Abstract—Accurate yield prediction to evaluate productivity, and to estimate production costs, is a critical issue in the highly competitive semiconductor industry. We propose yield models based on hierarchical Bayesian modeling of clustered spatial defects produced in integrated circuits (IC) manufacturing. We use spatial locations of the IC chips on the wafers as covariates, and develop four models based on Poisson regression, negative binomial (NB) regression, zero-inflated Poisson (ZIP) regression, and zero-inflated negative binomial (ZINB) regression. Along with the hierarchical Bayesian approaches, spatial variations of defects within one wafer as well as among different wafers are effectively incorporated in the yield models. Wafermap data obtained from an industrial collaborator are used to illustrate the proposed models. The results indicate that the Poisson regression model consistently underestimates the true yield because of extraneous Poisson variation caused by defect clustering. On the contrary, NB regression, ZIP regression, and ZINB regression models provide more reliable yield estimation and prediction in real applications.

Index Terms—Hierarchical bayesian model, spatial defects, yield prediction, zero-inflated models.

I. INTRODUCTION

A modern ultra-large-scale integrated circuit (ULSIC) consists of millions or billions of micro-/nano-scale devices, such as transistors, capacitors, and resistors, which are fabricated on a semiconductor wafer. The manufacturing of integrated circuits (ICs) is a highly complex, costly process that involves hundreds of chemical or physical processing steps. Among all the performance measures, yield is considered as a key in determining the success of semiconductor manufacturing [1]. Yield is usually defined as the ratio of the number of usable items after the completion of the manufacturing process to the number of potentially usable items at the beginning of the manufacturing process.
To secure high profits in the highly competitive semiconductor industry, the need for accurate yield prediction in order to evaluate productivity and estimate production costs is critically important [2].

The overall yield loss in semiconductor manufacturing can be broken down into three main components [3]: 1) loss due to errors such as breakage or wrong manufacturing sequence, 2) loss due to defects, and 3) packaging yield loss. Most semiconductor industries focus their efforts on reducing the yield loss due to defects, especially because the large profit from a relatively small investment can be attainable [1]. Today, even using highly automated, precisely positioned equipment located in a nearly dust-free clean room environment, the occurrence of defects still cannot be avoided [4]. A defect is defined as a variation of quality that may cause a circuit failure. If a defect occurs in a defect-sensitive area (called a “critical area”), it causes a failure, and is called a fault or fatal defect [1]. Wafer maps have been widely used by semiconductor manufacturers to monitor the manufacturing processes, and to provide information about defects on the wafers.

Defects are not randomly scattered on the wafer, but rather clustered [5]. Consequently, defective chips also tend to cluster [6]. The clustering is more evident as the wafer size increases. To accurately estimate and predict the wafer yield, it is imperative to explicitly consider the spatial clustering of defects or defective chips in the yield models.

There have been many attempts to model the yield of IC manufacturing since the 1960s. The classical Poisson and negative binomial yield models are the most frequently used. Let \( \lambda \) be the average number of faults per chip. When the faults are randomly distributed, and the occurrence of a fault at any location is \( s \)-independent of the occurrence of any other fault, the probability that a device contains \( k \) faults is calculated by the Poisson probability distribution. The probability that a chip does not contain any faults is then distributed according to the Poisson yield model. The Poisson yield model tends to underestimate the true yield due to its ignorance of the clustering effect of defects. To consider the defect density variation due to clustering, compound Poisson yield models have been introduced. Taking \( \lambda \) as a random variable following a gamma distribution with shape parameter \( \alpha \), and scale parameter \( \beta \),

\[
P_k = \int_{0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \cdot \frac{\lambda^{\alpha-1}e^{-\lambda/\beta}}{\Gamma(\alpha)\beta^\alpha} d\lambda = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(k+1)} \left( \frac{1}{1+\beta} \right)^\alpha \left( \frac{\beta}{1+\beta} \right)^k, \quad k=0,1,2,\ldots
\]

The number of faults per chip hence has a negative binomial distribution, and the negative binomial yield is given by

\[
Y_{rb} = P_0 = (1 + \beta)^{-\alpha} = \left( 1 + \frac{\overline{x}}{\alpha} \right)^{-\alpha},
\]

where \( \alpha \) is often referred to as a clustering parameter, which determines the degree of clustering of the yield model [1], and \( \overline{x} = \alpha/\beta \). Although the negative binomial yield model considers the clustering effect through the parameter \( \alpha \), it fails to completely capture the spatial feature of defects, which are commonly clustered at specific locations on the wafers, e.g., around the periphery of the wafer. Bae et al. [2] proposed spatial modeling approaches for clustered defects based on Poisson regression, negative binomial (NB) regression, and zero-inflated Poisson (ZIP) regression; and used a wafer map reported in the literature to illustrate their modeling approaches. They used the spatial location of each IC chip on the wafer as covariates for corresponding defect count listed in a wafer map. Analysis results indicate that yield estimation can be greatly improved by capturing spatial features of defects across the wafer map.

The aforementioned yield models estimate and predict yield by modeling the distribution of defects. Besides these statistical methods, there are many other methods which predict and improve yield by identifying critical manufacturing process attributes that cause yield loss, and exploring the relationship between yield and the process attributes. Critical process attributes may include history of tool usage, electrical and metrology measurements, etc. [7]. Artificial intelligence and data mining techniques are usually adopted to extract information from huge amounts of records and measurement data with many attributes. Lee and Ha [8] developed a hybrid yield prediction system combining neural networks and case-based reasoning. Chen and Hong [9] proposed a sample-efficient regression tree (SERT) approach for semiconductor yield loss analysis and prediction. Li et al. [10] presented a genetic programming data mining approach for constructing a semiconductor manufacturing yield prediction system. Kang et al. [11] used inductive decision trees, and neural networks with a back propagation algorithm, and a self-organizing mapping algorithm to manage yields over major manufacturing processes.

This study focuses on modeling the spatial distribution of defects, and extends the yield models developed by Bae et al. [2] by considering spatial variations of defect counts within one wafer, as well as across different wafers, and additionally proposes a new yield model based on zero-inflated negative binomial (ZINB) regression. The models in Bae et al. [2] considered only spatial variation of defects within one wafer. However, the spatial distribution of defects also varies across different wafers. Consideration of the spatial variation of defects across different wafers in a yield model can increase its capability of predicting the yield of future production. This may be achieved by introducing random effects to the regression models. For example, Poisson regression with random effects was considered by Boucher and Denuit [12] to analyze a Belgian motor third party liability portfolio. Hall [13] applied the ZIP regression with random effects to a case from horticulture. Yau and Lee [14] also applied the ZIP regression with random effects to evaluate an occupational injury prevention program. Yau et al. [15] used the ZINB regression with random effects to analyze a set of pancreas disorder inpatient length of stay data. All these studies used maximum likelihood (ML) methods to obtain estimation of the model parameters. In this study, we propose hierarchical Bayesian modeling of spatial variations of the defect distribution within one wafer, and across different wafers. There are three major motivations to use the Bayesian method instead of the ML method: 1) flexibility and generality of the Bayesian framework allows users to fit models with many parameters,
and complicated multi-layered probability specifications [16]; 2) the Bayesian method has better finite sample performance than the classical method with tighter interval estimates, and better coverage probabilities when analyzing zero-inflated regression models [17]; and 3) inference can always be provided without the need for approximations by using modern computational methods, e.g., Markov chain Monte Carlo (MCMC) [16].

The rest of this paper is organized as follows. Section II discusses the yield models based on Poisson regression, NB regression, ZIP regression, and ZINB regression using a hierarchical Bayesian modeling approach. Section III uses three real wafer maps as examples to illustrate the proposed Bayesian models. Finally, some concluding remarks are presented in Section IV.

II. BAYESIAN MODELING OF SPATIAL DEFECT COUNTS

The Poisson, NB, ZIP, and ZINB are the four most widely used distributions to model discrete count data. The negative binomial distribution is more flexible than the Poisson distribution in its ability to accommodate overdispersion; that is, the variance of the count data exceeds the Poisson variance λ. As indicated by (1), the negative binomial distribution can be derived by a Poisson-Gamma mixture model. If we assume, for a discrete nonnegative random variable X,

\[ X | u \sim \text{Poisson}(\lambda u) \quad \text{and} \quad u \sim \text{Gamma}(\alpha, \alpha), \quad (3) \]

the resulting marginal distribution of X is a negative binomial distribution with parameters \(\alpha/(\alpha + \lambda)\), and \(\alpha\) [18], i.e.,

\[ P(X = k) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\Gamma(k + 1)} \left( \frac{\lambda}{\alpha + \lambda} \right)^k \left( \frac{\alpha}{\alpha + \lambda} \right)^\alpha, \quad k = 0, 1, \ldots \]  

The inverse of the parameter \(\alpha\) is sometimes referred to as a dispersion parameter.

In some occasions, the count data may not follow the Poisson distribution, particularly inflated at zero. Extra dispersion appears as the number of observed zeros exceeds the number of expected zeros under the Poisson, or even the negative binomial distribution assumption. In recent years, there has been considerable interest in the zero-inflated distributions to model discrete count data to allow for the presence of excessive zeros. For example, Lewsey and Thomson [19] examined the utilization of the ZIP and ZINB models in a study of cross-sectional and longitudinal DMF (decayed, missing, filled) dental data. van den Broek [20] applied the ZIP model to analyze an HIV dataset. Yip and Yau [21] applied several zero-inflated count distributions to model claim frequency data in general insurance.

The ZIP distribution can be considered as a mixture of zero point mass, and a Poisson distribution:

\[ P(X = 0) = p + (1 - p) \exp(-\lambda), \]

\[ P(X = k) = (1 - p) \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 1, 2, \ldots \]  

Herein, \(0 \leq p \leq 1\), hence the extra zeros in the data are explicitly modeled. The ZINB distribution is a generalization of the ZIP model to account for overdispersion in the Poisson part of the ZIP distribution. The ZINB distribution has the probability mass function

\[ P(X = 0) = p + (1 - p) \left( 1 + \frac{\lambda}{\alpha} \right)^{-\alpha}, \]

\[ P(X = k) = (1 - p) \frac{\Gamma(k + \alpha)}{\Gamma(k + 1)\Gamma(\alpha)} \left( \frac{\lambda}{\lambda + \alpha} \right)^k \left( \frac{\alpha}{\lambda + \alpha} \right)^\alpha, \quad k = 1, 2, \ldots \]  

Using the data augmentation method discussed by Ghosh et al. [17], the random variable X in a zero-inflated model may be represented as \(X = V(1 - B)\), where \(B\) is a Bernoulli random variable with parameter \(p\), and \(V\), which is \(s\)-independent of \(B\), has a discrete distribution, e.g., Poisson for the ZIP model, and negative binomial for the ZINB model.

The four distributions can be extended to accommodate spatial coordinates as covariates in a regression setting, resulting in a Poisson regression model, NB regression model, ZIP regression model, and ZINB regression model. Lambert [23] proposed the ZIP regression model, and applied it to model defects in manufacturing. Martin et al. [24] reviewed the application of ZIP regression and ZINB regression to model zero inflation in ecological data. Bae et al. [2] applied the Poisson regression, NB regression, and ZIP regression to model spatial variation of defects within one semiconductor wafer, and developed yield models based on them. This section extends their work, and presents the hierarchical Bayesian approaches based on the four regression models.

A. Poisson Regression

Suppose that \(M\) integrated circuits are produced per wafer, and \(N_i\) defects are observed within mutually exclusive chip region \(A_i\) for \(i = 1, \ldots, M\). In the Poisson regression, we assume that \(N_i\) has a Poisson distribution with mean value that depends on the chip location as

\[ N_i \sim \text{Poisson}(\Lambda_i), \]

\[ \Lambda_i = \eta(\mathbf{f}(x_i)^T \beta), \quad \text{for} \quad i = 1, \ldots, M. \]  

Defect density shows radial and angular variation on the wafers [25]. Therefore, motivated by methods used in the field of spatial epidemiology [26], Bae et al. [2] proposed a link function of the form \(\eta(\mathbf{f}(x_i)^T \beta) = \exp(\mathbf{f}(x_i)^T \beta)\), where \(x_i = (r_i, \theta_i)\) is the polar coordinates of the \(i\)th chip center relative to the wafer center, and \(\mathbf{f}(x_i) = \{r_i \cos \theta_i, r_i \sin \theta_i, r_i r_i \cos \theta_i, r_i \sin \theta_i \}^T\).

Given the observed data \(n = \{n_1, n_2, \ldots, n_M\}\), the maximum likelihood estimate (MLE) of \(\beta\) can be obtained by maximizing the likelihood function

\[ L(\beta | n) = \prod_{i=1}^{M} \exp \left[ - \exp \left( \mathbf{f}(x_i)^T \beta \right) \right] \left[ \exp \left( \mathbf{f}(x_i)^T \beta \right) \right]^{n_i} / n_i!. \]

The Poisson regression model given by (7) only considers spatial variation of defect counts within one wafer. The Poisson regression model may be extended to include random effects such that spatial variations of defect counts within one wafer as well as across different wafers can be modeled. In this study, we propose to tackle the problem using the hierarchical Bayesian approach.
Assume that defect counts data on $K$ wafers are available, and the wafers have identical patterns of chips. Denote the data by $d \equiv \{n_1, n_2, \ldots, n_K\}$, where $n_j \equiv \{n_{ij_1}, n_{ij_2}, \ldots, n_{ij_M}\}$ denotes the defects data collected on the $j$th wafer, and $n_{ij}$ is the number of defects observed on the $i$th chip on the $j$th wafer. The defect counts on the $j$th wafer are modeled by the Poisson regression model with parameters $\beta_j$, i.e.,

$$N_{ij} \sim \text{Poisson}(\lambda_{ij}),$$

$$\lambda_{ij} = \exp \left( f(x_{ij})^T \beta_j \right), \quad \text{for } i = 1, \ldots, M, \quad j = 1, \ldots, K. \quad (8)$$

The hierarchical Bayesian approach assumes that all the parameter vectors $\beta_j$ form a random sample from a common population distribution, which is assumed to be the multivariate normal (MVN) distribution in this study, i.e.,

$$f(\beta_j | \mu_\beta, \Sigma_\beta) = \left(2\pi\right)^{-\nu/2} |\Sigma_\beta|^{-1/2} \exp \left[-\frac{1}{2} (\beta_j - \mu_\beta)^T \Sigma_\beta^{-1} (\beta_j - \mu_\beta) \right],$$

where $\nu$ is the dimensions of $\beta_j$.

To complete the model specification, the joint prior distribution for $\mu_\beta$ and $\Sigma_\beta$ needs to be specified. The conjugate prior for $\mu_\beta$, and $\Sigma_\beta$ is the $s$-independent normal-Wishart prior [27], i.e., a multivariate normal prior (with known mean $\theta_\beta$, and covariance matrix $C_\beta$) for $\mu_\beta$, and an inverse-Wishart prior for $\Sigma_\beta$. The inverse-Wishart distribution for a $\nu \times \nu$ symmetric positive-definite random matrix $X$ has the probability density function

$$f(X | \rho, S) = \left[2^{\nu/2} \pi^{(\nu+1)/2} \Gamma \left(\frac{\nu + 1}{2}\right) \right]^{-1} |S|^{\nu/2} |X|^{-(\nu + \rho + 1)/2} \exp \left[ -\frac{1}{2} \text{tr} \left( SX^{-1} \right) \right],$$

where $\rho$ is the degrees of freedom, and $S$ is a symmetric positive-definite $\nu \times \nu$ scale matrix. The inverse-Wishart prior for $\Sigma_\beta$ is computationally convenient; however, it is usually difficult to choose the parameters $\rho$ and $S$ if one wants to formulate a proper, noninformative prior. Letting $\rho$ equal $\nu + 1$, and $S$ be a $\nu \times \nu$ identity matrix $I_\nu$ is equivalent to setting a uniform prior defined over the interval $(-1, 1)$ to each correlation parameter [28]. The inverse-Wishart($\nu + 1, I_\nu$) prior may be reasonable for correlation estimation, but is quite constraining on the variance terms [29]. Therefore, we adopt the scaled inverse-Wishart prior (also called “separation-strategy prior” in [28], [30]) discussed in [29]. The covariance matrix $\Sigma_\beta$ is decomposed into variance and correlation components as

$$\Sigma_\beta = \Delta_\beta Q_\beta \Delta_\beta^T,$$

where the diagonal matrix $\Delta_\beta \equiv \text{diag}(\delta_{\beta_1}, \delta_{\beta_2}, \ldots, \delta_{\beta_\nu})$, and $\delta_{\beta_i} > 0$. and $Q_\beta$ is a symmetric positive-definite matrix. Then we assume an inverse-Wishart ($\rho_\beta, S_\beta$) prior distribution to the matrix $Q_\beta$, and a $s$-independent gamma prior distribution (with shape parameter $c_\beta$, and scale parameter $f_\beta$) to $\delta_{\beta_i}$, for $i = 1, \ldots, \nu$.

Applying Bayes’ theorem, we can derive the joint posterior distribution for all the model parameters conditioning on the data $d$ as

$$f(\beta_1, \ldots, \beta_K, \mu_\beta, Q_\beta, \delta_{\beta_1}, \ldots, \delta_{\beta_\nu} | d) \propto \prod_{j=1}^K \prod_{i=1}^M n_{ij}^{n_{ij}} e^{-\lambda_{ij}} |\Delta_\beta|^{-K/2} |Q_\beta|^{-K/2} |\Delta_\beta|^{-K/2} \times \exp \left[ -\frac{1}{2} \sum_{j=1}^K (\beta_j - \mu_\beta)^T \Delta_\beta^{-1} Q_\beta^{-1} \Delta_\beta^{-1} (\beta_j - \mu_\beta) \right] \times \exp \left[ -\frac{1}{2} (\mu_\beta - \theta_\beta)^T C_\beta^{-1} (\mu_\beta - \theta_\beta) \right] \times |Q_\beta|^{-1/2 + \rho_\beta + 1/2} \exp \left[ -\frac{1}{2} \text{tr} \left( S_\beta Q_\beta^{-1} \right) \right] \times \prod_{i=1}^\nu \delta_{\beta_i}^{-1} e^{-f_\beta \delta_{\beta_i}}.$$}

We use WinBUGS [31], a specialized software package for implementing Gibbs sampling [16], to perform MCMC simulation, and posterior inference. Each iteration of Gibbs sampling cycles through the unknown parameters, drawing a sample of one parameter conditional on the latest values of all the others. When the number of iterations is large enough, the sample draws on one parameter can be regarded as simulated observations from its marginal posterior distribution. Functions of the model parameters, such as the yield of the $j$th wafer, can also be conveniently sampled. It is usually straightforward to specify the conditional posterior distributions required for the Gibbs sampling. For the hierarchical Bayesian Poisson regression model, the conditional posterior distributions are given as follows.

- The conditional posterior distribution of $\beta_j$ is

$$f(\beta_j | \text{rest}, d) \propto \prod_{i=1}^M \exp \left[ -\exp \left( f(x_{ij})^T \beta_j \right) \right] \left[ \exp \left( f(x_{ij})^T \beta_j \right) \right]^{n_{ij}} \times \exp \left[ -\frac{1}{2} \sum_{j=1}^K (\beta_j - \mu_\beta)^T \Delta_\beta^{-1} Q_\beta^{-1} \Delta_\beta^{-1} (\beta_j - \mu_\beta) \right],$$

for $j = 1, \ldots, K$.

- The conditional posterior distribution of $\mu_\beta$ is

$$\mu_\beta | \text{rest}, d \sim \text{MVN} \left( \bar{\mu}_\beta \left( K \Delta_\beta^{-1} Q_\beta^{-1} \Delta_\beta^{-1} + C_\beta^{-1} \right)^{-1} \right),$$

where

$$\bar{\mu}_\beta = \left( K \Delta_\beta^{-1} Q_\beta^{-1} \Delta_\beta^{-1} + C_\beta^{-1} \right)^{-1} \left( K \Delta_\beta^{-1} Q_\beta^{-1} \Delta_\beta^{-1} \sum_{j=1}^K \beta_j + C_\beta^{-1} \theta_\beta \right).$$

- The conditional posterior distribution of $Q_\beta$ is

$$Q_\beta | \text{rest}, d \sim \text{inverse-Wishart} \left( \rho_\beta + K, S_\beta + \Delta_\beta^{-1} \sum_{j=1}^K (\beta_j - \mu_\beta)(\beta_j - \mu_\beta)^T \Delta_\beta^{-1} \right).$$

- The conditional posterior distribution of $\delta_{\beta_i}$ is

$$\delta_{\beta_i} | \text{rest}, d \sim \frac{\Gamma \left( \rho_\beta + c_\beta \right)}{\Gamma \left( \rho_\beta + c_\beta + 1 \right)} \frac{1}{\delta_{\beta_i}^2} e^{-\frac{1}{\delta_{\beta_i}^2} \left( \frac{1}{\delta_{\beta_i}} - \frac{f_\beta}{\delta_{\beta_i}} \right)}.$$
\[
f(\delta_{\mathbf{g}}|\text{rest}, \mathbf{d}) \\
\propto \delta_{\mathbf{g}}^{-K+\epsilon-1} \exp \left[ - \left( f_{\mathbf{g}} \delta_{\mathbf{g}} + \frac{V_{\mathbf{g},ij} \left[ Q_{\mathbf{g}}^{-1} \right]_{ij}}{2\delta_{\mathbf{g},ij}} ight) + \frac{1}{\delta_{\mathbf{g},ij}} \sum_{j \neq i} \left[ Q_{\mathbf{g}}^{-1} \right]_{ij} V_{\mathbf{g},ij}/\delta_{\mathbf{g},ij} \right],
\]
\tag{11}
\]

for \( i = 1, \ldots, \nu \), where \( \left[ Q_{\mathbf{g}}^{-1} \right]_{ij} \) is the \( i,j \)th element of \( Q_{\mathbf{g}}^{-1} \), and \( V_{\mathbf{g},ij} \) is the \( i,j \)th element of the matrix \( \sum_{j=1}^{K} (\beta_{j} - \mu_{g})(\beta_{j} - \mu_{g})^{T} \).

If we define the yield in the \( j \)th chip region on the \( j \)th wafer as the Poisson probability of having zero defects in that area, that is
\[
Y_{ij} = \exp(-\Lambda_{ij}) = \exp \left[ -\exp \left( f(x_{ij})^{T}\beta_{j} \right) \right],
\]
\( i = 1, \ldots, M, \quad j = 1, \ldots, K \),
then the wafer yield can be defined as the average portion of working chips per wafer:
\[
Y_{j} = \frac{1}{M} \sum_{i=1}^{M} Y_{ij} = \frac{1}{M} \sum_{i=1}^{M} \exp \left[ -\exp \left( f(x_{ij})^{T}\beta_{j} \right) \right],
\]
\( j = 1, \ldots, K. \) \tag{12}

At each iteration of the Gibbs sampling, \( Y_{j} \) is calculated based on the drawn at that iteration. Then a random sample for \( Y_{j} \) can be obtained after the completion of the Gibbs sampling. Point estimates based on the sample mean (or median) and Bayesian confidence interval for \( Y_{j} \) can then be conveniently computed.

The yield model given by (12) assumes that all the defects are faults. Because not all the defects cause immediate failures, the yield model can be easily modified to include a fault probability, which is the probability of a defect being fatal. Denoting the fault probability by \( \phi \), the yield is then defined as
\[
Y_{j} = \frac{1}{M} \sum_{i=1}^{M} \exp(-\phi \Lambda_{ij})
\]
\[
= \frac{1}{M} \sum_{i=1}^{M} \exp \left[ -\phi \exp \left( f(x_{ij})^{T}\beta_{j} \right) \right], \quad j = 1, \ldots, K.
\]
\tag{13}

The fault probability is the ratio of the critical area to the total area on which a defect can fall [1]. For simple devices, the critical areas may be calculated analytically [1]. For large commercial devices, advanced methods, such as the recent development of survey sampling based critical area estimation [32], can be used to extract the critical area information.

To predict the wafer yield for future production, denoted by \( Y_{p} \), we are interested in the posterior predictive distribution of \( Y_{p} \) (or equivalently \( \beta_{p} \), the predicted coefficient vector for a wafer from the future production). Random samples of \( Y_{p} \)

drawn from its posterior predictive distribution can be conveniently obtained using Gibbs sampling. At each iteration of the simulation, a random sample for \( \mu_{g} \), and a random sample for \( \Sigma_{g} \) are simulated. Then a random sample \( \beta_{p} \) is drawn from the MVN distribution using \( \mu_{g} \) and \( \Sigma_{g} \) at that iteration. Based on the simulated \( \beta_{p} \), \( Y_{p} \) is calculated according to the yield model given by (13). Upon the completion of the Gibbs sampling procedure, a random sample for \( Y_{p} \) has been obtained, and a posterior inference for \( Y_{p} \) can then be computed.

B. Negative Binomial Regression

The Poisson regression model discussed above extends the conventional Poisson yield model by including spatial covariates. The conventional negative binomial yield model given by (2) may also be improved by incorporating spatial covariates. Given the defect data \( \mathbf{d} \) on \( K \) wafers, a three-stage hierarchial Bayesian NB regression model can be setup as follows.

(i) \( P(N_{ij} = k) = (\Gamma(\alpha_{ij} + k)/\Gamma(k+1)) (\Lambda_{ij}/\alpha_{ij} + \alpha_{ij})^{-\alpha_{ij}} (\alpha_{ij}/\Lambda_{ij} + \alpha_{ij})^{\alpha_{ij}}, \quad k = 0, 1, \ldots, \lambda_{ij} = \exp(f(x_{ij})^{T}\beta_{j}), \quad i = 1, \ldots, M, \quad j = 1, \ldots, K \).

(ii) \( \beta_{j} \sim \text{MVN}(\mu_{\beta}, \Sigma_{\beta}), \quad \mu_{\beta} = \Delta_{g}\rho_{g}\Delta_{g}, \quad \Sigma_{g} = \text{diagonal}(\delta_{\beta,1}, \ldots, \delta_{\beta,\nu}), \quad \log(\alpha_{ij}) \sim \text{Normal}(\omega, \sigma^{2}), \quad f(x_{ij})^{T}\beta_{j} \).

(iii) \( \mu_{g} \sim \text{MVN}(\theta_{g}, C_{g}), \quad Q_{\beta} \sim \text{inverse - Wishart}(\rho_{g}, S_{g}), \quad \omega \sim \text{Normal}(a, b^{2}), \quad \sigma^{-2} \sim \text{Gamma}(c, d), \quad \delta_{\beta, i} \sim \text{Gamma}(e_{g}, f_{g}) \).

Herein, the first stage of the model uses the negative binomial model. One can also specify the first stage model based on the Poisson-Gamma mixture given by (3). Derivation of the joint posterior distribution, and the conditional posterior distributions of the NB regression model are presented in the Appendix.

The yield for the \( j \)th chip region on the \( j \)th wafer is defined as the negative binomial probability of having zero faults in that chip, that is
\[
Y_{ij} = \left( 1 + \frac{\phi \Lambda_{ij}}{\alpha_{ij}} \right)^{-\alpha_{ij}} \left( 1 + \frac{\phi \exp(f(x_{ij})^{T}\beta_{j})}{\alpha_{ij}} \right)^{-\alpha_{ij}}.
\]

Then the overall yield for the \( j \)th wafer is
\[
Y_{j} = \frac{1}{M} \sum_{i=1}^{M} Y_{ij} = \frac{1}{M} \sum_{i=1}^{M} \left( 1 + \frac{\phi \exp(f(x_{ij})^{T}\beta_{j})}{\alpha_{ij}} \right)^{-\alpha_{ij}}.
\]

C. Zero-Inflated Poisson Regression

As a result of defect clustering, IC chips with no defects are commonly observed, as well as chips with highly over-crowded defects in semiconductor manufacturing. The Poisson regression may not adequately account for clustered defect patterns with excessively many zero-defective chips. Therefore, yield models based on ZIP- regression were developed by Bae et al. [2]. The ZIP regression model proposed by Lambert [23] is very useful to model discrete count data with more zeros (called “inflated zeros”) than the Poisson model can accommodate.

Given the defect data \( \mathbf{d} \) on \( K \) wafers, a three-stage hierarchial Bayesian ZIP regression model can be setup as follows.

(i) \( P(N_{ij} = 0) = p_{ij} + (1 - p_{ij}) \exp(-\Lambda_{ij}), \quad P(N_{ij} = k) = (1 - p_{ij}) \exp(-\Lambda_{ij}) \Lambda_{ij}^{k}/k!, \quad k = 1, 2, \ldots, \log(\Lambda_{ij}) = f(x_{ij})^{T}\beta_{j}, \quad \logit(p_{ij}) = \).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
and \( \beta_j \), \( \Sigma_\beta \), \( \Delta_\beta \) are the covariates vectors evaluated at \( x_{ij} \), \( y_{ij} \), and \( g(x_{ij}) \) may be chosen to be the same (e.g., in [17]), or different (e.g., in [2]). In this study, we assume they are identical.

The yield in the \( j \)th chip region on the \( j \)th wafer is defined as

\[
Y_{ij} = P(N_{ij} = 0) = p_{ij} + (1 - p_{ij}) \exp(-\phi \Lambda_{ij}),
\]

and the overall yield for the \( j \)th wafer is

\[
Y_j = \frac{1}{M} \sum_{i=1}^{M} Y_{ij} = \frac{1}{M} \sum_{i=1}^{M} (p_{ij} + (1 - p_{ij}) \exp(-\phi \Lambda_{ij})).
\]

### D. Zero-Inflated Negative Binomial Regression

The ZIP regression model discussed in Section II-C can be generalized to the case of ZINB regression. Given the defect data \( d \) on \( K \) wafers, a three-stage hierarchical Bayesian ZINB regression model can be setup as follows.

(i) \( P(N_{ij} = 0) = p_{ij} + (1 - p_{ij})(1 + (\lambda_{ij}/\alpha_j))^{-\alpha_j} \),

\( P(N_{ij} = k) = (1 - p_{ij})(\lambda_{ij}/\alpha_j)^k/(k + 1)\Gamma(k + 1\alpha_j)/(\alpha_j + \lambda_{ij})^{k+\lambda_{ij}} \), \( k = 1, 2, \ldots \),

\[
\log(\lambda_{ij}) = g(x_{ij})^T \beta_j, \quad \text{logit}(p_{ij}) = g(x_{ij})^T \gamma_j,
\]

for \( i = 1, \ldots, M \) and \( j = 1, \ldots, K \).

(ii) \( \beta_j \sim \text{MVN}(\mu_\beta, \Sigma_\beta) \), \( \Sigma_\beta = \Delta_\beta Q_\beta \Delta_\beta \), and \( \Delta_\beta = \text{diag}(\delta_{\beta,1}, \ldots, \delta_{\beta,\nu}) \), \( \gamma_j \sim \text{MVN}(\mu_\gamma, \Sigma_\gamma) \), where \( \Sigma_\gamma = \Delta_\gamma Q_\gamma \Delta_\gamma \), and \( \Delta_\gamma = \text{diag}(\delta_{\gamma,1}, \ldots, \delta_{\gamma,\nu}) \). If \( p_{ij} = 0 \), then the ZIP regression model is reduced to the regular Poisson regression model discussed in Section II-A. Herein, \( f(x_{ij}, \gamma_j) \) and \( g(x_{ij}) \) are the covariates vectors evaluated at \( x_{ij} \), \( y_{ij} \), and \( g(x_{ij}) \) may be chosen to be the same (e.g., in [17]), or different (e.g., in [2]). In this study, we assume they are identical.

The yield in the \( j \)th chip region on the \( j \)th wafer is then defined as

\[
Y_{ij} = \frac{1}{M} \sum_{i=1}^{M} Y_{ij} = \frac{1}{M} \sum_{i=1}^{M} (p_{ij} + (1 - p_{ij}) \exp(-\phi \Lambda_{ij})).
\]
TABLE I
WAFER YIELD ESTIMATES BY THE CONVENTIONAL POISSON AND NEGATIVE BINOMIAL YIELD MODELS (??) AND (2), ALONG WITH THEIR RELATIVE BIASES

<table>
<thead>
<tr>
<th>Wafer</th>
<th>True yield</th>
<th>Poisson yield model</th>
<th>Negative binomial yield model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y$</td>
<td>$\hat{Y}_{\text{poisson}}$</td>
<td>$\hat{Y}_{\text{nb}}$</td>
</tr>
<tr>
<td>1</td>
<td>0.844</td>
<td>0.749</td>
<td>-11.26%</td>
</tr>
<tr>
<td>2</td>
<td>0.899</td>
<td>0.879</td>
<td>-2.22%</td>
</tr>
<tr>
<td>3</td>
<td>0.793</td>
<td>0.722</td>
<td>-8.95%</td>
</tr>
</tbody>
</table>

and the overall yield for the $j$th wafer is

$$Y_j = \frac{1}{M} \sum_{i=1}^{M} Y_{ij} = \frac{1}{M} \sum_{i=1}^{M} \left( p_{ij} + (1-p_{ij}) \left( 1 + \frac{\phi \Delta_{ij}}{\alpha_j} \right)^{-\alpha_j} \right).$$

III. EXAMPLES, AND RESULTS

A. Practical Examples

This section uses three real wafer maps provided by a DRAM manufacturer to illustrate the hierarchical Bayesian models developed in this study. Fig. 1 shows the spatial defect counts on the three wafers. The diameter of the wafers is 20 cm, and each wafer contains 473 potentially usable chips. For simplicity and convenience, we assume that the fault probability $\phi = 1$, that is, every defect will cause the failure of the chip. The first wafer shown in Fig. 1(a) contains 74 defective chips (a chip is considered as defective if it has at least one defect), and hence the actual wafer yield is $1 - \frac{74}{473} = 0.844$. Fig. 1(a) shows two defect clusters: one at the upper left portion, and another one at the upper right portion of the wafer. The second wafer shown in Fig. 1(b) has 48 defective chips, and the actual yield is 0.899. Most of the defective chips are located along the right edge of the wafer. Fig. 1(c) shows the defect counts on the third wafer with 98 defective chips, and the yield of 0.793. The third wafer has an obvious defect cluster with a linear pattern along its centerline. Wafer yields estimated by the conventional Poisson yield model, and the negative binomial yield model (2) are summarized in Table I. We define the relative bias of a yield estimation as

$$\text{relative bias} = \frac{\text{estimated yield} - \text{true yield}}{\text{true yield}} \times 100\%,$$

where a negative (positive) relative bias indicates an underestimation (overestimation) of the true yield. The Poisson yield model underestimates the true yield for all three wafers, especially for the first and the third wafers, which show higher clustering effect than the second wafer. The negative binomial yield model substantially enhances the accuracy of the yield estimation by considering the dispersion (or clustering) effect of defects scattered on the wafer in comparison with the Poisson model.

Maximum likelihood analysis for the Poisson regression, NB regression, and ZIP regression were summarized in Bae et al. [2], and for the ZINB regression in [21]. This study uses the R package zicounts [22] to perform the ML analysis for all the four regression models. The ML analysis considers each wafer individually, therefore it only considers the spatial variation of defect counts within an individual wafer. Table II shows the yield estimates obtained by the ML analysis. The Poisson regression model provides slightly better yield estimation than the conventional Poisson model. However, both models highly underestimate the true yields. The NB regression model, by introducing an over-dispersion factor, and the zero-inflated regression models, by explicitly modeling “extra” zero-defective chips, provide more accurate yield estimation than the Poisson regression model. The NB regression model provides the most accurate yield estimation among the four regression models. By incorporating spatial covariates, the NB regression model significantly improves accuracy of yield estimation over the conventional negative binomial yield model.

Table III shows the yields estimated by the hierarchical Bayesian models proposed in this study. The posterior medians and 95% confidence intervals constructed by using the 2.5% and 97.5% sample percentiles (shown in parenthesis) are listed. The prior distributions play an important role in the Bayesian analysis. They should reflect our honest knowledge of the model parameters. One needs to be cautious in the specification of the prior distributions because inappropriate priors may lead to wrong analysis results and conclusions. The prior knowledge may be obtained from historical data of the same or similar products, and from experts’ opinions. In this study, we specify the parameter values of the prior distributions to reflect our absence of prior knowledge of the model parameters, and hence posterior inference is made mainly based on information from the data. Especially, we assume $\theta_{\beta} = \theta_{\gamma} = (0, 0, 0, 0, 0, 0)$, $C_{\beta} = C_{\gamma} = 10^{6} I_{6}$, $Q_{\beta}$, and $Q_{\gamma}$ have $s$-independent inverse-Wishart $(\nu + 1, I_{\nu})$ prior distributions [29]; and $\theta_{\beta}$, and $\theta_{\gamma}$ have $s$-independent Gamma$(0.001, 0.001)$ prior distributions. For the NB regression, and the ZINB regression models, we also assume $\omega \sim \text{Normal}(0, 10^{6})$, and $\sigma^{-2} \sim \text{Gamma}(0.001, 0.001)$. Herein $\text{Normal}(0, 10^{6})$, and $\text{Gamma}(0.001, 0.001)$ are two noninformative, proper prior distributions that have been widely used in the Bayesian literature [33].

Table IV compares the average of absolute relative biases of the conventional yield models, the ML analysis of the four regression models, and the hierarchical Bayesian regression models. The ML analysis, and the hierarchical Bayesian approach produce comparable results. For the Poisson regression,
The NB regression, the ZIP regression, and the ZINB regression models provide comparable predictions, which are much higher than the yield predicted by the Poisson regression model. Note that the hierarchical Poisson regression model produces improper confidence intervals in comparison with the other three models.

**B. Simulation Study**

In this section, we use simulated wafer maps to compare the performance of the different yield models. The simulation study is conducted as follows. Five sets of wafer maps with typical clustering patterns are simulated. The number of simulated wafers in each set is randomly chosen between 3 and 10. Fig. 3 shows a representative set with five simulated wafers. The diameter of the wafers is 20 cm. The spatial defect patterns are simulated as the superposition of random defect patterns, and clustered defect patterns. The random defects are simulated according to the spatial homogeneous Poisson process. The expected number of random defects on a wafer is randomly distributed between 30 and 150. We simulate four typical types of clustering patterns: linear, curvilinear, amorphous, and ring-shaped. The number of defects in each cluster is randomly chosen between 50 and 150. Defects in amorphous or linear clusters are simulated from the bivariate normal distributions with random mean vectors, and variance-covariance matrices. Defects in curvilinear or ring-shaped clusters are generated from the assumption that the defects are distributed uniformly along and about an arc or circle. As shown in Fig. 3, wafer (a) does not contain clustered defects; wafer (b) has one amorphous cluster; wafer (c) has one amorphous cluster, and one curvilinear cluster; wafer (d) has two separated curvilinear clusters; and wafer (e) contains one ring-shaped cluster overlapped with a curvilinear cluster.

Assuming a die size of 1.5220 cm × 0.3480 cm, each wafer contains 473 chips. Defect counts data can be obtained by counting the number defects in each die. The simulated wafers in each set are analyzed using the ML regression models separately, and using the hierarchical Bayesian models simultaneously. The estimated yields are then compared with the observed actual yields. Table VI compares the average of absolute relative biases of the ML analysis of the four regression models and the hierarchical Bayesian regression models over all the five sets of simulated wafers. Again, the ML regression models and the hierarchical Bayesian regression models provide consistent, similar results. On average, the ZIP regression models perform the best. For simulated wafers without clustered defects, e.g., wafer (a) in Fig. 3, the Poisson regression models provide the most accurate yield estimation as expected, and ZINB regression models provide the worst estimation. For wafers with clustered defects, the ZINB regression models generally perform the best. For example, the hierarchical Bayesian ZINB regression model perfectly estimates the yields of wafer (c) and wafer (d) in the Fig. 3. In practice, this result suggests that yield estimation and prediction can be combined with the diagnostic analysis on wafer defect data. If the defect data do not show clear clustering patterns, the Poisson models may be used. On the other hand, if clustering effects are apparent, the

---

**TABLE IV**

<table>
<thead>
<tr>
<th>Conventional Models</th>
<th>Regression Models (Maximum Likelihood)</th>
<th>Regression Models (Hierarchical Bayesian)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>NB</td>
<td>Poisson NB ZIP ZINB</td>
</tr>
<tr>
<td>7.48</td>
<td>1.09</td>
<td>6.65 0.20 0.80 0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.64 0.61 0.51 0.20</td>
</tr>
</tbody>
</table>

---

**TABLE V**

<table>
<thead>
<tr>
<th>Wafer</th>
<th>True yield</th>
<th>Poisson</th>
<th>NB</th>
<th>ZIP</th>
<th>ZINB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Y rel. bias</td>
<td>Y rel. bias</td>
<td>Y rel. bias</td>
<td>Y rel. bias</td>
<td>Y rel. bias</td>
</tr>
<tr>
<td>1</td>
<td>0.844</td>
<td>0.763</td>
<td>0.835</td>
<td>0.837</td>
<td>0.839</td>
</tr>
<tr>
<td></td>
<td>-0.06%</td>
<td>-1.07%</td>
<td>0.835</td>
<td>0.837</td>
<td>-0.59%</td>
</tr>
<tr>
<td></td>
<td>(0.728, 0.795)</td>
<td>(0.803, 0.866)</td>
<td>(0.804, 0.866)</td>
<td>(0.808, 0.867)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.899</td>
<td>0.881</td>
<td>0.899</td>
<td>0.895</td>
<td>0.899</td>
</tr>
<tr>
<td></td>
<td>-2.00%</td>
<td>0.00%</td>
<td>0.895</td>
<td>0.44%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>(0.852, 0.907)</td>
<td>(0.872, 0.923)</td>
<td>(0.898, 0.923)</td>
<td>(0.898, 0.923)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.703</td>
<td>0.727</td>
<td>0.790</td>
<td>0.785</td>
<td>0.783</td>
</tr>
<tr>
<td></td>
<td>-8.32%</td>
<td>0.76%</td>
<td>0.785</td>
<td>0.25%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>(0.690, 0.763)</td>
<td>(0.763, 0.823)</td>
<td>(0.756, 0.880)</td>
<td>(0.758, 0.827)</td>
<td></td>
</tr>
</tbody>
</table>
TABLE V
POSTERIOR YIELD PREDICTION OF FUTURE PRODUCTION

<table>
<thead>
<tr>
<th>Model</th>
<th>Posterior median</th>
<th>95% Bayesian confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson regression</td>
<td>0.744</td>
<td>(0.000, 1.000)</td>
</tr>
<tr>
<td>NB regression</td>
<td>0.835</td>
<td>(0.775, 0.990)</td>
</tr>
<tr>
<td>ZIP regression</td>
<td>0.830</td>
<td>(0.808, 1.000)</td>
</tr>
<tr>
<td>ZINB regression</td>
<td>0.839</td>
<td>(0.786, 1.000)</td>
</tr>
</tbody>
</table>

ZINB models may be introduced to predict the yield of the wafers.

IV. CONCLUSION

In this study, we propose hierarchical Bayesian approaches for modeling the spatial defects over wafers in IC manufacturing. We use the spatial locations of the IC chips on the wafers as covariates, and develop four models based on Poisson regression, NB regression, ZIP regression, and ZINB regression. Three real wafer maps are used to illustrate the proposed approaches. Results from the hierarchical Bayesian models are compared to the results obtained from ML analysis of the four
regression models, as well as those from the conventional yield models. The ML method analyses each wafer individually, and does not consider random effects for the model parameters. In general, the hierarchical Bayesian approaches and the ML analysis produce very similar results for yield estimation. However, one distinct advantage of the hierarchical Bayesian approaches over the ML analysis is that the hierarchical Bayesian approaches can effectively model the spatial variations of defects within each wafer, as well as across different wafers. Therefore, the hierarchical Bayesian approaches are more desirable in predicting the yield for future production.

The Poisson regression models consistently underestimate the true yield due to extraneous Poisson variation caused by the defect clustering effect. The fundamental assumption of the Poisson yield models is that the defects are randomly distributed, and the occurrence of a defect at any location is s-independent of the occurrence of any other defect (i.e., conditional s-independence). If clusters of defects are apparent in the diagnostic analysis on wafer defect data, it is inappropriate to use the Poisson yield models. On the contrary, the NB regression, by taking clustering effects into consideration through the introduction of an overdispersion factor, the ZIP regression, by explicitly considering the excessive amount of zero-defective chips, and the ZINB regression, by considering both of them, provide more accurate yield estimation and prediction results. In particular, the ZINB regression model showed an attractive power as a yield model over conventional yield models by incorporating both over-dispersion effects and zero-inflated effects of defects on the wafers.

Defects on the semiconductor wafers are likely to be caused by different mechanisms. Random causes such as particles in the cleanrooms, and thermal variation in annealing processes generate global defects all over the wafer surfaces. Assignable causes such as human mistakes, particles from equipment, and chemical stains generate defects in clusters. Therefore, it will be desirable that a yield model separately captures yield losses due to different causes, hence providing valuable information for yield learning and improvement. This approach may be achieved by combining the proposed yield models with some defect clustering and classification techniques, i.e., the model-based clustering approach [5], which will be investigated in the future.

The yield models developed in this study are based on spatial modeling of defect counts on the wafers. It is also possible to construct yield models by directly modeling the defects using spatial point processes, e.g., spatial processes discussed by van Lieshout [34], and Daley and Vere-Jones [35]. There have been attempts to model yield using spatial marked point processes [36], and nonhomogeneous Poisson germ processes [37]. Further studies will be conducted to explore other spatial point processes for modeling spatial defects, and to apply the hierarchical Bayesian approach to the models.

### APPENDIX

**Posterior Derivation of the Hierarchical NB Regression Model**

The joint posterior distribution of all the model parameters involved in the three-stage hierarchical Bayesian NB regression model is given by

$$
f(\beta_1, \ldots, \beta_K, \alpha_1, \ldots, \alpha_K, \mu_0, \mathbf{Q}_0, \delta_{0,1}, \ldots, \delta_{0,i}, \omega, \sigma^2 | \mathbf{d})
$$

$$
\propto \prod_{j=1}^{K} \prod_{i=1}^{M} \frac{\Gamma(\alpha_j + n_{ij})}{\Gamma(\alpha_j)} \frac{1}{\Gamma(n_{ij} + 1)} \left( \frac{\Lambda_{ij}}{\Lambda_{ij} + \alpha_j} \right)^{n_{ij}} \left( \frac{\alpha_j}{\Lambda_{ij} + \alpha_j} \right)^{\alpha_j}
$$

$$
\times |\Delta_0|^{-K/2} |\mathbf{Q}_0|^{-K/2} |\Delta_0|^{-K/2}
$$

$$
\times \exp \left[ -\frac{\sum_{j=1}^{K} (\beta_j - \mu_0)^T \Delta_0^{-1} \mathbf{Q}_0^{-1} \Delta_0^{-1} (\beta_j - \mu_0)}{2} \right]
$$

$$
\times \exp \left[ -\frac{(\mu_0 - \theta_0)^T \mathbf{C}_0^{-1} (\mu_0 - \theta_0)}{2} \right]
$$

$$
\times |\mathbf{Q}_0|^{-\left(\nu + \rho_0 + 1\right)/2} \exp \left[ -\frac{\text{tr} \left( \mathbf{S}_0 \mathbf{Q}_0^{-1} \right)}{2} \right]
$$

$$
\times (\sigma^2)^{-K/2} \prod_{j=1}^{K} \alpha_j^{-1} \exp \left[ -\frac{\sum_{j=1}^{K} (\log \alpha_j - \omega)^2}{2\sigma^2} \right]
$$

$$
\times \prod_{i=1}^{\nu} \delta_{0,i}^{-1} \exp \left[ -f_0 \delta_{0,i} \right] \exp \left[ -\frac{(\omega - \delta_{0,i})^2}{2\sigma^2} \right]
$$

$$
\times (\sigma^2)^{-\left(\nu + 1\right)/2} \exp \left( -\frac{d}{\sigma^2} \right).
$$

The conditional posterior distributions of \( \mu_0, \mathbf{Q}_0, \) and \( \delta_{0,i} \) are the same as those given by (9)–(11), respectively. The condi-
tional posterior distributions of other parameters are given as follows.
• The conditional posterior distribution of $\beta_j$ is

$$f(\beta_j|\text{rest., d}) \propto \prod_{i=1}^{M} \left( \frac{\alpha_j}{\alpha_j + \beta_j} \right)^{n_{ij}} \left( \frac{\alpha_j}{\alpha_j + \beta_j} \right)^{\alpha_j} \exp \left[ -\frac{(\beta_j - \mu_\beta)^T \Delta_{\beta}^{-1} \Delta_{\beta}^{-1}(\beta_j - \mu_\beta)}{2} \right],$$

for $j = 1, \ldots, K$.

• The conditional posterior distribution of $\alpha_j$ is

$$f(\alpha_j|\text{rest., d}) \propto \prod_{i=1}^{M} \left( \frac{\Gamma(\alpha_j + n_{ij})}{\Gamma(\alpha_j)} \right) \frac{\alpha_j}{\alpha_j + \beta_j}^{n_{ij}} \exp \left[ -\frac{(\log \alpha_j - \omega)^2}{2\tau^2} \right],$$

for $j = 1, \ldots, K$.

• The conditional posterior distribution of $\omega$ is

$$\omega|\text{rest., d} \sim \text{Normal}\left((K\sigma^2 + b^{-2})^{-1} \times \left( \sigma^2 - \sum_{j=1}^{K} \log \alpha_j + b^{-2} \right), (K\sigma^2 + b^{-2})^{-1}\right),$$

(14)

• The conditional posterior distribution of $\sigma^2$ is

$$\sigma^2|\text{rest., d} \sim \text{Inverse-Gamma}\left(c + \frac{K}{2}, d + \sum_{j=1}^{K} (\log \alpha_j - \omega)^2 \right).$$

(15)

### Posterior Derivation of the Hierarchical ZIP Regression Model

The joint posterior distribution of all the model parameters in the three-stage hierarchical Bayesian ZIP regression model, shown at the bottom of the page, where $I(\cdot)$ is the indicator function. The conditional posterior distributions of $\mu_\beta$, $Q_\beta$, and $\delta_\beta$ are given by (9)–(11), respectively. The conditional posterior distributions of $\mu_\gamma$, $Q_\gamma$, and $\delta_\gamma$ can be obtained by replacing $\beta$ with $\gamma$ in (9)–(11). The conditional posterior distributions of $\beta_j$ and $\gamma_j$ are given as follows.

• The conditional posterior distribution of $\beta_j$ is

$$f(\beta_j|\text{rest., d}) \propto \prod_{i=1}^{M} \left[ p_{ij} + (1 - p_{ij}) \exp(-\Lambda_{ij}) \right]^{n_{ij} - \eta_{ij}} \times \left[ 1 - p_{ij} \exp(-\Lambda_{ij}) \right]^{n_{ij} \cdot I(n_{ij} = 0)} \times \exp \left[ -\frac{\sum_{j=1}^{K} (\beta_j - \mu_\beta)^T \Delta_{\beta}^{-1} \Delta_{\beta}^{-1}(\beta_j - \mu_\beta)}{2} \right],$$

for $j = 1, \ldots, K$.

• The conditional posterior distribution of $\gamma_j$ is

$$f(\gamma_j|\text{rest., d}) \propto \prod_{i=1}^{M} \left[ p_{ij} + (1 - p_{ij}) \exp(-\Lambda_{ij}) \right]^{n_{ij} - \eta_{ij}} \times \left[ 1 - p_{ij} \exp(-\Lambda_{ij}) \right]^{n_{ij} \cdot I(n_{ij} = 0)} \times \exp \left[ -\frac{\sum_{j=1}^{K} (\gamma_j - \mu_\gamma)^T \Delta_{\gamma}^{-1} \Delta_{\gamma}^{-1}(\gamma_j - \mu_\gamma)}{2} \right],$$

for $j = 1, \ldots, K$. 

---

$$f(\beta_1, \ldots, \beta_K, \gamma_1, \ldots, \gamma_K, \mu_\beta, Q_\beta, \delta_\beta, \mu_\gamma, Q_\gamma, \delta_\gamma|\text{d}) \propto \prod_{j=1}^{K} \prod_{i=1}^{M} \left[ p_{ij} + (1 - p_{ij}) \exp(-\Lambda_{ij}) \right]^{n_{ij} - \eta_{ij}} \times \left[ 1 - p_{ij} \exp(-\Lambda_{ij}) \right]^{n_{ij} \cdot I(n_{ij} = 0)} \times \exp \left[ -\frac{\sum_{j=1}^{K} (\beta_j - \mu_\beta)^T \Delta_{\beta}^{-1} \Delta_{\beta}^{-1}(\beta_j - \mu_\beta)}{2} \right] \times \exp \left[ -\frac{\sum_{j=1}^{K} (\gamma_j - \mu_\gamma)^T \Delta_{\gamma}^{-1} \Delta_{\gamma}^{-1}(\gamma_j - \mu_\gamma)}{2} \right] \times \exp \left[ -\frac{\sum_{j=1}^{K} (\eta_{ij} - \mu_\gamma)^T \Delta_{\gamma}^{-1} \Delta_{\gamma}^{-1}(\eta_{ij} - \mu_\gamma)}{2} \right] \times \exp \left[ -\frac{\sum_{j=1}^{K} (\delta_{ij} - \mu_\beta)^T \Delta_{\beta}^{-1} \Delta_{\beta}^{-1}(\delta_{ij} - \mu_\beta)}{2} \right] \times \exp \left[ -\frac{\sum_{j=1}^{K} (\delta_{ij} - \mu_\gamma)^T \Delta_{\gamma}^{-1} \Delta_{\gamma}^{-1}(\delta_{ij} - \mu_\gamma)}{2} \right] \times \frac{\nu^{(\nu + \mu_\beta + 1)/2} \exp \left[ -\frac{\nu}{2} \right]}{\Gamma(\nu/2)} \times \frac{\nu^{(\nu + \mu_\gamma + 1)/2} \exp \left[ -\frac{\nu}{2} \right]}{\Gamma(\nu/2)} \times \prod_{i=1}^{M} \exp(-f_{\beta_i} \delta_{\beta_i}) \times \prod_{i=1}^{M} \exp(-f_{\gamma_i} \delta_{\gamma_i}).$$
Posterior Derivation of the Hierarchical ZINB Regression Model

The joint posterior distribution of all the model parameters in the three-stage hierarchical Bayesian ZINB regression model, shown at the bottom of the page. The conditional posterior distributions of $\mu_\beta$, $Q_\beta$, $\delta_\beta$, $\mu_\gamma$, $Q_\gamma$, and $\delta_\gamma$ are the same as those in the hierarchical ZIP regression model. The conditional posterior distributions of $\omega$, and $\sigma^2$ are given by (14), and (15) respectively. The conditional posterior distributions of other parameters are given as follows.

- The conditional posterior distribution of $\beta_j$ is

$$f(\beta_j|\mathbf{d}) \propto \prod_{i=1}^{M} \left[ p_{ij} + (1-p_{ij}) \left( 1 + \frac{A_{ij}}{\alpha_j} \right)^{-\alpha_j} \right]^{I(n_{ij}=0)} \times \left[ \frac{\Gamma(\alpha_j+n_{ij})}{\Gamma(n_{ij}+1)\Gamma(\alpha_j)} \alpha_j^{\frac{n_{ij}}{2}} A_{ij}^{\alpha_j-1} \right]^{1-I(n_{ij}=0)} \times \exp \left[ -\sum_{j=1}^{K} (\beta_j - \mu_\beta)^T \delta_\beta^{-1} Q_\beta^{-1} \delta_\beta^{-1} (\beta_j - \mu_\beta) \right] \text{ for } j = 1, \ldots, K.

- The conditional posterior distribution of $\gamma_j$ is

$$f(\gamma_j|\mathbf{d}) \propto \prod_{i=1}^{M} \left[ p_{ij} + (1-p_{ij}) \left( 1 + \frac{A_{ij}}{\alpha_j} \right)^{-\alpha_j} \right]^{I(n_{ij}=0)} \times \left[ \frac{\Gamma(\alpha_j+n_{ij})}{\Gamma(n_{ij}+1)\Gamma(\alpha_j)} \alpha_j^{\frac{n_{ij}}{2}} A_{ij}^{\alpha_j-1} \right]^{1-I(n_{ij}=0)} \times \exp \left[ -\sum_{j=1}^{K} (\gamma_j - \mu_\gamma)^T \delta_\gamma^{-1} Q_\gamma^{-1} \delta_\gamma^{-1} (\gamma_j - \mu_\gamma) \right] \text{ for } j = 1, \ldots, K.

$$

\[f(\beta_1, \ldots, \beta_K; \gamma_1, \ldots, \gamma_K; \mu_\beta, Q_\beta, \delta_\beta, 1, \ldots, \delta_\gamma, \mu_\gamma, Q_\gamma, \delta_\gamma, \alpha_1, \ldots, \alpha_K, \omega, \sigma^2|\mathbf{d}) \propto \prod_{j=1}^{K} \prod_{i=1}^{M} \left[ p_{ij} + (1-p_{ij}) \left( 1 + \frac{A_{ij}}{\alpha_j} \right)^{-\alpha_j} \right]^{I(n_{ij}=0)} \times \left[ \frac{\Gamma(\alpha_j+n_{ij})}{\Gamma(n_{ij}+1)\Gamma(\alpha_j)} \alpha_j^{\frac{n_{ij}}{2}} A_{ij}^{\alpha_j-1} \right]^{1-I(n_{ij}=0)} \times \exp \left[ -\sum_{j=1}^{K} (\beta_j - \mu_\beta)^T \delta_\beta^{-1} Q_\beta^{-1} \delta_\beta^{-1} (\beta_j - \mu_\beta) \right] \times \exp \left[ -\sum_{j=1}^{K} (\gamma_j - \mu_\gamma)^T \delta_\gamma^{-1} Q_\gamma^{-1} \delta_\gamma^{-1} (\gamma_j - \mu_\gamma) \right] \times \exp \left[ -\frac{\text{tr}(S_\beta Q_\beta)}{2} \right] \times \exp \left[ -\frac{\text{tr}(S_\gamma Q_\gamma)}{2} \right] \times \prod_{i=1}^{M} \exp \left( -f_\beta \delta_{\beta,i} \right) \times \prod_{i=1}^{M} \exp \left( -f_\gamma \delta_{\gamma,i} \right) \times \exp \left( -\frac{(\omega - \omega)^2}{2\sigma^2} \right) \times \exp \left( -(\omega - \omega)^2 \right) \times \exp \left( -\frac{d}{\sigma^2} \right), \]

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
REFERENCES


Tao Yuan (S’03–M’06) is an Assistant Professor in the Department of Industrial and Systems Engineering at Ohio University. He received the B.E. degree in thermal engineering from Tsinghua University, China; the M.S. degree in aerospace engineering, and the M.E. degree in industrial engineering from Texas A&M University, College Station; and the Ph.D. degree in industrial engineering from the University of Tennessee, Knoxville. His research interests are in reliability modeling and prediction, as well as reliability testing and characterization of micro and nano devices. He is a member of IEEE, IIE, and INFORMS.

Saleem Z. Ramadan is an Assistant Professor in the Department of Mechanical and Industrial Engineering Applied Science University, Shafa Badran, Jordan. He received his Ph.D. from the Department of Industrial and Systems Engineering at Ohio University in 2011. His research interests are in Bayesian statistics, reliability, genetic algorithms, and operations research.

Suk Joo Bae is an Assistant Professor in the Department of Industrial Engineering at Hanyang University, Seoul, Korea. He received his Ph.D. from the School of Industrial and Systems Engineering at Georgia Institute of Technology in 2003. He worked as a reliability engineer at SDI, Korea, from 1996 to 1999. His research interests are in reliability evaluation of light displays, and nano-devices via accelerated life and degradation testing, statistical robust parameter design, and process control for large-volume on-line processing data. He is a member of INFORMS, ASA, and IMS.