterson and R. Thompson, "Recovery of inter-block information when block sizes are unequal," *Biometrika*, vol. 58, no. 3, pp. 545–554, 1971.

[27] P. Diggle, P. Heagerty, K. Y. Liang, and S. Zeger, *Analysis of Longitudinal Data*, 2nd ed. New York, NY, USA: Oxford Univ. Press, 2002.

[28] C. A. McGilchrist, "Regression analysis of dependent error models," *Bull. Aust. Math. Soc.*, vol. 34, pp. 199–209, 1986.

[29] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, *Longitudinal Data Analysis*. Boca Raton, FL, USA: CRC Press, 2009.

[30] D. P. Kennedy, "A note on stochastic search methods for global optimization," *Adv. Appl. Probab.*, vol. 20, no. 2, pp. 476–478, 1988.

[31] R. H. Myers and D. C. Montgomery, *Response Surface Methodology*, 2nd ed. New York, NY, USA: Wiley, 2002.

[32] K. Y. Liang and S. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[33] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.

[34] L. R. Fabrigar and D. T. Wegener, *Exploratory Factor Analysis* (Understanding Statistics). New York, NY, USA: Oxford Univ. Press, 2012.

[35] G. M. Furnival, "All possible regressions with less computation," *Technometrics*, vol. 13, no. 2, pp. 403–408, 1971.

[36] R. R. Hocking, "A biometrics invited paper. The analysis and selection of variables in linear regression," *Biometrics*, vol. 32, no. 1, pp. 1–49, 1976.

[37] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[38] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 4th ed. Hoboken, NJ. USA: Wiley, 2006.

[39] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*. New York, NY, USA: Wiley, 1987.

[40] L. Breiman. (2002). *Manual On Setting Up, Using, and Understanding Random Forests V3.1* [Online]. Available: http://oz.berkeley.edu/users/breiman/UsingrandomforestsV3.1.pdf

[41] C.-F. Chien, Y.-J. Chen, C.-Y. Hsu, and H.-K. Wang, "Overlay error compensation using advanced process control with dynamically adjusted proportional-integral R2R controller," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 473–484, Apr. 2014.

Chen-Fu Chien received the B.S. (Phi Tao Phi Hons.) degree with double majors in industrial engineering and electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 1990, the M.S. degree in industrial engineering, and the Ph.D. degree in operations research and decision sciences from the University of Wisconsin-Madison, Madison, WI, USA, in 1994 and 1996, respectively, and the PCMPCL Training from Harvard Business School, Boston, MA, USA, in 2007. He is a Tsing Hua Chair Professor with NTHU. He is also the Director of the NTHU-Taiwan Semiconductor Manufacturing Company (TSMC) Center for Manufacturing Excellence and the Principal Investigator for the NSC Semiconductor Technologies Empowerment Partners Consortium. From 2002 to 2003, he was a Fulbright Scholar with the University of California, Berkeley, Berkeley, CA, USA. From 2005 to 2008, he had been on-leave as a Deputy Director with Industrial Engineering Division, TSMC. His research efforts center on decision analysis, modeling and analysis for semiconductor manufacturing, manufacturing strategy, and manufacturing intelligence. He has received eight invention patents on semiconductor manufacturing and published four books, over 120 journal papers, and a number of case studies in Harvard Business School. He has been invited to give keynote lectures at several conferences, including APIEMS, C&IE, IEEM, IML, and leading universities worldwide. He was the recipient of the National Quality Award, the Distinguished Research Awards, and the Tier 1 Principal Investigator (Top 3%) from the National Science Council, the Distinguished University-Industry Collaborative Research Award from the Ministry of Education, the University Industrial Contribution Awards from the Ministry of Economic Affairs, the Distinguished University-Industry Collaborative Research Award and the Distinguished Young Faculty Research Award from NTHU, the Distinguished Young Industrial Engineer Award, the Best IE Paper Award, and the IE Award from Chinese Institute of Industrial Engineering, the Best Engineering Paper Award and the Distinguished Engineering Professor by Chinese Institute of Engineers in Taiwan, and the 2012 Best Paper Award of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING. He is currently an Area Editor of the *Flexible Services and Manufacturing Journal*, an Editorial Board Member of *Computers and Industrial Engineering*, and an Advisory Board Member of *OR Spectrum*.

Shih-Chung Chuang received the Ph.D. degree from the Institute of Statistics, National Central University, Jhongli, Taiwan, in 2009. He is a Post-Doctoral Fellow with National Tsing Hua University-Taiwan Semiconductor Manufacturing Company (NTHU-TSMC) Center for Manufacturing Excellence, Department of Industrial Engineering and Engineering Management, NTHU, Hsinchu, Taiwan. His research interests include multivariate data analysis, data mining, queueing system, and experimental design.